Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тульский государственный университет»
Институт права и управления

УТВЕРЖДАЮ

Директор Института права и управления

____________ М.А. Берестнев

«26» января 2023 г.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ПРАКТИЧЕСКИХ РАБОТ ПО ДИСЦИПЛИНЕ

ОП.02 СТАТИСТИКА

для специальности

38.02.06 Финансы

Тула 2023

РАССМОТРЕНА
цикловой комиссией общепрофессиональных дисциплин

Протокол от «26» января 2023 г. №6

Председатель цикловой комиссии
А.Л.Сабинина, д.э.н., проф.каф.ФиМ

Автор:
Макарова Н.Н., к.т.н., доцент кафедры Финансы и менеджмент

Занятие 1. Статистическая сводка и группировка данных

Группировка – это разбиение совокупности на группы, однородные по какому-либо признаку. С точки зрения отдельных единиц совокупности группировка – это объединение отдельных единиц совокупности в группы, однородные по каким-либо признакам.

Группировочный признак – это признак, по которому происходит объединение отдельных единиц совокупности в однородные группы.

Интервал очерчивает количественные границы групп. Как правило, он представляет собой промежуток между максимальными и минимальными значениями признака в группе.

При проведении группировки приходится решать ряд задач:
1) выделение группировочного признака;
2) определение числа групп и величины интервалов;
3) при наличии нескольких группировочных признаков описание того, как они комбинируются между собой;
4) установление показателей, которыми должны характеризоваться группы, т.е. сказуемого группировки.

Статистические группировки и классификации преследуют цели выделения качественно однородных совокупностей, изучения структуры совокупности, исследования существующих зависимостей. Каждой из этих целей соответствует особый вид группировки: типологическая, структурная, аналитическая (факторная).

Типологическая группировка решает задачу выявления и характеристики социально-экономических типов (частных подсовокупностей).

Структурная дает возможность описать составные части совокупности или строение типов, а также проанализировать структурные сдвиги.

Аналитическая (факторная) группировка позволяет оценивать связи между взаимодействующими признаками.

Структурная группировка применяется для характеристики структуры совокупности и структуры сдвигов.

Структурной называется группировка, в которой происходит разделение выделенных с помощью технологической группировки типов явлений, однородных совокупностей на группы, характеризующие их структуру по какому-либо варьирующему признаку. Например, группировка населения по размеру среднедушевого дохода.

Определение числа групп. Здесь необходимо учитывать несколько условий:
а) число групп детерминируется уровнем колеблемости группировочного признака. Чем значительнее вариация признака, тем больше при прочих равных условиях должно быть число групп;
б) число групп должно отражать реальную структуру изучаемой совокупности;
в) не допускается выделение пустых групп;
г) группы должны быть взаимоисключающими;
д) группы должны быть исчерпывающими.

Для нахождения числа групп может быть использована формула Стерджесса:

$$n = 1 + 3{,}322 \cdot \lg N,$$

где N – количество элементов совокупности.

В случае равных интервалов величина интервала может быть определена как

$$i = \frac{X_{max} - X_{min}}{n}$$

или

$$i = \frac{X_{max} - X_{min}}{1 + 3{,}322 \cdot \lg N}.$$

Показатель численности групп представлен либо частотой (количеством единиц в каждой группе), либо частотностью (удельным весом каждой группы).

Среди простых группировок особо выделяют ряды распределения.

Ряд распределения – это группировка, в которой для характеристики групп (упорядоченно расположенных по значению признака) применяется один показатель – численность группы. Другими словами, это ряд чисел, показывающий, как распределяются единицы некоторой совокупности по изучаемому признаку.

Ряды, построенные по атрибутивному признаку, называются *атрибутивными рядами распределения.*

Ряды распределения, построенные по количественному признаку, называются *вариационными рядами.*

Примером атрибутивных рядов могут служить распределения населения по полу, занятости, национальности, профессии и т.д.

Пример атрибутивного ряда

Категория преподавателей	Число преподавателей	В % к итогу
Стажеры	20	5
Ассистенты	100	25
Старшие преподаватели	40	10
Доценты	160	40
Профессора	40	10
Итого	400	100

Примером вариационного ряда распределения могут служит распределения населения по возрасту, рабочих – по стажу работы, заработной плате и т.д.

Вариационные ряды распределения состоят их двух элементов вариантов и частот.

Вариантами называются числовые значения количественного признака в ряду распределения, они могут быть положительными и отрицательными, абсолютными и относительными.

Частоты – это численности отдельных вариантов или каждой группы вариационного ряда. Сумма всех частот называется объемом совокупности и определяет число элементов всей совокупности.

Вариационные ряды в зависимости от характера вариации подразделяются на дискретные и интервальные.

Дискретные вариационные ряды. Пусть $a_1,...,a_k$ - возможные значения дискретного признака, при этом $a_1 < a_2 < ... < a_k$. Исследуемая статистическая совокупность содержит статистические единицы. В этой совокупности n_1 единиц имеют значение признака a_1, n_2 единиц - a_2, ... , n_k единиц - a_k. Числа $n_1,...,n_k$ называются абсолютными частотами.

Если N - общее количество единиц совокупности, то

$0 < n_j < n$, $N = \sum_{j=1}^{k} n_j$.

Обычно при статистическом исследовании вместо абсолютных рассматриваются относительные частоты:

$$f_j = \frac{n_j}{N} \cdot 100\%.$$

Для них выполняются соотношения:

$0 \le f_j \le 100\%$, $\sum_{j=1}^{k} f_j = 100\%$.

Рассмотрим пример группировки дискретного ряда.

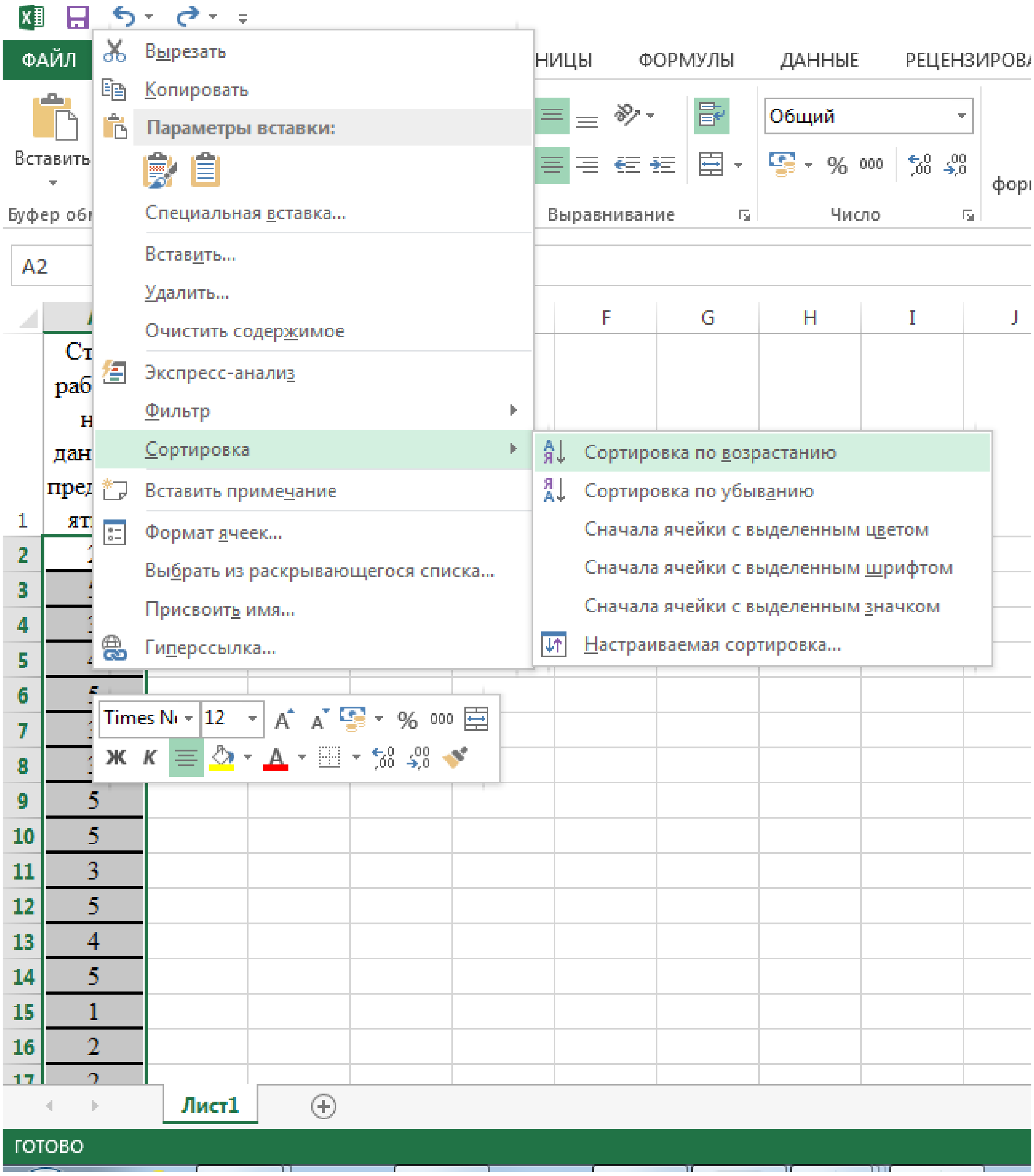
На предприятии работает 25 человек, стаж работы которых представлен в таблице:

Порядковый номер рабочего	Стаж работы на данном предприятии
1	5
2	3
3	4
4	5
5	3
6	3
7	5
8	5
9	2
10	5
11	4
12	5
13	1
14	2
15	2

16	2
17	4
18	1
19	2
20	3
21	2
22	2
23	2
24	3
25	1

Данный ряд является дискретным, число групп определяется числом вариантов: от 1 до 5.

Для удобства определения абсолютных частот воспользуемся функцией MicrosoftExcelСортировка и Счетесли (также можно использовать функцию Гистограмма):



Получим:

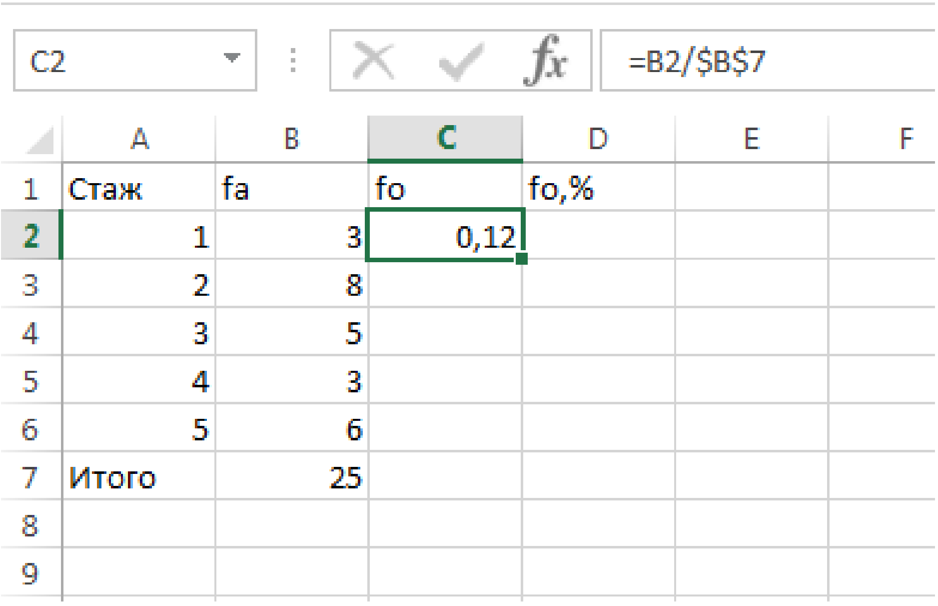
Стаж работы	Число рабочих в каждой группе (абсолютная частота), fa, чел	Отн. частота (частость)fo	Отн. частота, %	Кумул. частота снизу, чел.	Кумул. частота снизу, %	Кумул. частота сверху, чел.	Кумул. частота сверху, %
1	3	0,12	12	3	12	25	100

2	8	0,32	32	11	44	22	88
3	5	0,2	20	16	64	14	56
4	3	0,12	12	19	76	9	36
5	6	0,24	24	25	100	6	24
ИТОГО	25	1	100	-	-	-	-

Абсолютная частота – число наблюдений в каждой группе.

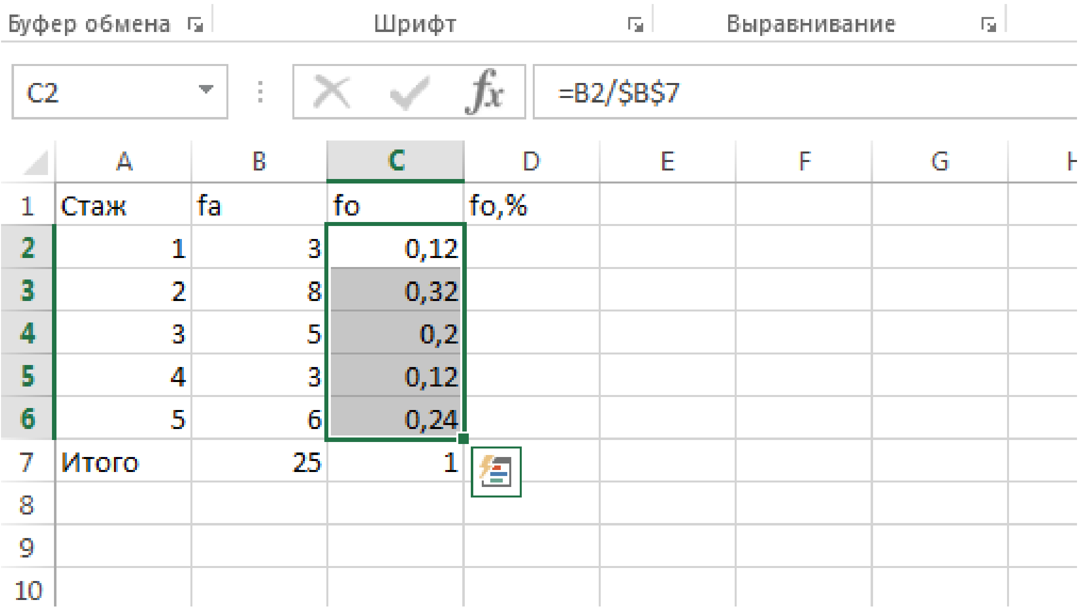
Относительная частота может определяться в долях (частость) и в процентах:

Для нахождения относительных частот применим функцию АВТОЗАПОЛНЕНИЕ:



C2 =B2/$B$7

	A	B	C	D	E	F
1	Стаж	fa	fo	fo,%		
2	1	3	0,12			
3	2	8				
4	3	5				
5	4	3				
6	5	6				
7	Итого	25				
8						
9						



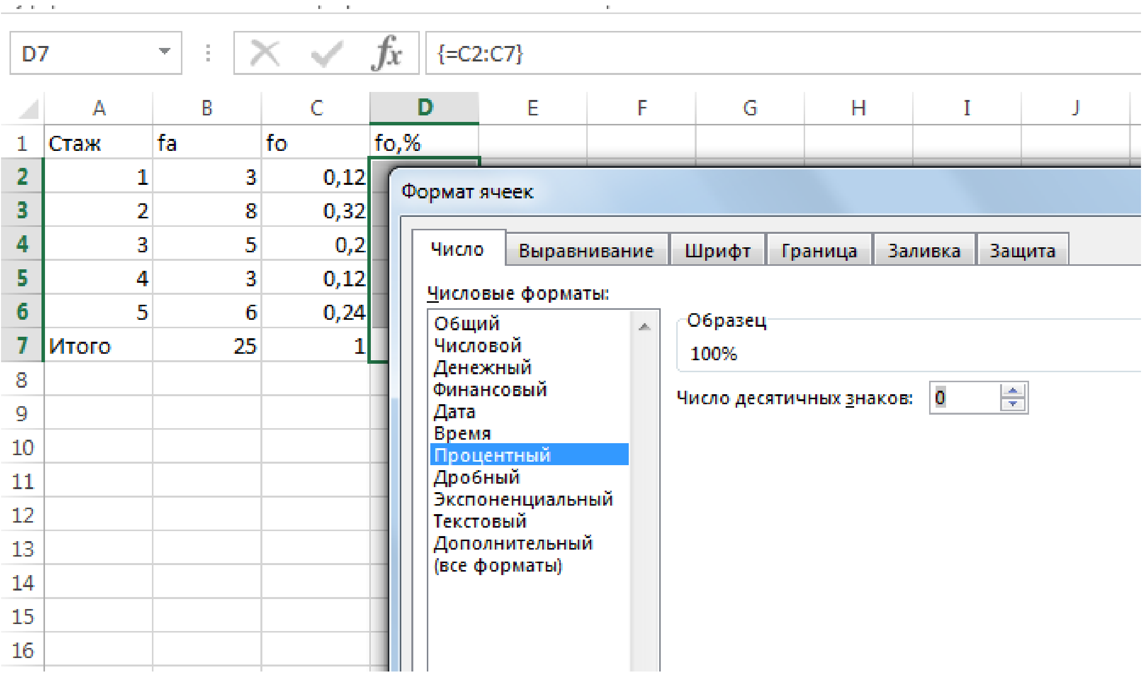
Буфер обмена　Шрифт　Выравнивание

C2 =B2/$B$7

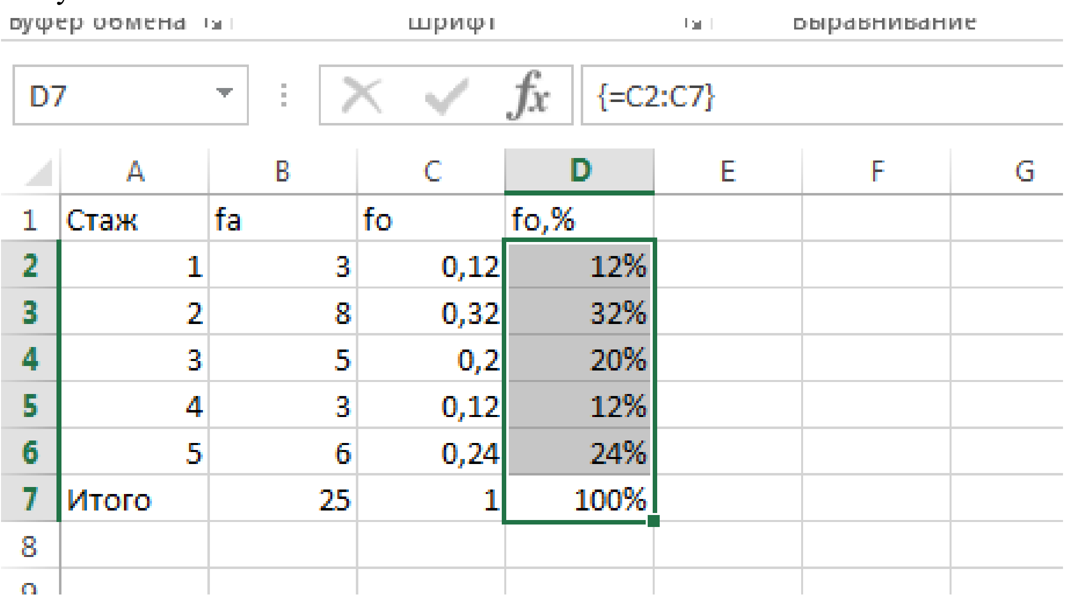
	A	B	C	D	E	F	G
1	Стаж	fa	fo	fo,%			
2	1	3	0,12				
3	2	8	0,32				
4	3	5	0,2				
5	4	3	0,12				
6	5	6	0,24				
7	Итого	25	1				
8							
9							
10							

Естественно сумма относительных частот в долях составит 1.

Для определения относительных частот в процентах можно найденные частоты задать формате «Процентный»:



Получим:



	A	B	C	D
1	Стаж	fa	fo	fo,%
2	1	3	0,12	12%
3	2	8	0,32	32%
4	3	5	0,2	20%
5	4	3	0,12	12%
6	5	6	0,24	24%
7	Итого	25	1	100%

Таким образом, можно сделать вывод, что наибольшее число сотрудников предприятия (32%, т.е. почти треть) имеет стаж работы 2 года.

Кумулятивная (накопленная, интегральная) частота характеризует число наблюдений, накопленное до уровня соответствующей группы, и определяется суммированием абсолютных или относительных частот нарастающим итогом:

Буфер обмена | Шрифт | Выравнивание

B3 | =E2+B3

	A	B	C	D	E	F
1	Стаж	fa	fo	fo,%	fк	
2	1	3	0,12	12%	3	
3	2	8	0,32	32%	=E2+B3	
4	3	5	0,2	20%		
5	4	3	0,12	12%		
6	5	6	0,24	24%		
7	Итого	25	1	100%		
8						

Применив функцию АВТОЗАПОЛНЕНИЕ, получим:

	A	B	C	D	E	F
1	Стаж	fa	fo	fo,%	fк	
2	1	3	0,12	12%	3	
3	2	8	0,32	32%	11	
4	3	5	0,2	20%	16	
5	4	3	0,12	12%	19	
6	5	6	0,24	24%	25	
7	Итого	25	1	100%		
8						

Кумулятивная частота в %:

Буфер обмена | Шрифт | Выравнивание

F2 | =E2/$B$7*100

	A	B	C	D	E	F	G
1	Стаж	fa	fo	fo,%	fк	fк,%	
2	1	3	0,12	12%	3	12	
3	2	8	0,32	32%	11	44	
4	3	5	0,2	20%	16	64	
5	4	3	0,12	12%	19	76	
6	5	6	0,24	24%	25	100	
7	Итого	25	1	100%			
8							

Таким образом можно сделать вывод о том, что больше половины сотрудников (64%) работают на предприятии не более 3-х лет, при этом не более 2 лет 44%.

Также кумулятивная частота может накапливаться в обратном порядке, сверху вниз:

Стаж	fa, чел	fo	fo, %	fk, чел.	fk, %	fk, чел., сверху
1	3	0,12	12	3	12	=B7
2	8	0,32	32	11	44	
3	5	0,2	20	16	64	
4	3	0,12	12	19	76	
5	6	0,24	24	25	100	
ИТОГО	25	1	100	-	-	

Стаж	fa, чел	fo	fo, %	fk, чел.	fk, %	fk, чел., сверху
1	3	0,12	12	3	12	25
2	8	0,32	32	11	44	=G2-B2
3	5	0,2	20	16	64	
4	3	0,12	12	19	76	
5	6	0,24	24	25	100	
ИТОГО	25	1	100	-	-	

Применив к последней ячейке функцию АВТОЗАПОЛНЕНИЕ, получим:

Стаж	fa, чел	fo	fo, %	fk, чел.	fk, %	fk, чел., сверху
1	3	0,12	12	3	12	25
2	8	0,32	32	11	44	22
3	5	0,2	20	16	64	14
4	3	0,12	12	19	76	9
5	6	0,24	24	25	100	6
ИТОГО	25	1	100	-	-	

Аналогично в процентах:

	A	B	C	D	E	F	G	H	I
1	Стаж	fa, чел	fo	fo, %	fk, чел.	fk, %	fk, чел., сверху	fk, % сверху	
2	1	3	0,12	12	3	12	25	100	
3	2	8	0,32	32	11	44	22	=H2-D2	
4	3	5	0,2	20	16	64	14		
5	4	3	0,12	12	19	76	9		
6	5	6	0,24	24	25	100	6		
7	ИТОГ О	25	1	100	-	-			
8									

1	Стаж	fa, чел	fo	fo, %	fk, чел.	fk, %	fk, чел., сверху	fk, % сверху	
2	1	3	0,12	12	3	12	25	100	
3	2	8	0,32	32	11	44	22	88	
4	3	5	0,2	20	16	64	14	56	
5	4	3	0,12	12	19	76	9	36	
6	5	6	0,24	24	25	100	6	24	
7	ИТОГ О	25	1	100	-	-			
8									

С помощью кумулятивной частоты сверху можно получить дополнительную информацию, например в данном примере более половины а именно 56% (14 человек) сотрудников, работают не менее 3 лет.

Интервальные вариационные ряды. Пусть исследуемый признак может принимать значения из непрерывного промежутка $[a,b]$. Произведем группировку данных. Выберем интервалы $(a..a_1),(a_1..a_2),\dots,(a_{n-1},b)$. Обычно заранее решают, к какой именно группе относить границы интервалов $a_1,a_2,\dots,a_{n-1}$ (например, пусть интервал включает свою верхнюю границу, тогда получаем интервалы $[a..a_1],[a_1..a_2],\dots,[a_{n-1},b]$). Подсчитывается число объектов, попавших в каждый из интервалов, и рассчитываются относительные частоты.

Интервалы группировок выделяют из логических соображений. Иногда используются "принцип равных интервалов", о котором говорилось выше, и "принцип равных частот", по которому интервалы выбираются таким образом, чтобы относительные частоты были примерно одинаковы.

Рассмотрим пример группировки интервального ряда.

Возраст сотрудников рассмотренного выше предприятия представлен в таблице:

Порядковый номер рабочего	Возраст (в годах)
1	22
2	34

3	28
4	22
5	40
Б	38
7	32
8	30
9	23
10	25
11	25
12	27
13	23
14	26
15	28
16	29
17	20
18	20
19	22
20	21
21	22
22	22
23	23
24	19
25	25

Здесь для определения числа групп применим формулу Стерджесса:

$$n = 1 + 3{,}322 \cdot \lg N,$$

В данном случае n=1+3.322lg(25)=5.65

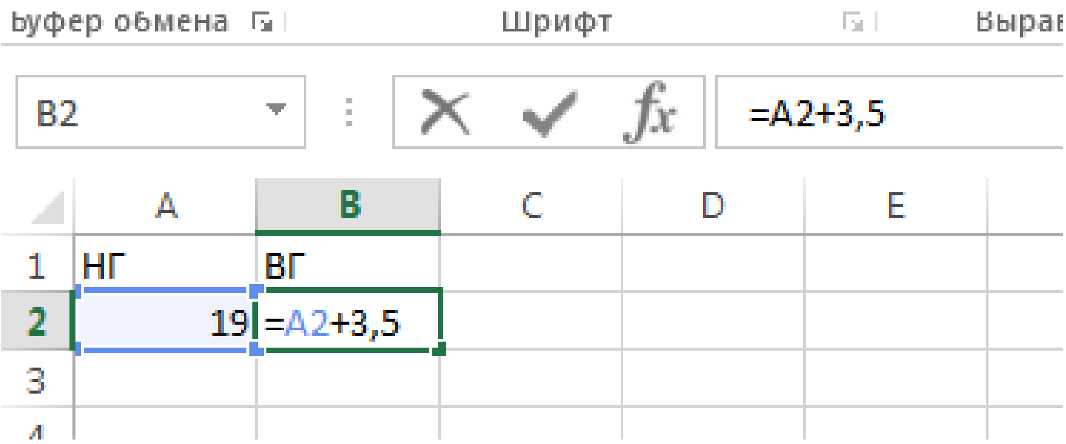
Так как невозможно создать такое число групп, округлим до 6, тогда длина интервалов будет равна:

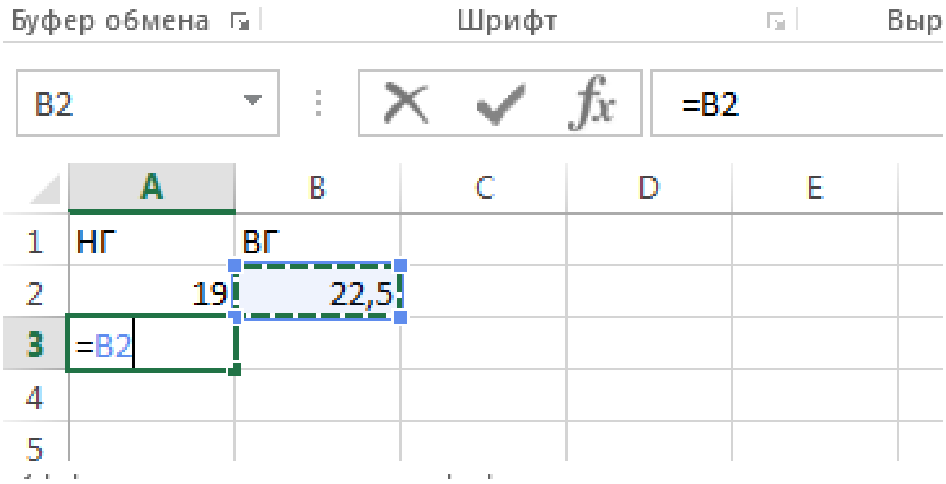
i=(40-19)/6=3,5.

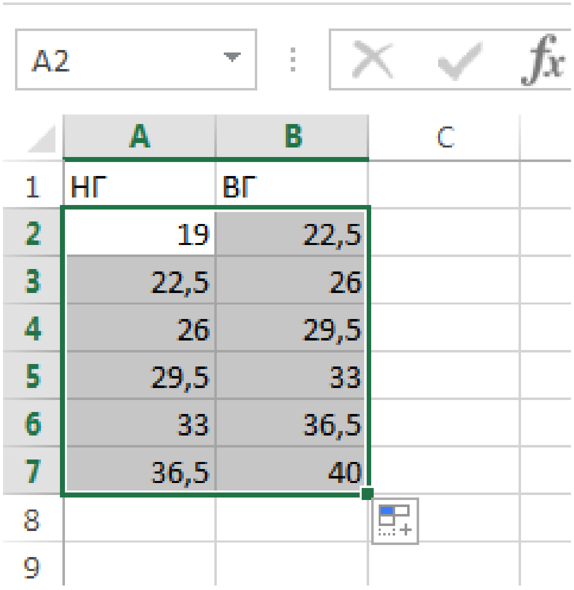
Получим группировку:

Возраст рабочих	Число рабочих в каждой группе (абсолютная частота), fa, чел	Относительная частота (частость), fo	Относительная частота, %	Кумулятивная частота снизу, чел.	Кумулятивная частота снизу, %	Кумулятивная частота сверху, чел.	Кумулятивная частота сверху, %
19-22,5	9	0,36	36%	9	36%	25	100%
22,5-26	7	0,28	28%	16	64%	16	64%
26-29,5	4	0,16	16%	20	80%	9	36%
29,5-33	2	0,08	8%	22	88%	5	20%
33-36,5	1	0,04	4%	23	92%	3	12%
36,5-40	2	0,08	8%	25	100%	2	8%
ИТОГО	25	1	100%	-	-		

Нижние и верхние границы интервалов можно определить, используя функцию АВТОЗАПОЛНЕНИЕ:

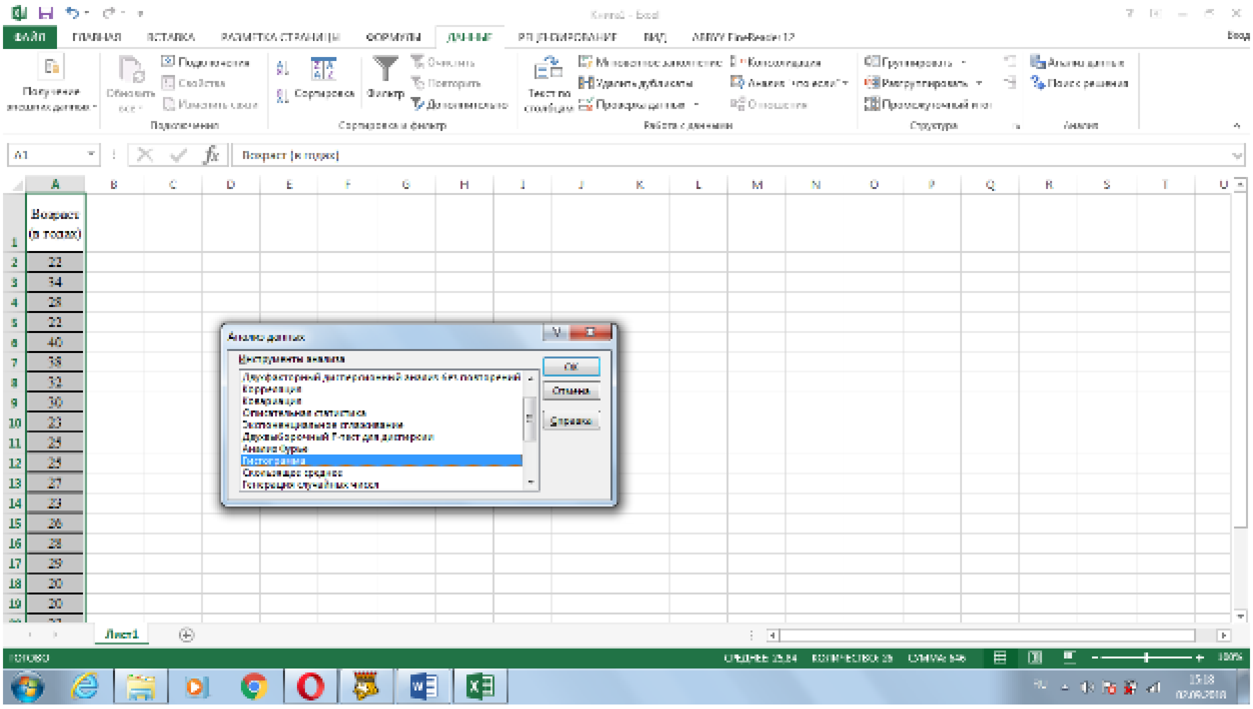




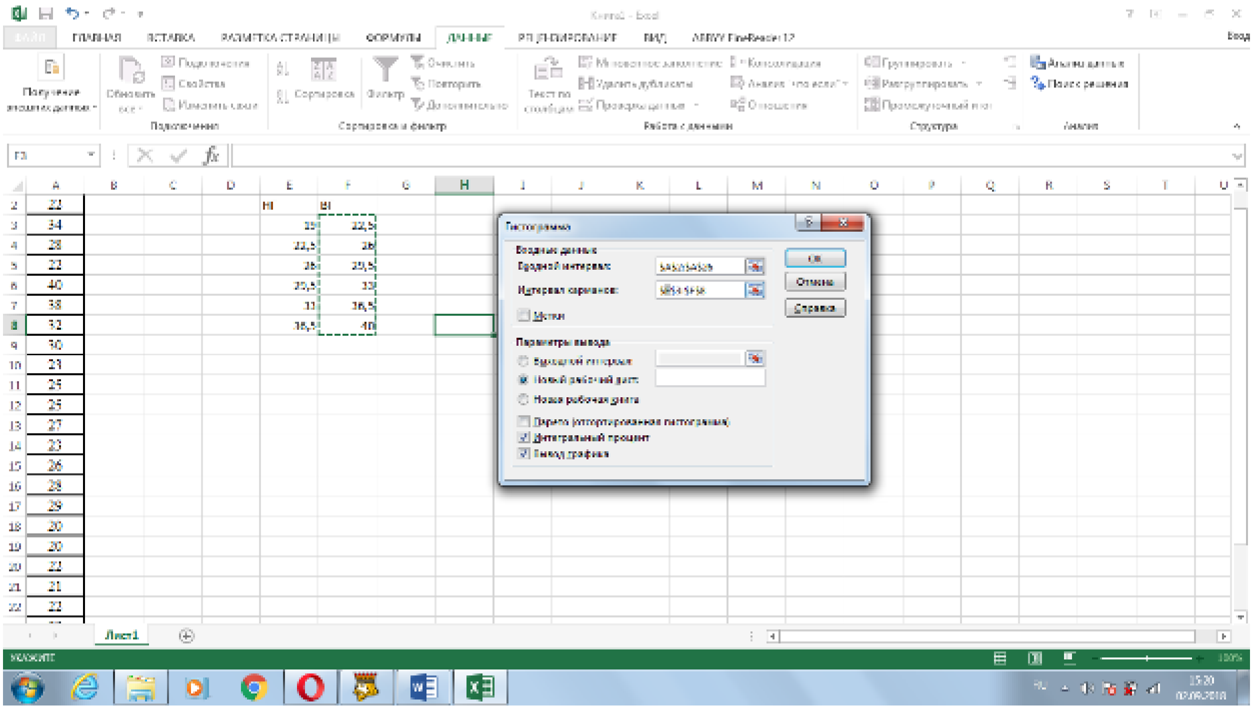


Далее аналогично группировке дискретного ряда. Интерпретация кумулятивной частоты несколько отличается в интервальных группировках: соответствующая частота показывает, сколько наблюдений накоплено до верхней границы данного интервала, например в изучаемой совокупности 80% сотрудников не старше 29,5 лет и только 12% сотрудников старше 33 лет.

Рассмотрим применение функции «Гистограмма»:

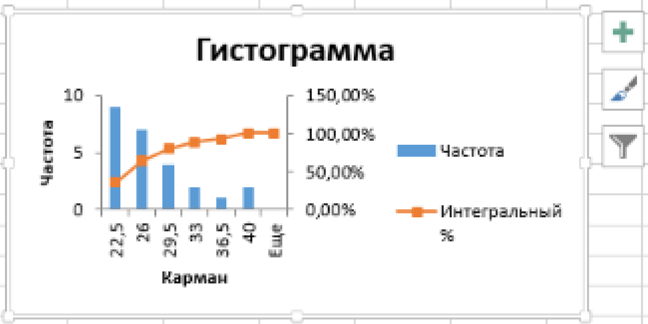


Входной интервал – исходные данные, карманы – верхние границы интервалов (или дискретные значения, если группировка по дискретному признаку):



Получим абсолютную частоту и интегральную частоту в процентах, совмещенные в одном графике:

Карман	Частота	Интегральный %
22,5	9	36,00%
26	7	64,00%
29,5	4	80,00%
33	2	88,00%
36,5	1	92,00%
40	2	100,00%
Еще	0	100,00%



Решить задачи:

Задача 1.1.К каким группировочным признакам – атрибутивным или количественным – относятся:
а) возраст человека;
б) национальность;
в) балл успеваемости;
г) доход сотрудника фирмы;
д) форма собственности?

Задача 1.2.

Какие из указанных ниже группировок являются типологическими:

а) группировка населения по полу;
б) группировка населения по занятости в отраслях народного хозяйства;
в) группировка капитальных сложений в строительство объектов производственного и непроизводственного назначения;
г) группировка предприятий общественного питания по формам собственности.

Задача 1.3. Пользуясь формулой Стерджесса, определить интервал группировки сотрудников фирмы по уровню доходов, если общая численность сотрудников фирмы составляет 200 человек, а минимальный и максимальный доход соответственно равны 5000 р. и 100000 р.

Задача 1.4.В таблице ниже указано количество баллов, набранных студентами группы по дисциплине «Статистика» на контрольной работе к 1-ой текущей аттестации:

18, 16, 19, 17, 18, 20, 15, 13, 5, 15, 16, 8, 9, 20, 4, 11, 12, 18, 15, 9, 13, 4, 5.

1) Сгруппировать данные.

2) Построить:
а) ряд распределения абсолютных частот;
б) ряд распределения относительных частот;
в) ряд распределения интегральных частот.

Задача 1.5.

За отчетный период работа предприятий торговли района характеризуется данными :

Предприятия	Розничный товарооборот, тыс. руб.	Издержки обращения, тыс. руб.
1	511	30,0
2	560	34,0
3	800	46,0
4	465	30,9
5	228	15,9
6	392	25,2
7	640	42,0
8	404	27,0
9	200	16,4
10	425	34,8
11	570	37,0
12	472	28,6
13	250	18,7
14	665	39,0
15	650	36,0
16	620	36,0
17	383	25,0
18	550	38,5
19	750	44,0
20	660	37,0
21	452	27,0
22	563	35,0

1) Построить:
а) ряд распределения абсолютных частот;
б) ряд распределения относительных частот;
в) ряд распределения интегральных частот.

2)Для изучения зависимости между объемом розничного товарооборота и издержками обращения провести группировку предприятий торговли по объему товарооборота. Каждую группу предприятий и совокупность в целом охарактеризовать – числом предприятий, объемом товарооборота, издержками обращения.

Примечание: при выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».
Задача 1.6.

Произведите группировку сравнымиинтервалами по среднегодовой стоимости основных фондов (ОФ). Вкаждойгруппеивцеломпо всем предприятиям подсчитайте:
1) количествопредприятий;
2) среднегодовую стоимость основных фондов;
3) объем товарной продукции за год (ТП);
4) фондоотдачу.

№ п/п	ОФ, млн руб.	ТП, млн руб	№ п/п	ОФ, млн руб.	ТП, млн руб
1	164	369	11	225	399
2	147	134	12	189	354
3	171	194	13	227	630
4	267	377	14	216	453
5	211	223	15	343	661
6	123	91	16	296	1072
7	238	545	17	246	711
8	109	31	18	150	270
9	176	213	19	204	388
10	255	791	20	157	124

Занятие 2. Статистические графики и таблицы.

Практикой выработаны определенные требования к составлению и оформлению таблиц.

1. Таблица по возможности должна быть краткой.

2. Каждая таблица должна иметь подробное название, из которого становится известно:
а) какой круг вопросов излагает и иллюстрирует таблица;
б) каковы географические границы представленной статистической совокупности;
в) за какой период времени, которому они относятся;
г) каковы единицы измерения (если они одинаковы для всех табличных клеток). Если единицы измерения неодинаковы, то в верхних или боковых заголовках обязательно следует указывать, в каких единицах приводятся статистические данные (тонн, штук, рублей и пр.).

3. Таблица может сопровождаться примечаниями, в которых указываются источники данных, более подробно раскрывается содержание показателей, даются и другие пояснения, а также оговорки в случае, если таблица содержит данные, полученные в результате вычислений.

4. При оформлении таблиц обычно применяются такие условные обозначения: знак тире (-) – когда явление отсутствует; х – если явление не имеет осмысленного содержания; многоточие (...) – когда отсутствуют сведения о его размере (или делается запись «Нет сведений»). Если сведения имеются, но числовое значение меньше принятой в таблице точности, оно выражается дробным числом (0,0).

Округленные числа приводятся в таблице с одинаковой степенью точности (до 0,1; до 0,01 и т.п.). Если в таблице приводятся проценты роста, то во многих случаях целесообразно проценты от 300 и более заменять отношениями в разах. Например, писать не «1000 %», а «в 10,0 раз».

Использование графиков для изложения статистических показателей позволяет придать последним наглядность и выразительность, облегчить их восприятие, а во многих случаях помогает уяснить сущность изучаемого явления, его закономерности и особенности, увидеть тенденции его развития, взаимосвязь характеризующих его показателей.

Статистические графики можно классифицировать по разным признакам: назначению (содержанию), способу построения и характеру графического образа.

По *содержанию или назначению* можно выделить графики сравнения в пространстве, графики различных относительных величин (структуры, динамики и т.п.), графики вариационных рядов, графики размещения по территории, графики взаимосвязанных показателей. Возможны и комбинации этих графиков, например графическое изображение вариации в динамике или динамики взаимосвязанных показателей и т.п.

По *способу построения* графики можно разделить на диаграммы, картодиаграммы и картограммы.

По *характеру графического образа* различают графики точечные, линейные, плоскостные (столбиковые, почасовые, квадратные, круговые, секторные, фигурные) и объемные.

Полигон в основном применяют для дискретных рядов. По оси абсцисс откладывают варианты признака, а по оси ординат – частоты или частости.

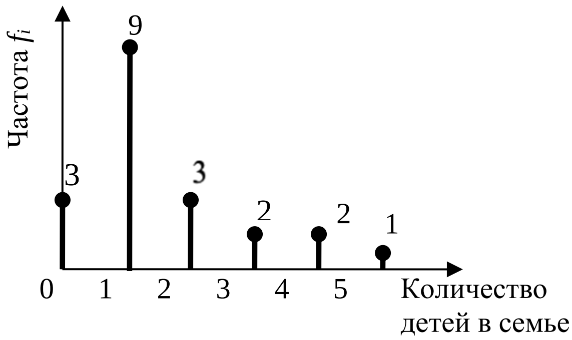


Рис. 2.1. Полигон распределения сотрудников по количеству детей в семье.

Также полигон может выглядеть следующим образом:

Рис. 2.2. Полигон распределения по стажу.

Гистограмма частот (частостей) – это столбиковая диаграмма. Гистограмму применяют для интервальных рядов. Если интервалы равные, то основания столбцов по оси абсцисс – это интервалы изучаемого признака, а высоты столбиков – это частоты (частости).

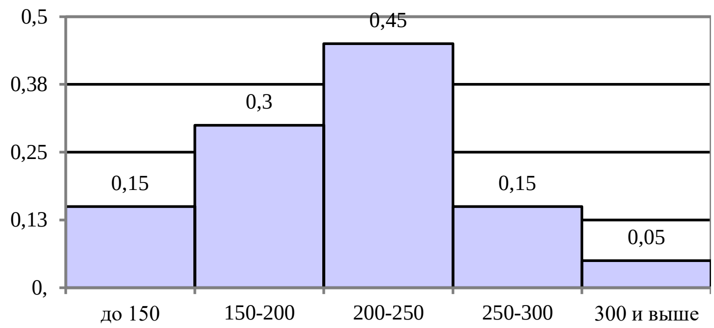


Рис. 2.3. Гистограмма частостей.

Если интервалы неравные то, чтобы площади столбцов равнялись частоте или частости высоту i-го столбца bi рассчитывают по формуле

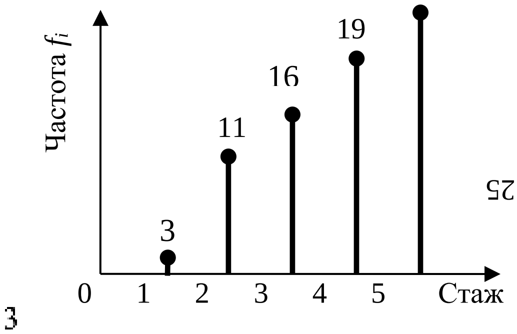
$$b_i = \frac{f_i}{h_i} \quad \text{или} \quad \bar{b}_i = \frac{w_i}{h_i}$$

где $b_i\ (\bar{b}_i)$ – абсолютная (относительная) плотность; fi (wi) – частота (частость) i-ого интервала; hi – величина i-ого интервала.

Кумулята (огива) – это графики кумулятивного ряда снизу (сверху). Кумулятивный ряд – это ряд накопленных частот (частостей). Его получают путем объединения последовательных вариант или групповых интервалов и суммированием соответствующих им частот (частостей).

Кумулята для интервального ряда:

Для дискретного ряда кумуляту целесообразней изобразить следующим образом:



Решить задачи:

Задача 2.1. По данным задачи 1.4 построить гистограммы абсолютных и относительных частот, кумуляту и огиву. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 2.2. По данным задач 1.5 и 1.6 построить гистограммы абсолютных и относительных частот, кумуляту и огиву. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 2.3.Учебные достижения учащихся группы по статистике на 1 октября характеризуются данными, представленными в таблице:

Баллы	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	1	1	2	3	4	4	6	5	3	3	2	1

Построить полигон частот. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 2.4. В таблице представлено распределение личного состава подразделения по воинским званиям:

Звание	Число военнослужащих
Рядовой	25
Ефрейтор	18
Младший сержант	7
Сержант	5
Старший сержант	2

Построить ряд относительных частот. Представить данные графически в виде круговой диаграммы. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 2.5.По результатам тестирования студентов по дисциплине «Статистика» получены данные о доступности заданий теста (отношение числа учащихся, правильно выполнивших задания, к общему числу тестировавшихся). Тест содержал 25 заданий:

Доступность задания, %	25-35	35-45	45-55	55-65	65-75	75-85	85-95
Количество задач	2	1	4	7	6	4	1

1) Построить гистограмму.
2) По данным таблицы составить кумулятивный вариационный ряд, для которого построить кумуляту и огиву.
Примечание. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 2.6. Разработать макет статистической таблицы, характеризующей зависимость успеваемости студентов вашей группы от посещаемости учебных занятий и занятости внеучебной деятельностью.

Задача 2.7. Разработать макет статистической таблицы, характеризующейзависимость спроса на продукт от пола и семейного положения потребителей.

Занятие 3. Степенные средние, выбор формулы расчета

Для того, чтобы средний показатель был действительно типизирующим, он должен рассчитываться с учетом определенных принципов:

1. Средняя должна определяться для совокупностей, состоящих из качественно однородных единиц.
2. Средняя должна исчисляться для совокупности, состоящей из достаточно большого числа единиц.
3. Средняя должна рассчитываться для совокупности, единицы которой находятся в нормальном, естественном состоянии.
4. Средняя должна вычисляться с учетом экономического содержания исследуемого показателя.

Виды степенных средних

Вид степенной средней	Показатель степени (m)	Формула расчета	
		Простая	Взвешенная
Гармоническая	-1	$\overline{X} = \dfrac{n}{\sum \dfrac{1}{x}}$	$\overline{X} = \dfrac{\sum m}{\sum \dfrac{m}{x}}$ $m = xf$
Геометрическая	0	$\overline{X} = \sqrt[n]{\Pi x} = \sqrt[n]{x_1 x_2 \ldots x_n}$	$\overline{X} = \sqrt[\sum f]{\Pi x^f} = \sqrt[\sum f]{x_1^{f_1} x_2^{f_2} \ldots x_n^{f_n}}$
Арифметическая	1	$\overline{X} = \dfrac{\sum x}{n}$	$\overline{X} = \dfrac{\sum xf}{\sum f}$
Квадратическая	2	$\overline{X} = \sqrt{\dfrac{\sum x^2}{n}}$	$\overline{X} = \sqrt{\dfrac{\sum x^2 f}{\sum f}}$
Кубическая	3	$\overline{X} = \sqrt[3]{\dfrac{\sum x^3}{n}}$	$\overline{X} = \sqrt[3]{\dfrac{\sum x^3 f}{\sum f}}$

Простые формулы используются для несгруппированных данных, а взвешенные – для сгруппированных. (f – частоты). В случае, если исходные данные представлены в виде интервального ряда распределения, то в качестве вариантов усредняемого признака (xi) принимают середины интервалов, вычисляемые по каждой группе. Серединное значение интервала может определяться несколькими способами:

1) середина закрытого интервала = полусумма верхней и нижней границ интервала;

2) середина первого (открытого) интервала = середина второго интервала – величина второго интервала;

3) середина последнего (открытого) интервала = середина предпоследнего интервала + величина предпоследнего интервала.

Наиболее часто используется средняя арифметическая. Среднюю гармоническую применяют для расчетов тогда, когда в качестве весов используются не единицы совокупности – носители признака, а произведения этих единиц на значения признака (т.е. m = Xf). К средней гармонической простой следует прибегать в случаях определения, например, средних затрат труда, времени, материалов на единицу продукции, на одну деталь по двум (трем, четырем и т.д.) предприятиям, рабочим, занятым изготовлением одного и того же вида продукции, одной и той же детали, изделия.

Пример 1. Автомобиль от склада до магазина проезжает 20 км со скоростью 40 км/ч, а обратно на склад со скоростью – 60 км/ч. Необходимо рассчитать среднюю скорость автомобиля.

Средняя скорость ($\bar{v}$) равна отношению пройденного пути (s) ко времени (t), затраченному на поездку. Тогда средняя скорость ($\bar{v}$) равна

$$v = \frac{s}{t} = \frac{20+20}{\frac{20}{40}+\frac{20}{60}} = \frac{1+1}{\frac{1}{40}+\frac{1}{60}} = 48\text{км/ч}.$$

В этом случае была использована средняя гармоническая простая.

Пример 2. Автомобиль в течение первого часа едет по трассе со скоростью 40 км/ч, а в течение второго часа – скоростью 60 км/ч. Найти среднюю скорость автомобиля.

$$v = \frac{s}{t} = \frac{40+60}{1+1} = 50\text{ км/ч}$$

В данном случае использована средняя арифметическая простая.

Таким образом, эти два примера еще раз наглядно демонстрируют, что выбор той или иной формы средней зависит от имеющихся исходных данных.

Средняя гармоническая – это превращенная форма средней арифметической, когда частоты fi не заданы (не известны), а известен сложный показатель qi=xi×fi. Тогда fi =qi/xi×и

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i \cdot f_i}{\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} \frac{q_i}{x_i}} = \bar{x}_{гарм}.$$

Средняягармоническаявзвешеннаянаходитболееширокоеприменениевстатистическойпрактикепосравнениюспростой. Использование средней гармонической целесообразно и обосновано для всех показателей интенсивности: цена, скорость, производительность труда, плотность населения и т.п.

Средняя геометрическая обычно применяется в тех случаях, когда варианты ряда представлены относительными показателями динамики. Эта средняя выражает, как правило, средний темп относительного роста или спала.

Пример. Темп роста цен в январе – 105%, в феврале – 98% и в марте – 112%. Найти средний темп роста цен в I квартале.

Используем среднюю геометрическую простую

$$x_{геом} = \sqrt[3]{105\% \cdot 98\% \cdot 112\%} = 104{,}84\%.$$

☛ При выполнении расчетов на калькуляторе более удобно использовать следующий вариант этой формулы

$$\bar{x}_{геом} = \sqrt[3]{1,05 \cdot 0,98 \cdot 1,12} \cdot 100\% = 104,84\%.$$

Главное требование к формуле расчета среднего значения заключается в том, чтобы все этапы расчета имели реальное содержательное обоснование; полученное среднее значение должно заменить индивидуальные значения признака у каждого объекта без нарушения связи индивидуальных и сводных показателей. Иначе говоря, средняя величина должна исчисляться так, чтобы при замене каждого индивидуального значения осредняемого показателя его средней величиной оставался без изменения некоторый итоговый сводный показатель, связанный тем или другим образом с осредняемым. Этот итоговый показатель называется *определяющим,* поскольку характер его взаимосвязи с индивидуальными значениями определяет конкретную формулу расчета средней величины.

Решить задачи:

Задача 3.1. По данным задачи 1.4 определить среднюю арифметическую величину. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel» (воспользоваться функцией «срзнач»). Сравнить с результатом применения средней арифметической взвешенной по сгруппированным данным.

Задача 3.2. По данным задачи 1.5 определить средний товарооборот и средние издержки обращения. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel» (воспользоваться функцией «срзнач»). Сравнить с результатом применения средней арифметической взвешенной по сгруппированным данным.

Задача 3.3.По данным задачи 1.6 определить средние:

- среднегодовую стоимость основных фондов;
- объем товарной продукции за год (ТП);
- фондоотдачу.

Указать, какой вид средней необходимо применять для вычисления и почему.

При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 3.4. Имеются следующие данные 5% выборочного механического обследования студентов одного из вузов о затратах времени на дорогу до института:

Затраты на дорогу до института, час.	До 0,5	0,5 – 1,0	1,0 – 1,5	1,5 – 2,0	Более 2,0
Число студентов в % к итогу	7	18	32	36	7

По этим данным определить средние затраты времени на дорогу, сделать выводы. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 3.5. По данным таблицы:

Предприятие	Численность промышленно-производственного персонала, чел.	Месячный фонд заработной платы, тыс. руб.	Средняя заработная плата, руб.
1	540	564,84	1046
2	275	332,75	1210
3	458	517,54	1130

определить среднюю заработную плату.

Задача 3.6.

Имеются данные о работе малых предприятий за текущий период:

Предприятия	Фактический объем реализации, руб.	Средний объем реализации на 1 работника, руб.	Прибыль в % к объему реализации
1	19000	3800	19
2	16000	4000	20
3	20000	5000	26

Рассчитать по малым предприятиям в целом:

1. Средний объем реализации на 1 работника.
2. Среднюю рентабельность реализованной продукции (прибыль/объем реализованной продукции*100%).

Указать, какой вид средней необходимо применять для вычисления и почему.

Задача 3.7. В отделе заказов торговой фирмы занято трое работников, имеющих 8-часовой рабочий день. Первый работник на оформление одного заказа затрачивает в среднем 14 мин., второй – 15, третий – 19. Определить средние затраты времени в среднем по отделу.

Задача 3.8.

Использование складских помещений города характеризуется следующими данными:

Группы складских помещений, тыс. м^2	Число помещений	Общая занятая площадь, тыс. м^2

До 5 5-10 10-15 15-20 20-25 25-30 30-35 35 и более	3 21 17 9 5 3 4 2	5,2 108,0 163,6 101,2 65,3 40,6 55,4 29,0

Определить средний процент загрузки.При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Занятие 4. Структурные средние и способы их определения

В качестве структурных средних чаще всего используют показатели *моды* – наиболее часто повторяющегося значения признака – и *медианы* – величины признака, которая делит упорядоченную последовательность его значений на две равные по численности части. В итоге у одной половины единиц совокупности значение признака не превышает медианного уровня, а у другой – не меньше его.

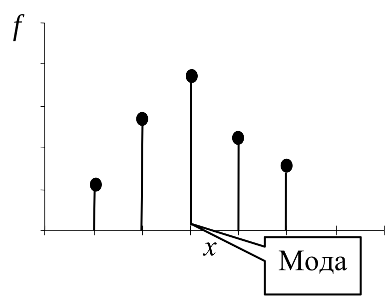
Если же данные о значениях признака X представлены в виде упорядоченных интервалов его изменения (интервальных рядов), расчет моды и медианы осуществляется по специальным формулам.

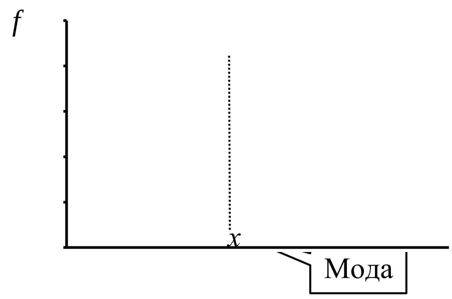
Для интервального ряда с равными интервалами величина моды определяется как

$$\mathbf{Mo} = \mathrm{X}_{\mathrm{Mo}} + \mathbf{h}\frac{\mathbf{m}_{\mathrm{Mo}} - \mathbf{m}_{\mathrm{Mo}-1}}{\left(\mathbf{m}_{\mathrm{Mo}} - \mathbf{m}_{\mathrm{Mo}-1}\right) + \left(\mathbf{m}_{\mathrm{Mo}} - \mathbf{m}_{\mathrm{Mo}+1}\right)},$$

где X_{Mo} – нижнее значение модального интервала;
m_{Mo} – число наблюдений или объем взвешивающего признака в модальном интервале (в абсолютном либо относительном выражении);
m_{Mo-1} – то же для интервала, предшествующего модальному;
m_{Mo+1} – то же для интервала, следующего за модальным;
h – величина интервала изменения признака в группах.

Моду можно определить графически по полигону (рис. 4.1, а) или гистограмме (рис. 4.1, б) распределения.





а) б)

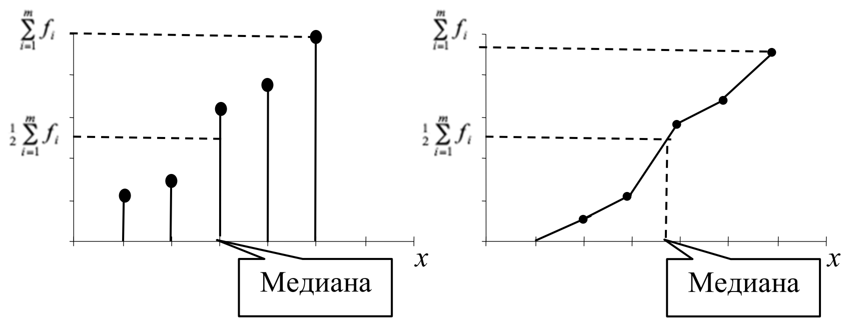
Рис. 4.1. Графическое определение моды по:
а) полигону дискретного ряда; б) гистограмме интервального ряда

Поскольку медианное значение делит всю совокупность на две равные по численности части, оно оказывается в каком-то из интервалов признака X. С помощью интерполяции в этом медианном интервале находят значение медианы:

$$Me = X_{Me} + h_{Me} \cdot \frac{\frac{\sum m}{2} - S_{Me-1}}{m_{Me}},$$

где X_{Me} – нижняя граница медианного интервала;
h_{Me} – его величина;
$(\text{Sum } m)/2$ – половина от общего числа наблюдений или половина объема того показателя, который используется в качестве взвешивающего в формулах расчета средней величины (в абсолютном или относительном выражении);
S_{Me-1} – сумма наблюдений (или объема взвешивающего признака), накопленная до начала медианного интервала;
m_{Me} – число наблюдений или объем взвешивающего признака в медианном интервале (также в абсолютном либо относительном выражении).

Графически медиану можно определить по кумуляте (рис.4.2).



а) б)

Рис. 4.2. Графическое определение медианы по кумуляте:
а) дискретного ряда; б) интервального ряда

Квартили делят ранжированный ряд на четыре части. Различают первый (нижний) квартиль, второй (центральный) квартиль (совпадает с медианой) и третий (верхний) квартиль.

Первый квартиль – это варианта ранжированного ряда, превышающая 1/4 единиц совокупности и меньшая, чем 3/4 единиц совокупности.

Третий квартиль – это варианта ранжированного ряда, превышающая 3/4 единиц совокупности и меньшая, чем 1/4 единиц совокупности.

Для интервального ряда квартили находят по формулам

$$Q_1 = x_{Q_1} + h\frac{\frac{1}{4}\sum_{i=1}^{m} f_i - S_{Q_1-1}}{f_{Q_1}},$$

$$Q_3 = x_{Q_3} + h\frac{\frac{3}{4}\sum_{i=1}^{m} f_i - S_{Q_3-1}}{f_{Q_3}},$$

где Q1, Q3 – первый и третий квартили; xQ1, xQ3 – нижние границы квартильных интервалов; h – величина квартильного интервала; SQ1-1, SQ3-1 – члены кумулятивного ряда, предшествующие квартильному интервалу; fQ1, fQ3 – частоты квартильных интервалов.

Квартили также можно определить по кумуляте (рис.4.3).

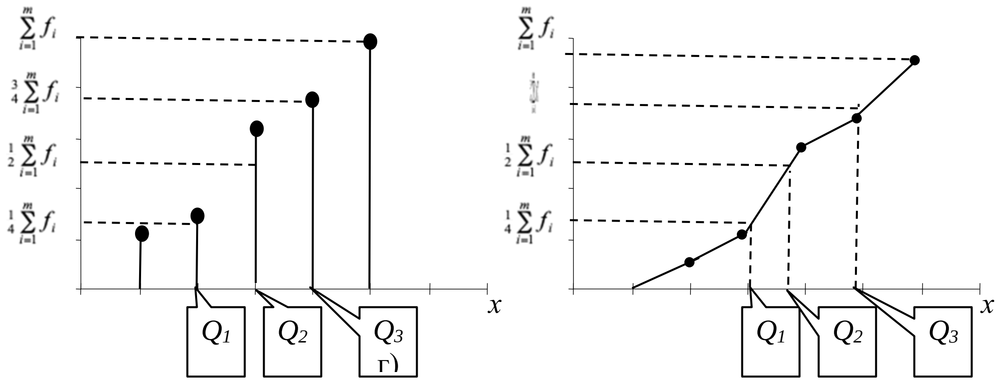


Рис. 4.3. Графическое определение квартилей по кумуляте:
а) дискретного ряда; б) интервального ряда

Децили делят ранжированный ряд на десять равных частей. Всего возможно 9 децилей. Например, первый дециль превышает 1/10 единиц совокупности и меньше, чем 9/10 единиц совокупности.

В случае интервального ряда децили d_j рассчитывают по формуле

$$d_j = x_{d_j} + h\frac{\frac{j}{10}\sum_{i=1}^{m} f_i - S_{d_j-1}}{f_{d_j}}, \quad j=1,\ldots,9,$$

где x_{dj}– нижние границы децильных интервалов; h – величина децильного интервала; S_{d_j-1} – член кумулятивного ряда, предшествующий децильному интервалу; f_{d_j} – частота децильного интервала.

Перцентили (процентиль) делят ранжированный ряд на десять равных частей. Всего возможно 99 перцентилей. Например, седьмой перцентиль превышает 7/100 единиц совокупности и меньше, чем 93/100 единиц совокупности.

Нахождение децилей и перцентилей возможно сделать графически на основе кумуляты по аналогии с медианой и квартилями.

На практике наиболее часто из средних структурных используют моду и медиану.

Решить задачи:

Задача 4.1. Найти моду, медиану и квартили по следующим исходным данным:
В таблице ниже указано количество баллов, набранных студентами группы по дисциплине «Статистика» на контрольной работе к 1-ой текущей аттестации:

18, 16, 19, 17, 18, 20, 15, 13, 5, 15, 16, 8, 9, 20, 4, 11, 12, 18, 15, 9, 13, 4, 5.

При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel» (Надстройка «Анализ данных» - «Описательная статистика» или функции КВАРТИЛЬ, МОДА, МЕДИАНА).

Задача 4.2. Найти моду, медиану и квартилипо следующим исходным данным:

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	1	1	2	3	4	4	6	5	3	3	2	1

Задача 4.3. Найти моду, медиану, квартили и децили по следующим исходным данным:

Доступность задания, %	25-35	35-45	45-55	55-65	65-75	75-85	85-95
Количество задач	2	1	4	7	6	4	1

Задача 4.4.Найти моду, медиану,квартили и децили по следующим исходным данным:

Затраты на дорогу до института, час.	До 0,5	0,5 – 1,0	1,0 – 1,5	1,5 – 2,0	Более 2,0
Число студентов в % к итогу	7	18	32	36	7

Занятие 5. Показатели вариации

Для измерения вариации в статистике применяют несколько способов.

Наиболее простым является расчет показателя *размаха вариации* Н как разницы между максимальным (X_{max}) и минимальным (X_{min}) наблюдаемыми значениями признака:

$$R = X_{max} - X_{min}.$$

Децильный размах

$$D = d_9 - d_1,$$

где d_1иd_9 – первая (нижняя) и девятая (верхняя) децили.

Квартильный размах или интерквартильный разброс (interquartilerange, IQR)

$$IQR = Q_3 - Q_1,$$

где Q_1,Q_3 – первый (нижний) и третий (верхний) квартили. Среди показателей размаха наиболее часто в практическом анализе используют квартильный размах.

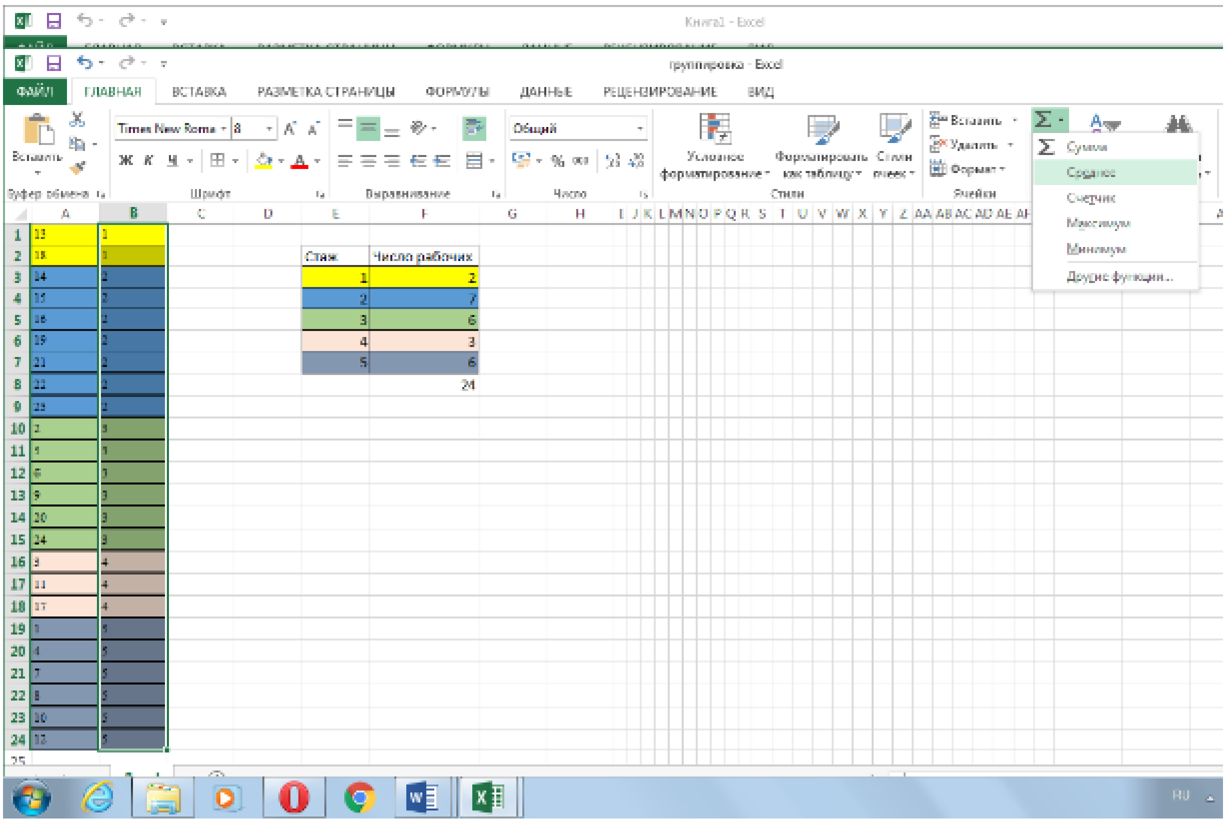
Однако размах вариации показывает лишь крайние значения признака. Повторяемость промежуточных значений здесь не учитывается.

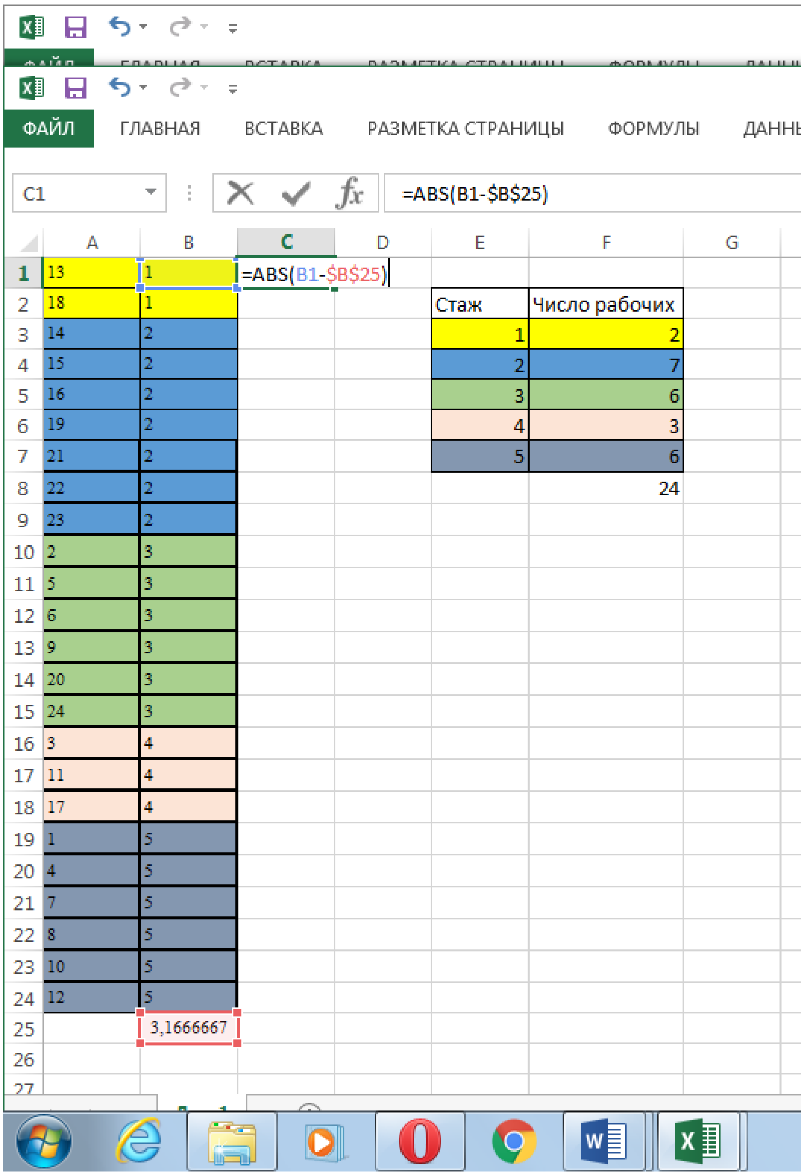
Более строгими характеристиками являются показатели колеблемости относительно среднего уровня признака. Простейший показатель такого типа – *среднее линейное отклонение* Л как среднее арифметическое значение абсолютных отклонений признака от его среднего уровня:

$$l = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

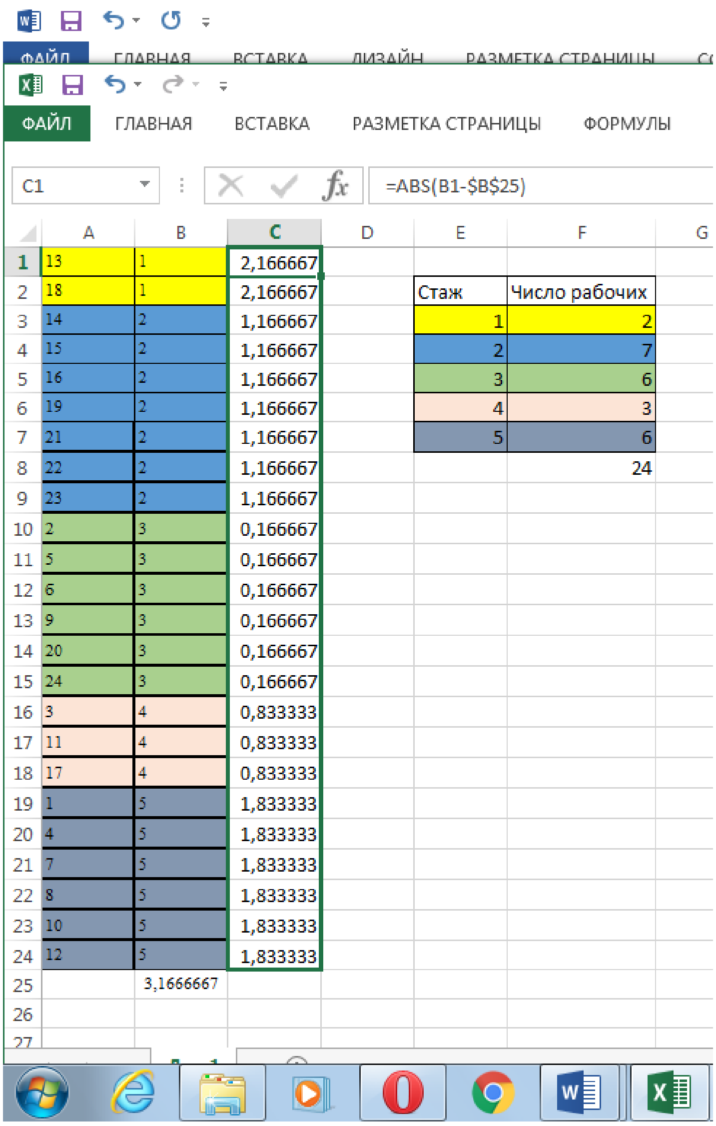
Рассмотрим пример расчета:

Сначала определим среднее значение:

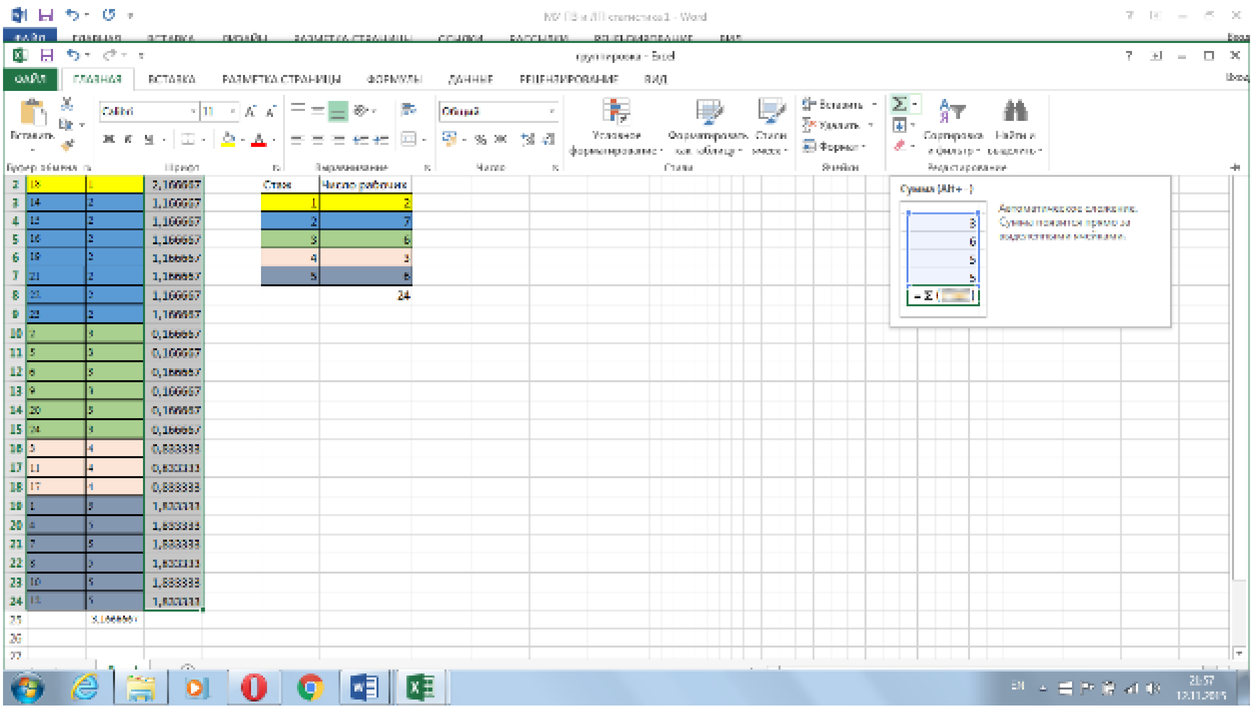




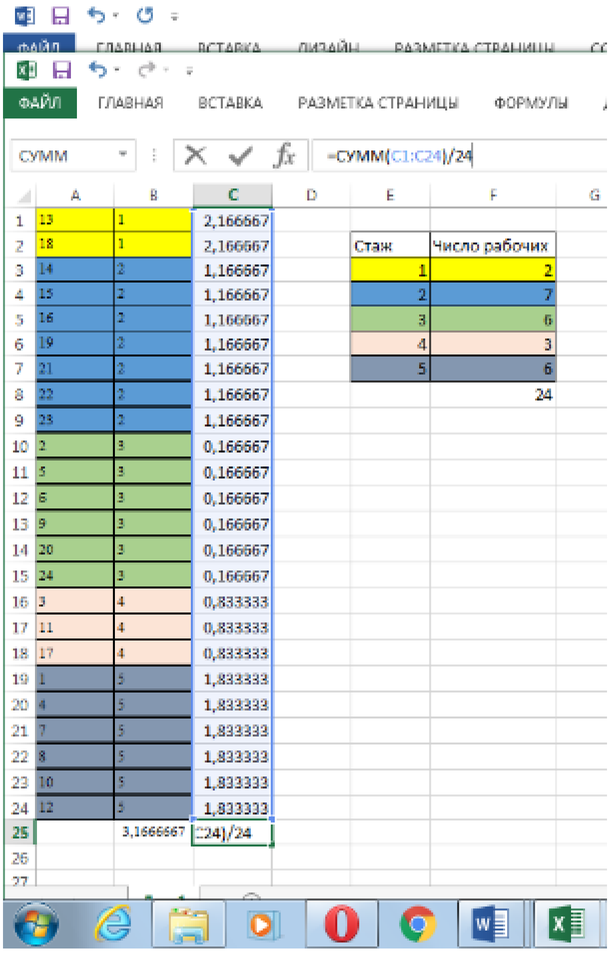
Используем функцию АВТОЗАПОЛНЕНИЕ:


ФАЙЛ
ГЛАВНАЯ
ВСТАВКА
РАЗМЕТКА СТРАНИЦЫ
ФОРМУЛЫ
C1
=ABS(B1-$B$25)
A B C D E F G
1 13 1 2,166667
2 18 1 2,166667 Стаж Число рабочих
3 14 2 1,166667 1 2
4 15 2 1,166667 2 7
5 16 2 1,166667 3 6
6 19 2 1,166667 4 3
7 21 2 1,166667 5 6
8 22 2 1,166667 24
9 23 2 1,166667
10 2 3 0,166667
11 5 3 0,166667
12 6 3 0,166667
13 9 3 0,166667
14 20 3 0,166667
15 24 3 0,166667
16 3 4 0,833333
17 11 4 0,833333
18 17 4 0,833333
19 1 5 1,833333
20 4 5 1,833333
21 7 5 1,833333
22 8 5 1,833333
23 10 5 1,833333
24 12 5 1,833333
25 3,1666667
26
27

и автосумма:



Осталось поделить на количество элементов в ряду:



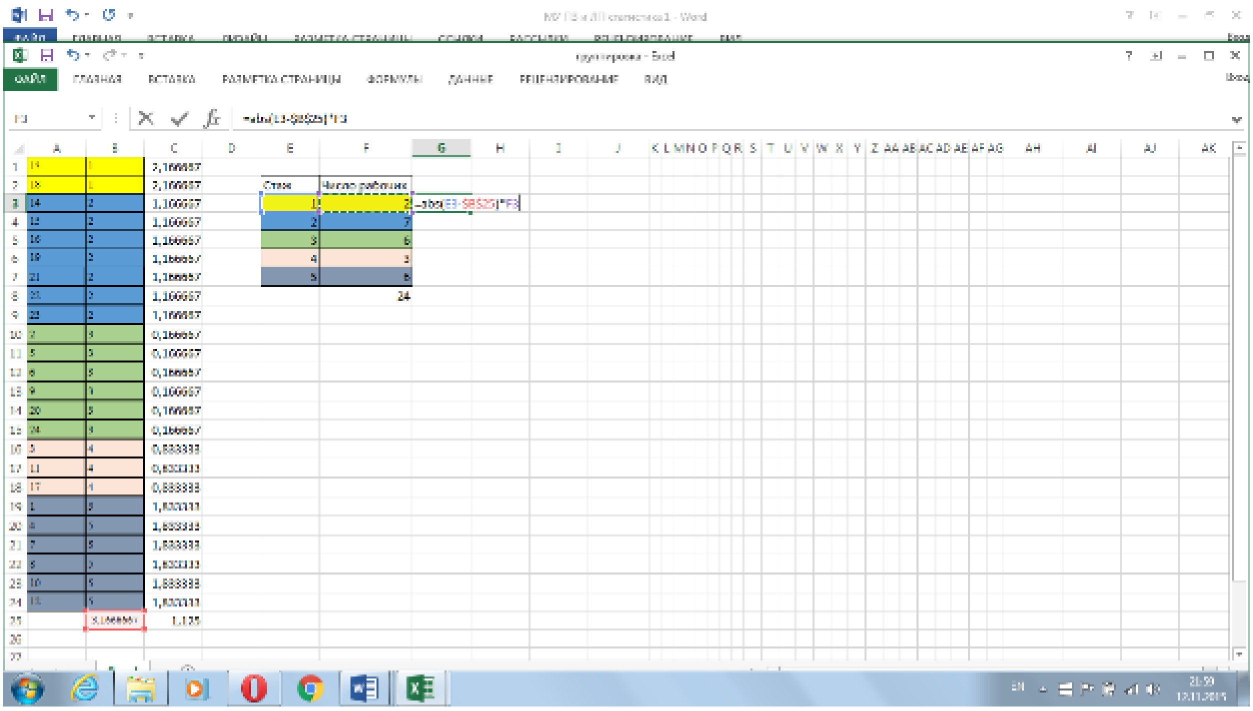
Получили среднее линейное отклонение 1,125. Также можно использовать функцию Excel СРОТКЛ.

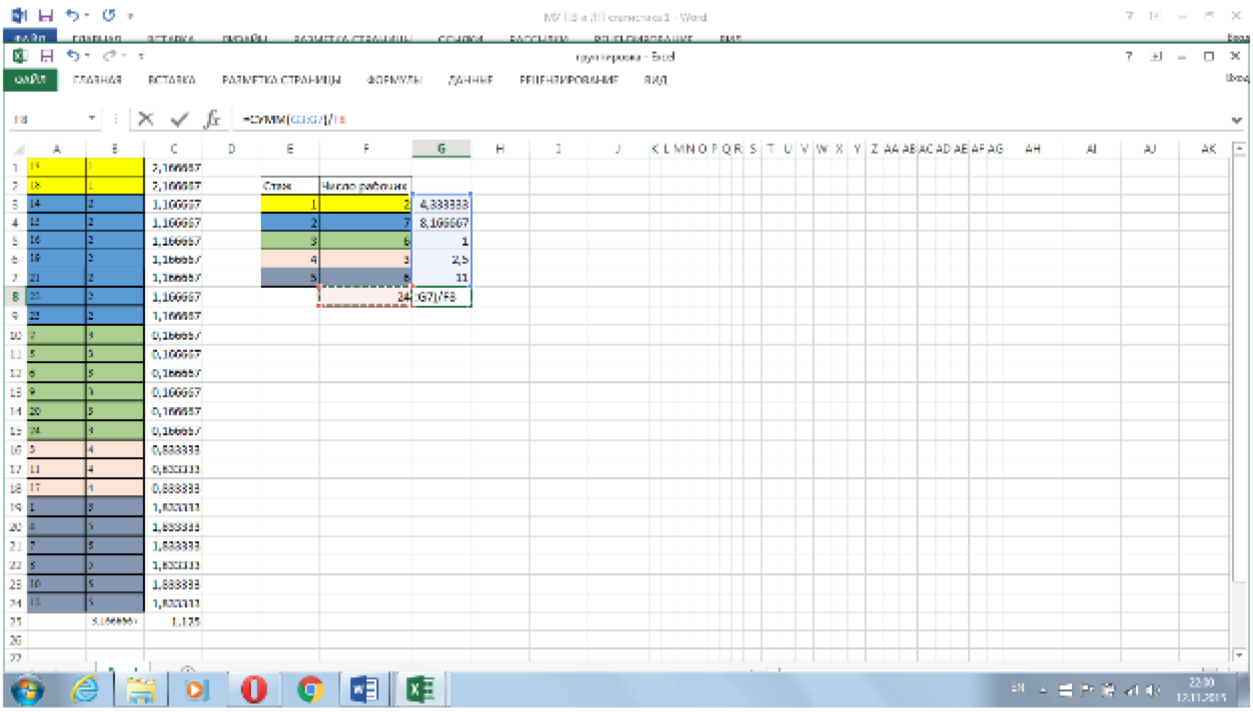
При повторяемости отдельных значений X используют формулу средней арифметической взвешенной:

$$I = \frac{\sum_{i=1}^{m} |x_i - \bar{x}| f_i}{\sum_{i=1}^{m} f_i}$$

(Напомним, что алгебраическая сумма отклонений от среднего уровня равна нулю.)

Рассмотрим тот же пример, но сгруппированные данные:





Значение среднего линейного отклонения также равно 1,125.

В случае расчета по интервальной группировке используются середины интервалов, и расчет по сгруппированным и несгруппированным данным будет давать некоторое расхождение в результатах.

Показатель среднего линейного отклонения нашел широкое применение на практике. С его помощью анализируются, например, состав работающих, ритмичность производства, равномерность поставок материалов, разрабатываются системы материального стимулирования. Но, к сожалению, этот показатель усложняет расчеты вероятностного типа, затрудняет применение методов математической статистики. Поэтому в статистических научных исследованиях для измерения вариации чаще всего применяют показатель *дисперсии.*

Дисперсия признака определяется на основе квадратической степенной средней:

- **простая** для несгруппированных данных

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

- **взвешенная** для сгруппированных данных

$$\sigma^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^{m} f_i}.$$

Можно отметить следующий недостаток этого показателя вариации – если варианты xi имеют некоторую размерность (метр, рубль, килограмм и т.д.), то дисперсия имеет размерность в квадрате, что затрудняет ее интерпретацию (например, если средняя зарплата составляет 18 тысяч рублей, то соответствующая дисперсия может составить 500 тысяч рублей в квадрате, что лишено экономического смысла).

Этого недостатка лишено **среднее квадратическое отклонение**

- **простое** для несгруппированных данных

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

- **взвешенное** для сгруппированных данных

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^{m} f_i}}$$

Достоинством этого показателя вариации является то, что он выражается в тех же единицах, что и варианты, поэтому экономически хорошо интерпретируется

В общей теории статистики показатель дисперсии является оценкой одноименного показателя теории вероятностей и (как сумма квадратов отклонений) оценкой дисперсии в математической статистике, что позволяет использовать положения этих теоретических дисциплин для анализа социально-экономических процессов.

Если вариация оценивается по небольшому числу наблюдений, взятых из неограниченной генеральной совокупности, то и среднее значение признака определяется с некоторой погрешностью. Расчетная величина дисперсии оказывается смещенной в сторону уменьшения. Для получения несмещенной оценки выборочную дисперсию, полученную по приведенным ранее формулам, надо умножить на величину n / (n - 1). В итоге при малом числе наблюдений (< 30) дисперсию признака рекомендуется вычислять по формуле

$$\sigma^2 = \frac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1} \quad \text{или} \quad \sigma^2 = \frac{n}{n-1}\left(\overline{X^2} - \left(\overline{X}\right)^2\right).$$

Обычно уже при n > (15÷20) расхождение смещенной и несмещенной оценок становится несущественным. По этой же причине обычно не учитывают смещенность и в формуле сложения дисперсий.

Если из генеральной совокупности сделать несколько выборок и каждый раз при этом определять среднее значение признака, то возникает задача оценки колеблемости средних. Оценить дисперсию *среднего значения* можно и на основе всего одного выборочного наблюдения по формуле

$$\sigma^2\left(\overline{X}\right) = \sigma^2 \;/\; n,$$

где n – объем выборки; s^2 – дисперсия признака, рассчитанная по данным выборки.

Величина $\mu = \sqrt{\sigma^2\left(X\right)} = \sqrt{\sigma^2 \;/\; n}$ носит название *средней ошибки выборки* и является характеристикой отклонения выборочного среднего значения признака X от его истинной средней величины. Показатель средней ошибки используется при оценке достоверности результатов выборочного наблюдения.

Вышеперечисленные показатели (кроме среднего линейного отклонения) можно определить по несгруппированным данным с помощью приложения «Описательная статистика» (см. выше).

Показатели относительного рассеивания. Для характеристики меры колеблемости изучаемого признака исчисляются показатели колеблемости в относительных величинах. Расчет показателей меры относительного рассеивания осуществляют как отношение абсолютного показателя рассеивания к средней арифметической, умножаемое на 100%.

Относительный размах (коэффициент осцилляции)

$$\nu_R = \frac{R}{\bar{x}}100\%.$$

Относительное квартильное расстояние

$$\nu_q = \frac{IQR}{\bar{x}}\cdot 100\%.$$

Относительное линейное отклонение

$$\nu_l = \frac{l}{\bar{x}}100\%.$$

Коэффициент вариации

$$\nu = \frac{\sigma}{\bar{x}}100\%.$$

Последний является наиболее распространенным показателем колеблемости, используемым для оценки типичности средних величин.В статистике совокупности, имеющие коэффициент вариации больше 30–35 %, принято считать неоднородными.

Вариация значений признака обусловлена как воздействием случайных факторов (случайная вариация), так и воздействием неслучайных факторов (систематическая вариация). Изучение вариации позволяет вскрыть сущность изучаемого явления – выявить каковы существенные факторы и оценить степень их влияния.

Для оценки влияние отдельных факторов на вариацию осуществляют группировку, разбивая изучаемую совокупность на группы, однородные по изучаемому признаку. Изучение вариации проводят путем исчисления и анализа следующих видов дисперсий: общей, межгрупповой и внутригрупповой.

Общая дисперсия измеряет вариацию признака, обусловленную влиянием всех факторов (случайных и неслучайных) на данную совокупность. Может быть рассчитана по формуле простой или взвешеннойдисперсии.

Межгрупповая дисперсия измеряет систематическую вариацию, т.е. оценивает влияние признака-фактора, положенного в основание группировки, на вариацию изучаемого (результативного) признака.

$$\delta^2 = \frac{\sum_{j=1}^{m}(\bar{x}_j - \bar{x})^2 f_j}{\sum_{j=1}^{m} f_j}$$

где $\bar{x}_j$ – групповая (частная) средняя *j*-й группы; $\bar{x}$ – общая средняя всей совокупности; f_j – частота *j*-й группы.

Внутригрупповые (частные) дисперсии σ_j^2 отражают случайную вариацию, т.е. часть вариации, обусловленную влиянием других неучтенных факторов. Внутригрупповая дисперсия *j*-й группы σ_j^2 вычисляется на основе отклонений отдельных значений признака внутри *j*-й группы от средней арифметической этой группы. В зависимости от имеющихся данных может использоваться формула простой или взвешенной дисперсии.

Средняя из внутригрупповых дисперсий $\overline{\sigma^2}$ – это средняя арифметическая взвешенная из внутригрупповых дисперсий.

$$\overline{\sigma^2} = \frac{\sum_{j=1}^{m}\sigma_j^2 \cdot f_j}{\sum_{j=1}^{m} f_j}$$

Согласно **правилу сложения дисперсий** общая дисперсия равна сумме межгрупповой дисперсии и средней из внутригрупповых дисперсий

$$\sigma^2 = \delta^2 + \overline{\sigma^2}$$

Правило сложения дисперсий позволяет оценить степень влияния группировочного признака-фактора на изучаемый результативный показатель. Для оценки тесноты связи этих факторов служат **коэффициент детерминации** и **эмпирическое (выборочное) корреляционное отношение**.

Коэффициент детерминации

$$\eta^2 = \frac{\delta^2}{\sigma^2}$$

Эмпирическое (выборочное) корреляционное отношение

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$$

Коэффициент детерминации h2 и эмпирическое корреляционное отношение h принимают значения в диапазоне от 0 до 1. При отсутствии влияния группировочного признака-фактора на вариацию результативного показателя эти показатели равны нулю. Чем ближе значения показателя к единице, тем сильнее связь.

Для качественной оценки тесноты связи на основе показателя эмпирического корреляционного отношения h использовать шкалу Чэддока:

h	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Теснота связи	Слабая	Умеренная	Заметная	Тесная	Весьма тесная

Решить задачи:

Задача 5.1. В таблице ниже указано количество баллов, набранных студентами группы по дисциплине «Статистика» на контрольной работе к 1-ой текущей аттестации:

18, 16, 19, 17, 18, 20, 15, 13, 5, 15, 16, 8, 9, 20, 4, 11, 12, 18, 15, 9, 13, 4, 5.

Рассчитать показатели вариации. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 4.2. Рассчитать показатели вариации по следующим исходным данным:

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	1	1	2	3	4	4	6	5	3	3	2	1

При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 4.3.Рассчитать показатели вариации по следующим исходным данным:

Доступность задания, %	25-35	35-45	45-55	55-65	65-75	75-85	85-95
Количество задач	2	1	4	7	6	4	1

При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 4.4.Рассчитать показатели вариации по следующим исходным данным:

Затраты на дорогу до института, час.	До 0,5	0,5 – 1,0	1,0 – 1,5	1,5 – 2,0	Более 2,0
Число студентов в % к итогу	7	18	32	36	7

При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Занятие 6. Динамические ряды.

Ряд динамики (динамический ряд, временной ряд) – это статистические данные, отображающие развитие изучаемого явления во времени.

В каждом ряду динамики имеются два основных элемента:

1. Показатели времени *t*. В качестве показателей времени в рядах динамики выступают либо определенные даты (моменты) времени, либо отдельные периоды (годы, кварталы, месяцы, сутки).

2. Соответствующие им уровни развития изучаемого явления y_t. Они могут выражаться абсолютными, относительными и средними величинами.

В зависимости от характера изучаемого явления уровни рядов динамики могут относиться или к определенным датам (моментам) времени, либо к отдельным периодам (интервалам) времени. В соответствии с этим выделяют:

- **моментные**ряды динамики, которые отображают состояние изучаемых явлений на определенные даты (моменты) времени (например, число работающих на 20 ноября 2018 г., товарные запасы на 01.04.2019 г., величина банковских депозитов на 15.07.2019 г. и т.д.);
- **интервальные** ряды динамики отображают итоги развития явления за определенные периоды (интервалы) времени. Особенность интервальных рядов – каждый уровень ряда складывается из данных за более короткие периоды времени. Примеры интервальных рядов: объем товарооборота по месяцам (кварталам, годам); суммы выплаченной заработной платы по месяцам и т.д.
- Для проведения статистического анализа ряда динамики исчисляют систему показателей, сравнивая уровни ряда между собой.
- В зависимости от выбора базы сравнения уровней ряда динамики различают две системы расчета показателей:
- Базисная система, при которой каждый уровень ряда динамики yi сравнивается с уровнем, принятым за постоянную базу сравнения (за базу сравнения обычно принимают первый уровень y1).
- Цепная система, при которой каждый уровень ряда динамики yi сравнивается с его предыдущим уровнем yi-1.

Расчет показателей динамики представлен в следующей таблице:

Показатель	**Базисный**	**Цепной**
Абсолютный прирост $(\Delta_{i_{баз}}; \Delta_{i_{цеп}})_*$	Y_i-Y_0	Y_i-Y_{i-1}

Коэффициент роста ($К_р$)	$Y_i : Y_0$	$Y_i : Y_{i-1}$
Темп роста ($Т_р$)	$(Y_i : Y_0)\times 100$	$(Y_i : Y_{i-1})\times 100$
Коэффициент прироста ($К_{пр}$)**	$K_p - 1; \quad \frac{Y_i - Y_0}{Y_0};$ $\Delta_{баз}/Y_0$	$K_p - 1; \quad \frac{Y_i - Y_{i-1}}{Y_{i-1}};$ $\Delta_{цеп}/Y_{i-1}$
Темп прироста ($Т_{пр}$)	$K_{пр}\cdot 100; \quad T_p - 100$	$K_{пр}\cdot 100; \quad T_p - 100$
Абсолютное значение одного процента прироста (А)	$Y_0 \div 100$	$Y_{i-1}/100; \ \Delta/T_{пр};$ $\frac{Y_i - Y_{i-1}}{T_p - 100}$

* $\Delta_{i_{баз}} = \sum \Delta_{i_{баз}}$.

** $K_p^{баз} = \prod_{i=1} K_p^{цеп}$.

В случае, когда сравнение проводится с периодом (моментом) времени, начальным в ряду динамики, получают **базисные показатели.** Если же сравнение производится с предыдущим периодом или моментом времени, то говорят о **цепных показателях.**

Рассмотрим пример. Имеются данные об объемах и динамике продаж акций на 15 крупнейших биржах России за пять месяцев.

Показатель	Март	Апрель	Май	Июнь	Июль	Август
Объем продаж, млн. руб.	709,98	1602,61	651,83	220,80	327,68	277,12
Абсолютный прирост:						
цепной,	-	892,63	-950,78	-431,03	106,88	-50,56
базисный	-	892,63	-58,15	-489,18	-382,3	-432,86
Коэффицент (индекс) роста цепной	-	2,257	0,407	0,339	1,484	0,846
Темп роста, %:						
цепной,	-	225,7	40,7	33,9	148,4	84,6
базисный	100	225,7	91,8	31,1	46,2	39,0
Темп прироста						
цепной, %	-	125,7	-59,3	-66,1	48,4	-15,4
базисный, %	-	125,7	-8,2	-68,9	-53,8	61,0
Абсолютное значение 1% прироста (цепной)	-	7,10	16,03	6,52	2,21	3,28

Система средних показателей динамики включает:
средний уровень ряда,
средний абсолютный прирост,
средний темп роста,
средний темп прироста.

Средний уровень интервального ряда динамики вычисляют по формуле средней арифметической:

- если **интервалы равные**, то применяется **средняя арифметическая простая**:

$$\bar{y} = \frac{y_1 + y_2 + \boxed{?} + y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$$

где y_i – уровни ряда; n – число уровней ряда;

- если **интервалы неравные**, то применяется **средняя арифметическая взвешенная**:

$$\bar{y} = \frac{y_1 \cdot t_1 + y_2 \cdot t_2 + \boxed{?} + y_n \cdot t_n}{t_1 + t_2 + \boxed{?} + t_n} = \frac{\sum_{i=1}^{n} y_i \cdot t_i}{\sum_{i=1}^{n} t_i}$$

где y_i – уровни ряда, сохраняющиеся без изменения в течение промежутка времени t_i.

Средний уровень моментного ряда динамики вычисляют с помощью средней хронологической:

- если ряд динамики с **равноотстоящими уровнями**, то применяется **средняя хронологическая простая**:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \boxed{?} + \frac{y_{n-1} + y_n}{2}}{n-1} = \frac{\frac{y_1}{2} + y_2 + \boxed{+} y_{n-1} + \frac{y_n}{2}}{n-1};$$

- если ряд динамики с **неравноотстоящими уровнями**, то применяется **средняя хронологическая простаявзвешенная**:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} \cdot t_1 + \frac{y_2 + y_3}{2} \cdot t_2 + \boxed{?} + \frac{y_{n-1} + y_n}{2} \cdot t_{n-1}}{t_1 + t_2 + \boxed{?} + t_{n-1}} =$$

$$= \frac{\sum_{i=1}^{n-1} \frac{y_i + y_{i+1}}{2} \cdot t_i}{\sum_{i=1}^{n-1} t_i} = \frac{\sum_{i=1}^{n-1} (y_i + y_{i+1}) \cdot t_i}{2 \sum_{i=1}^{n-1} t_i}.$$

Средний абсолютный прирост рассчитывается по формуле:

$$\overline{\Delta y_{n/1}} = \frac{y_n - y_1}{n-1} = \frac{\Delta y_{n/1}}{n-1} = \frac{\Delta y_{2/1} + \Delta y_{3/2} + \boxed{?} + \Delta y_{n/n-1}}{n-1} = \frac{\sum_{k=1}^{n-1} \Delta y_{k+1/k}}{n-1},$$,

где

$\Delta y_{n/1}$ – абсолютный базисный прирост;

$\Delta y_{k+1/k}$ – абсолютные цепные приросты.

Для определения среднего абсолютного прироста через абсолютные цепные приросты $\Delta y_{k+1/k}$ применяется формула средней арифметической простой.

Средний коэффициент роста вычисляется по формуле

$$\overline{K}_{p\,n/1} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{K_{p\,n/1}} =$$

$$= \sqrt[n-1]{K_{p\,2/1} \cdot K_{p\,3/2} \cdot \boxed{?} \cdot K_{p\,n/n-1}} = \sqrt[n-1]{\prod_{k=1}^{n-1} K_{p\,k+1/k}},$$

где $K_{p\,n/1}$ – базисный коэффициент роста; $K_{p\,k+1/k}$ – цепные коэффициенты роста.

Для определения среднего коэффициента роста через цепные коэффициенты роста $K_{p\,k+1/k}$ применяется формула средней геометрической простой.

Средний темп роста:

$$\overline{\text{Т}_\text{р}} = \overline{\text{К}_\text{р}} \cdot 100,$$

Средний темп прироста (%) определяется по единственной методологии:

$$\overline{\text{Т}_\text{пр}} = \overline{\text{Т}_\text{р}} - 100.$$

Рассмотрим пример.

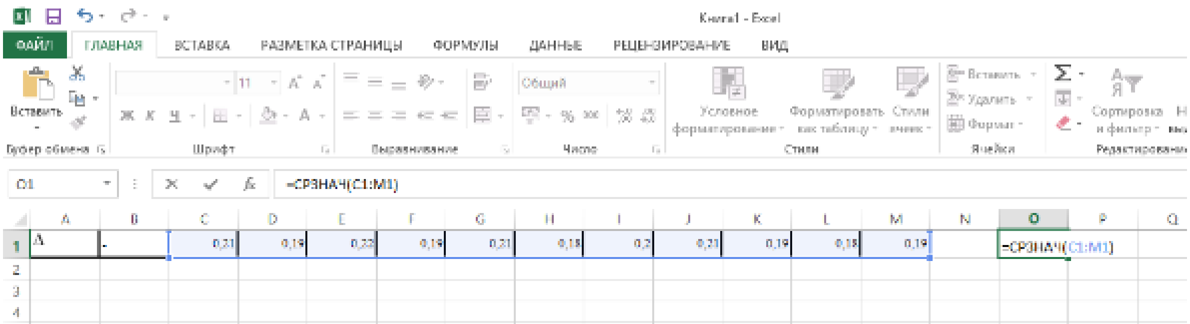
Товарооборот предприятия характеризуется следующими показателями (млн. руб.):

Неделя	1	2	3	4	5	6	7	8	9	10	11	12
ТО, млн. руб.	2,01	2,22	2,41	2,625	2,819	3,03	3,211	3,41	3,62	3,81	3,99	4,18
Базисные показатели												
Δ	-	0,21	0,4	0,615	0,809	1,02	1,201	1,4	1,61	1,8	1,98	2,17
K_p	-	1,10	1,20	1,31	1,40	1,51	1,60	1,70	1,80	1,90	1,99	2,08
T_p	-	110,4%	119,9%	130,6%	140,2%	150,7%	159,8%	169,7%	180,1%	189,6%	198,5%	208,0%
$K_{пр}$	-	0,10	0,20	0,31	0,40	0,51	0,60	0,70	0,80	0,90	0,99	1,08
$T_{пр}$	-	10,4%	19,9%	30,6%	40,2%	50,7%	59,8%	69,7%	80,1%	89,6%	98,5%	108,0%
А	0,0201											
Цепные показатели												
Δ	-	0,21	0,19	0,22	0,19	0,21	0,18	0,2	0,21	0,19	0,18	0,19

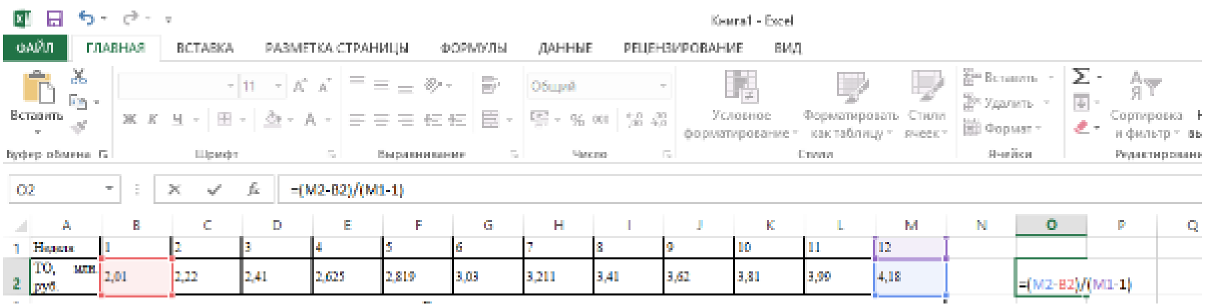
K_p	-	1,10	1,09	1,09	1,07	1,07	1,06	1,06	1,06	1,05	1,05	1,05
T_p	-	110	109	109	107	107	106	106	106	105	105	105
$K_{пр}$	-	0,10	0,09	0,09	0,07	0,07	0,06	0,06	0,06	0,05	0,05	0,05
$T_{пр}$	-	10,45	8,56	8,92	7,39	7,48	5,97	6,20	6,16	5,25	4,72	4,76
A	-	0,0201	0,0222	0,0241	0,02625	0,02819	0,0303	0,03211	0,0341	0,0362	0,0381	0,0399

Рассчитаем средний абсолютный прирост:

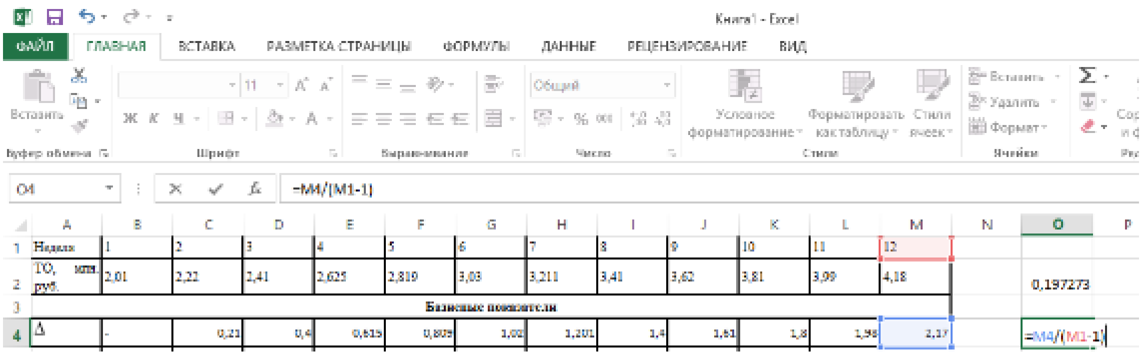
Через цепные приросты:



Через первый и последний уровни ряда:



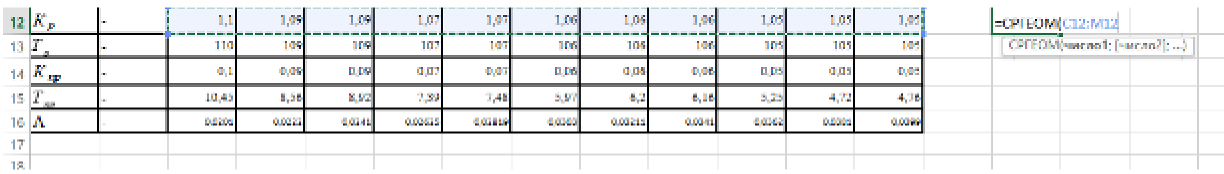
Через последний базисный прирост:



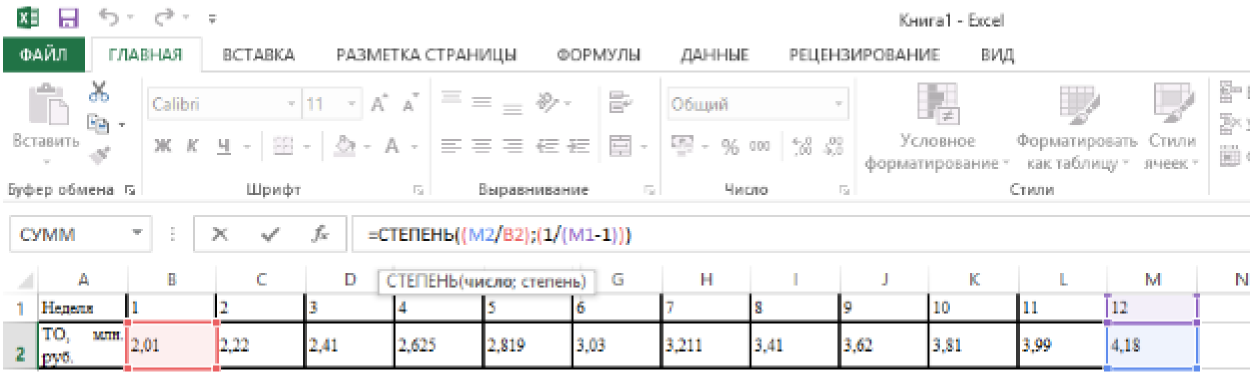
Отметим, что все три способа расчета дали один и тот же результат 0,197273.

Рассчитаем теперь средний коэффициент роста:

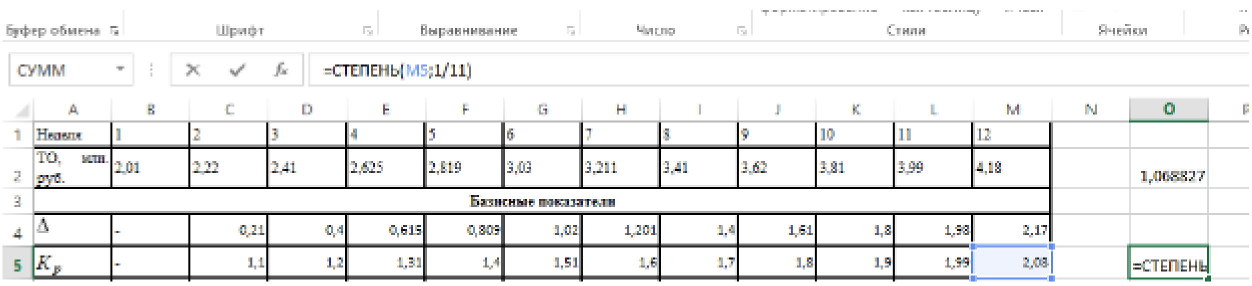
Как среднюю геометрическую цепных коэффициентов роста:



Через первый и последний уровни ряда:



Через последний базисный коэффициент роста:



Отметим, что все три способа расчета дали один и тот же результат 1,068827.

Остальные средние показатели определяются через средний коэффициент роста:

Средний темп роста: 1,068827*100%=106,88%

Средний коэффициент прироста: 1,068827-1=0,068827

Средний темп прироста: 106,88%-100%=6,88%

Решить задачи:

Задача 6.1. Динамика численности работников фирмы представлена в таблице:

Период	1	2	3	4	5	6	7	8	9
Численность персонала	420	429	427	431	442	450	460	465	475

Рассчитать базисные, цепные и средние показатели динамики, сделать выводы.При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 6.2.Розничный товарооборот торговой организации характеризуется следующими данными:

День	1	2	3	4	5	6	7	8	9	10	11	12
Млн. руб.	1580	1584	1582	1581	1587	1585	1582	1581	1586	1584	1582	1587

Определить базисные и цепные показатели динамического ряда, средние показатели динамики. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 6.3.Имеются следующие данные, характеризующие общий объем продукции промышленности региона, млн. руб.:

Год	2006	2007	2008	2009	2010	2011	2012
В старых границах региона	20,1	20,7	21,0	21,2	-	-	-
В новых границах региона				23,8	24,6	25,5	27,2

Привести ряд динамики к сопоставимому виду через коэффициент соотношения и в процентах, определить цепные и средние показатели динамики.При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Занятие 7. Индексы.

Относительная величина, получаемая при сравнении уровней, называется **индивидуальным индексом,** если исследователь не интересуется структурой изучаемого явления и количественную оценку уровня в данных условиях сравнивает с такой же конкретной величиной уровня этого явления в других условиях.

Так, уровень товарооборота в виде суммы выручки от продажи товара в условиях отчетного года Q_1 сравнивается с аналогичной суммой выручки базисного года Q_0. В итоге получаем индивидуальный индекс товарооборота

$$i_Q = Q_1 / Q_0.$$

Аналогичные индивидуальные индексы можно рассчитать и для любого интересующего нас показателя. В частности, поскольку сумма выручки определяется ценой товара (p) и количеством продаж в натуральном измерении (q), можно определить индивидуальные индексы цены i_p и количества проданных товаров – i_q :

$$\mathbf{i}_p = \mathbf{p}_1 / \mathbf{p}_0, \ \mathbf{i}_q = \mathbf{q}_1 / \mathbf{q}_0.$$

С аналитической точки зрения i_q показывает, во сколько раз увеличилась (или уменьшилась) общая сумма выручки под влиянием изменения объема продажи в натуральных единицах.

Аналогично i_p показывает, во сколько раз изменилась общая сумма выручки под влиянием изменения цены товара. Очевидно, что

$$\mathbf{i}_Q = \mathbf{i}_q \cdot \mathbf{i}_p, \text{ или } \mathbf{Q}_1 = \mathbf{Q}_0 \mathbf{i}_q \cdot \mathbf{i}_p.$$

Вторая формула представляет двухфакторную индексную мультипликативную модель итогового показателя, в данном случае – объема товарооборота. Посредством такой модели находят прирост итога под влиянием каждого фактора в отдельности.

Так, если выручка от продажи некоторого товара возросла с 8 млн. руб. в предыдущем периоде до 12,180 млн. руб. в последующем и известно, что это объясняется увеличением количества проданного товара на 5 % при цене на 45 % большей, чем в предыдущем периоде, то можно записать следующее соотношение:

$$12{,}180 = 8 \times 1{,}05 \times 1{,}45 \text{ (млн. руб.)}.$$

Очевидно, что общий прирост выручки в сумме 12,180-8 = 4,180 млн. руб. объясняется изменением объема продажи и цены. Прирост выручки за счет изменения объема продажи (в натуральном выражении) составит

$$\Delta Q(q) = Q_0 \cdot (i_q - 1),$$

или в нашем примере

$$\Delta Q(q) = 8 \cdot (1{,}05 - 1) = +0{,}40 \ \text{млн руб}.$$

Тогда за счет изменения цены данного товара сумма выручки изменилась на

$$\Delta Q(q) = (Q_1 - Q_0) - \Delta Q(q) = Q_1 - Q_0 \cdot i_q = Q_0 \cdot i_q \cdot (i_p - 1),$$

или $\Delta Q(p) = 8 \cdot 1{,}05 \cdot (1{,}45 - 1) = +3{,}78 \ \text{млн руб}.$

Очевидно, что общий прирост товарооборота складывается из приростов, объясняемых каждым фактором в отдельности, т.е.

$$\Delta Q = Q_1 - Q_0 = \Delta Q(q) + \Delta Q(p),$$

или $\Delta Q = 12{,}18 - 8 = 0{,}40 + 3{,}78 = 4{,}18$ млн руб.

Можно заметить, что существует и другой способ распределения общего прироста по факторам в двухфакторной индексной мультипликативной модели, а именно:

$$\Delta Q(q) = Q_0 \cdot i_p \cdot (i_q - 1) \text{ и } \Delta Q(p) = Q_0 \cdot (i_p - 1).$$

В нашем примере общий прирост выручки (4,18 млн. руб.) объясняется теперь:

изменением цены

$$\Delta Q(p) = 8 \cdot (1{,}45 - 1) = 3{,}60 \text{ млн руб.},$$

изменением объема продажи

$$\Delta Q(q) = 8 \cdot 1{,}45 \cdot (1{,}05 - 1) = 0{,}58 \text{ млн руб.}$$

Выбор конкретной формы разложения общего прироста итога должен определяться конкретными условиями развития изучаемого показателя, в данном случае – конъюнктурой спроса-предложения. В экономической практике и большинстве научных рекомендаций в настоящее время преобладает первое направление, когда сначала выясняют вклад в общий прирост количественного фактора при базисном уровне качественного признака (цен), а затем – вклад качественного фактора (цены) в расчете на отчетный уровень количественного показателя (объема – q).

Индекс становится общим, когда в расчетной формуле показывается неоднородность изучаемой совокупности. Примером неоднородной совокупности является общая масса проданных товаров всех или нескольких видов. Тогда сумму выручки можно записать в виде **агрегата** (суммы произведений взвешивающего показателя на объемный), например:

$$Q = \sum p \cdot q.$$

Отношение агрегатов, построенных для разных условий, дает общий индекс показателя в агрегатной форме. Так, например, получают индекс общего объема товарооборота в агрегатной форме:

$$I_Q = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0}.$$

При анализе прироста общего объема товарооборота этот прирост также объясняется изменением уровня цен и количества проданных товаров.

Влияние на прирост товарооборота общего изменения цен выражается *агрегатным индексом цен* I_p, который в предположении первичности изменения количественного показателя (q) и вторичности – качественного (p) имеет вид

$$I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}.$$

Влияние на прирост товарооборота изменения количества проданных товаров отражается *агрегатным индексом физического объема* I_q , который строится также в предположении первичности изменения количественных показателей (q) и вторичности влияния качественных (p):

$$I_q = \frac{\sum p_0 \cdot q_1}{\sum p_0 \cdot q_0}.$$

В форме мультипликативной индексной модели динамика товарооборота будет выражаться соотношениями

$$I_Q = I_q \cdot I_p \text{ или } Q_1 = Q_0 \cdot I_q \cdot I_p,$$

где $Q_0 = \sum p_0 \cdot q_0; Q_1 = \sum p_1 \cdot q_1$.

Если принимается предположение об очередности влияния факторов – сначала q, а затем p, то общий прирост товарооборота будет распределяться по факторам следующим образом:

$$\Delta Q(q) = Q_0 \cdot (I_q - 1);$$

$$\Delta Q(p) = Q_0 \cdot I_q \cdot (I_p - 1).$$

Если же принимается предположение об обратной последовательности влияния факторов – сначала p, затем q, то меняются и формулы разложения прироста и формулы расчета индексов I_q и I_p . Тогда

$$\Delta Q(q) = Q_0 \cdot I_p \cdot (I_q - 1);$$

$$\Delta Q(p) = Q_0 \cdot (I_p - 1).$$

где

$$I_p = (\sum p_1 \cdot q_0) / (\sum p_0 \cdot q_0);$$

$$I_q = (\sum p_1 \cdot q_1) / (\sum p_1 \cdot q_0).$$

Примером мультипликативной индексной модели с большим числом факторов является изменение общей суммы материальных затрат на производство продукции. Сумма затрат зависит от количества выпущенной продукции (индекс I_q), удельных расходов (норм) материала на единицу продукции (индекс I_n) и цены на материалы (индекс I_p). Прирост общей суммы затрат распределяется следующим образом:

$$\Delta M(q) = M_0 \cdot (I_q - 1);$$

$$\Delta M(n) = M_0 \cdot I_q \cdot (I_n - 1);$$

$$\Delta M(p) = M_0 \cdot I_q \cdot I_n \cdot (I_p - 1),$$

где $M_0 = \sum q_0 \cdot n_0 \cdot p_0$, а величины индексов таковы:

индекс увеличения суммы затрат в связи с изменением объемов производства продукции (индекс физического объема)

$$I_q = \frac{\sum q_1 \cdot n_0 \cdot p_0}{\sum q_0 \cdot n_0 \cdot p_0};$$

индекс изменения суммы затрат за счет изменения удельных расходов материала (индекс удельных расходов)

$$I_n = \frac{\sum q_1 \cdot n_1 \cdot p_0}{\sum q_1 \cdot n_0 \cdot p_0};$$

индекс изменения общей суммы затрат, объясняемого изменением цен на материалы (индекс цен на материалы)

$$I_n = \frac{\sum q_1 \cdot n_1 \cdot p_1}{\sum q_1 \cdot n_1 \cdot p_0};$$

Приведем формулы расчета некоторых наиболее употребительных агрегатных индексов.

Индекс изменения общей суммы затрат на производство продукции в зависимости от объема производства (q) *и затрат на единицу* (z):

$$I_c = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_0} = \frac{\sum z_0 \cdot q_1}{\sum z_0 \cdot q_0} \cdot \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1} = I_q \cdot I_z.$$

Индекс изменения общего фонда оплаты труда в связи с изменением общей численности работающих (T) *и заработной платы* (f):

$$I_f = \frac{\sum f_1 \cdot T_1}{\sum f_0 \cdot T_0} = \frac{\sum f_0 \cdot T_1}{\sum f_0 \cdot T_0} \cdot \frac{\sum f_1 \cdot T_1}{\sum f_0 \cdot T_1} = I_T \cdot I_f.$$

Индекс изменения объема продукции в связи с изменением численности работающих (T) *и уровня их выработки* (w):

$$I_Q = \frac{\sum W_1 \cdot T_1}{\sum W_0 \cdot T_0} = \frac{\sum W_0 \cdot T_1}{\sum W_0 \cdot T_0} \cdot \frac{\sum W_1 \cdot T_1}{\sum W_0 \cdot T_1} = I_T \cdot I_W.$$

Индекс изменения объема продукции в связи с изменением объема основных производственных фондов (Ф) *и показателя эффективности их использования – фондоотдачи* (Н):

$$I_Q = \frac{\sum H_1 \cdot Ф_1}{\sum H_0 \cdot Ф_0} = \frac{\sum H_0 \cdot Ф_1}{\sum H_0 \cdot Ф_0} \cdot \frac{\sum H_1 \cdot Ф_1}{\sum H_0 \cdot Ф_1} = I_ф \cdot I_H.$$

Аналогичным образом находят общие агрегатные индексы и по многим другим экономическим показателям. Нетрудно заметить, что используемые в приведенных формулах индексы I_q, I_T, $I_ф$ получаются по методу индекса физического объема, а индексы I_z, I_f, I_W, I_H – по методу индекса цен. Таким образом, рассмотренная выше методика распределения общего прироста товарооборота полностью приложима к анализу прироста продукции, изменения общих затрат на производство, изменения общего фонда оплаты труда и т.д.

На практике расчет общих индексов в агрегатной форме во многих случаях оказывается невозможным. Это связано с тем, что количественный учет осуществляется не везде. Например, в сфере розничной торговли легче получить сведения не о количестве проданных товаров, а об их стоимости. Однако, при этом может быть доступна информация об индивидуальных индексах цен и/или физического объема товарооборота. В связи с этим исчисление общих индексов в виде средних из индивидуальных индексов получило широкое применение.

Рассмотрим случай, когда мы не располагаем данными о физическом объеме товарооборота (q_0 и q_1), а имеем информацию: 1) о товарообороте в действующих ценах (p_0q_0 и p_1q_1) и 2) об индивидуальных индексах цен (i_p). По этим данным, учитывая, что $i_p=p_1/p_0$ и, следовательно, $p_0=p_1/i_p$, рассчитаем следующие общие индексы:

Общий индекс цен в среднегармонической форме

$$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\Sigma p_1 q_1}{\Sigma \frac{p_1}{i_p} q_1} = \frac{\Sigma p_1 q_1}{\Sigma \frac{p_1 q_1}{i_p}}.$$

Учитывая, что между индексами существует взаимосвязь $I_{pq}= I_p\ I_q$, найдем **общий индекс физического объема товарооборота (в сопоставимых ценах)**

$$I_q = \frac{I_{pq}}{I_p} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} : \frac{\Sigma p_1 q_1}{\Sigma \frac{p_1 q_1}{i_p}} = \frac{\Sigma \frac{p_1 q_1}{i_p}}{\Sigma p_0 q_0}.$$

Теперь рассмотрим случай, когда мы располагаем информацией: 1) о товарообороте в действующих ценах (p_0q_0 и p_1q_1) и 2) об индивидуальных индексах физического объема (i_q). По этим данным, учитывая, что $i_q=q_1/q_0$ и, следовательно, $q_0=q_1/i_q$, рассчитаем следующие общие индексы:

Общий индекс физического объема товарооборота (в сопоставимых ценах) в среднеарифметической форме

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{\Sigma i_q q_0 p_0}{\Sigma q_0 p_0}.$$

Учитывая, что между индексами существует взаимосвязь $I_{pq}= I_p I_q$, найдем **общий индекс цен**.

$$I_p = \frac{I_{pq}}{I_q} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} : \frac{\Sigma i_q p_0 q_0}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_1}{\Sigma i_q p_0 q_0}.$$

Отметим, что **общий индекс цен в среднегармонической форме** и **общий индекс физического объема в среднеарифметической форме** являются преобразованными формами **общего индекса в агрегатной форме**, поэтому для них справедливо соотношение факторного анализа в абсолютном выражении

$$Dpq(pq)= Dpq(p)+Dpq(q),$$

где абсолютные приросты D$pq(p)$ и D$pq(q)$ находят как разницу между числителем и знаменателем соответствующих индексов.

Еще раз отметим, что выбор формы индекса (агрегатная или средняя из индивидуальных индексов) зависит от имеющихся исходных данных.

Кроме того, вычисление общих индексов через индивидуальные позволяет наглядно представить динамику цен и физических объемов товарооборота по отдельным товарам, их роль в формировании общего индекса.

Индексный метод позволяет анализировать изменения средних величин (средняя цена единицы товара $\bar{p}$, средняя себестоимость единицы продукции $\bar{z}$ и т.д.). Обычно рассматривают средние величины для однородных единиц совокупности (однородные товары, однородная продукция).

1. Рассмотрим **индекс средней цены**

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0},$$

где $\bar{p}_1$, $\bar{p}_0$ – средняя цена в текущем и базисном периоде соответственно.

$$\bar{p}_0 = \frac{\Sigma p_0 q_0}{\Sigma q_0} = \Sigma p_0 d_0, \quad \bar{p}_1 = \frac{\Sigma p_1 q_1}{\Sigma q_1} = \Sigma p_1 d_1,$$

где $d_1 = \frac{q_1}{\Sigma q_1}$, $d_0 = \frac{q_0}{\Sigma q_0}$ – удельный вес (доля) отдельных разновидностей товаров в общей совокупности в текущем и базисном периодах соответственно.

Отметим, что $\Sigma d_0 = \Sigma d_1 = 1$.

Отсюда индекс средней цены

$$I_{\bar p} = \frac{\bar p_1}{\bar p_0} = \frac{\dfrac{\Sigma p_1 q_1}{\Sigma q_1}}{\dfrac{\Sigma p_0 q_0}{\Sigma q_0}} = \frac{\Sigma p_1 d_1}{\Sigma p_0 d_0}.$$

Такие индексы называют **индексами переменного состава**, так как они отражают изменение не только индексируемого показателя (цены *p*), но и изменение структуры совокупности (*d*).

2. **Индекс постоянного (фиксированного) состава** отражает изменение только индексируемого показателя при постоянстве структуры совокупности

$$I_p = \frac{\dfrac{\Sigma p_1 q_1}{\Sigma q_1}}{\dfrac{\Sigma p_0 q_1}{\Sigma q_1}} == \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\Sigma p_1 d_1}{\Sigma p_0 d_1}$$

Обратите внимание, что индекс постоянного состава совпадает с общим индексом цен в агрегатной форме.

3. **Индекс структурных сдвигов** характеризует влияние изменения структуры совокупности на изучаемый показатель (среднюю цену).

$$I_{стр} = \frac{\dfrac{\Sigma p_0 q_1}{\Sigma q_1}}{\dfrac{\Sigma p_0 q_0}{\Sigma q_0}} = \frac{\Sigma d_1 p_0}{\Sigma d_0 p_0}.$$

Между индексами переменного состава и постоянного состава существует взаимосвязь

$$I_{\bar p} = I_p \cdot I_{стр}.$$

По такой же методике строят системы индексов при анализе динамики любых средних показателей.

Так как измерение динамики цен является одной из важнейших задач статистического анализа, поэтому различными исследователями предпринимались многочисленные попытки разработать «идеальный» индекс цен. Например, можно отметить индексы цен, предложенные в в 1735 г. Дюто ($I_D = \frac{\sum p_1}{\sum p_0}$), Карли в 1751 г. ($I_K = \frac{\sum \frac{p_1}{p_0}}{n} = \frac{\sum i_p}{n}$) и Джевонсом в 1863 г. ($I_G = \sqrt[n]{\prod \frac{p_1}{p_0}} = \sqrt[n]{\prod i_p}$). Существенный недостаток этих индексов – игнорирование удельных весов товаров в товарообороте.

Широкий класс агрегатных индексов цен можно получить на основе индекса Лоу, впервые предложившего этот индекс в 1823 г.

$$I_{Lo} = \frac{\Sigma p_1 q}{\Sigma p_0 q},$$

где *q* – объемы продаж, задаваемые в различной форме.

Например, в качестве объемов продаж *q* можно использовать данные за базисный период времени ($q=q_0$), тогда получаем индекс цен Ласпейреса

$$I_L = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}.$$

Если в качестве *q* выступают продажи текущего периода ($q=q_1$), то получаем индекс Пааше

$$I_P = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}.$$

Если $q = \frac{q_0 + q_1}{2}$, то получаем индекс Эджворта-Маршалла

$$I_{\partial M} = \frac{\Sigma p_1 \frac{q_0 + q_1}{2}}{\Sigma p_0 \frac{q_0 + q_1}{2}} = \frac{\Sigma p_1 (q_0 + q_1)}{\Sigma p_0 (q_0 + q_1)}.$$

Если $q = \sqrt{q_0 \cdot q_1}$, то получаем индекс Уолша

$$I_W = \frac{\Sigma p_1 \sqrt{q_0 \cdot q_1}}{\Sigma p_0 \sqrt{q_0 \cdot q_1}}.$$

Общий индекс цен можно так же получить как среднюю геометрическую индексов Пааше и Ласпейреса – это индекс Фишера («идеальный» индекс цен)

$$I_F = \sqrt{I_L \cdot I_P} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}.$$

На практике наиболее часто используются индексы Ласпейреса, Пааше и Фишера.

Часто при анализе динамических рядов можно столкнуться с таким явлением как сезонность.

Сезонные колебания обусловлены действием факторов, проявляющихся периодически – зима, весна, лето, осень. При изучении социально-экономических явлений обычно ограничиваются рассмотрением сезонных колебаний, обусловленных сменой времен года. Однако при анализе некоторых экономических процессов необходимо учитывать циклические колебания с периодичность более года (пятидесятилетние циклы Кондратьева, цикличность развития мировой экономики и т.д.).

Для обеспечения корректности анализа сезонных колебаний (и вообще анализа рядов динамики) необходимо обеспечить сопоставимость уровней ряда динамики.

Причины несопоставимости в рядах динамики:

Различная продолжительность месяцев (28, 29, 30, 31 день), кварталов (90, 91 или 92 дня), лет (365 или 366 дней). Для устранения влияния этой причины объемные показатели (выручка, прибыль) пересчитываются в средние показатели (обычно среднесуточные или за условный месяц 30,5 дней).

Изменение цен. Для устранения влияния этой причины стоимостные показатели пересчитывают в неизменные (сопоставимые) цены.

Неоднородность состава изучаемых совокупностей во времени (например, из Алтайского края выделилась Республика Алтай).

Изменение методики учета изучаемого показателя.

Для измерения сезонных колебаний исчисляют индексы сезонности.

В стационарных (стабильных) рядах, в которых нет явно выраженного тренда, индексы сезонности рассчитывают по формуле

$$i_{st} = \frac{y_t}{\bar{y}},$$

где yt – фактический уровень ряда; $\bar{y}$ – средний уровень всего ряда динамики.

Для того чтобы повысить устойчивость оценки сезонных колебаний, индексы сезонности рекомендуется рассчитывать за несколько лет по следующей формуле

$$\overline{i}_{s\,t} = \frac{\overline{y}_i}{y},$$

где $\overline{y}_i$ – средний уровень ряда по одноименным внутригодовым отрезкам времени (месяцам, кварталам).

Для наглядного изображения сезонной волны индексы сезонности изображают в виде графика.

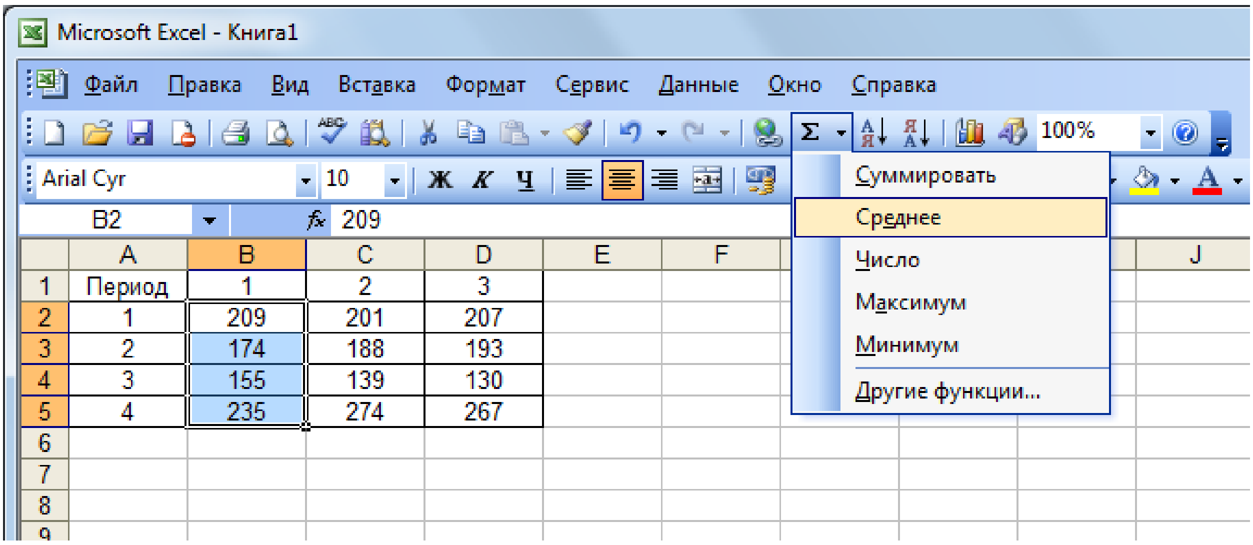
Пример расчетаиндексов сезонности в среде MicrosoftExcel

Дана динамика некого условного показателя за 3 года по кварталам:

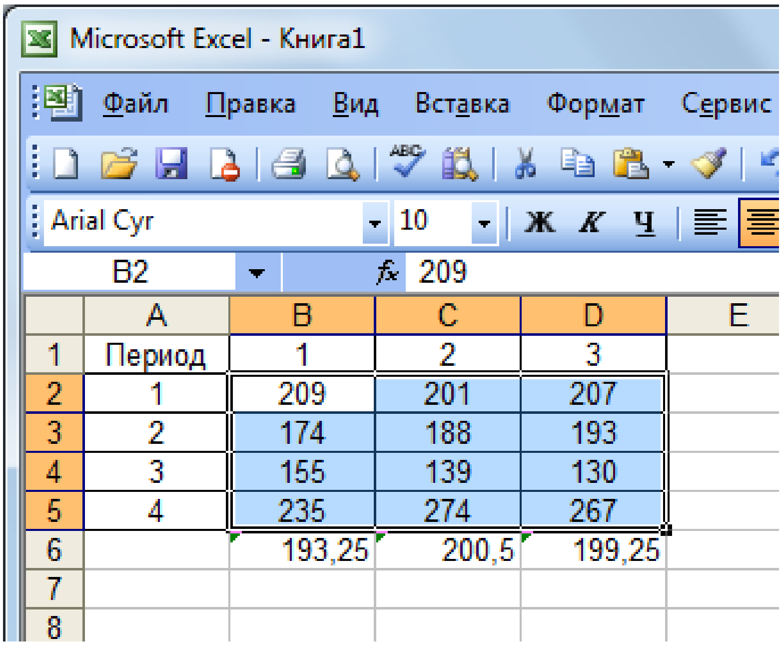
Период	1	2	3
1	209	201	207
2	174	188	193
3	155	139	130
4	235	274	267

Предварительный графический анализ дает возможность сделать вывод о наличии динамики:

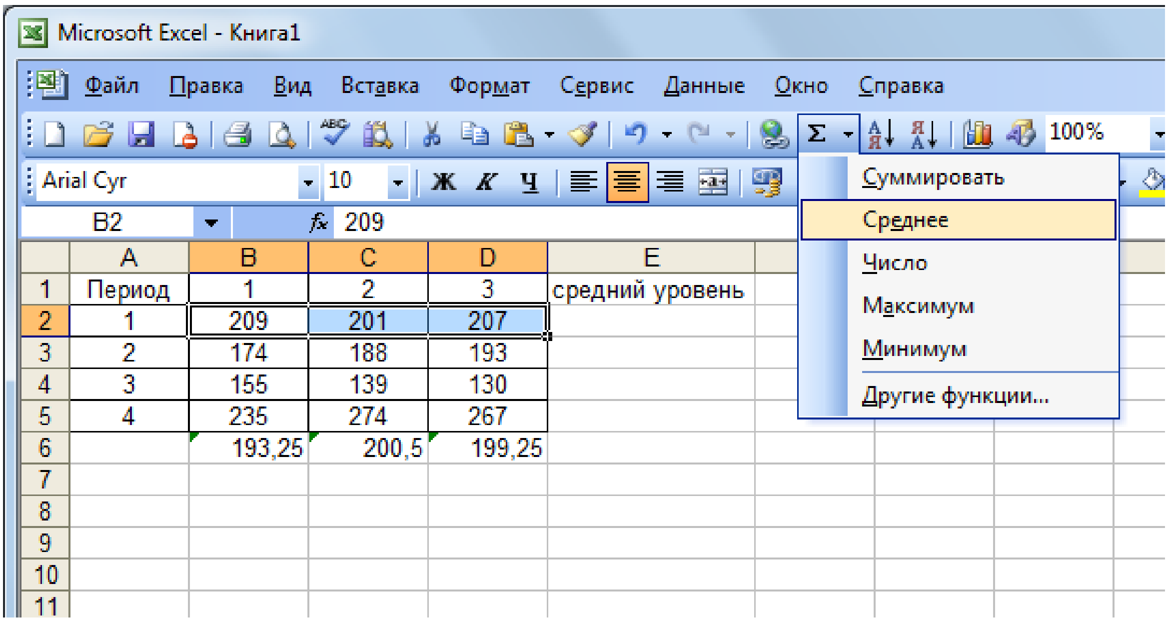
Определим, есть ли рост среднего показателя по годам:



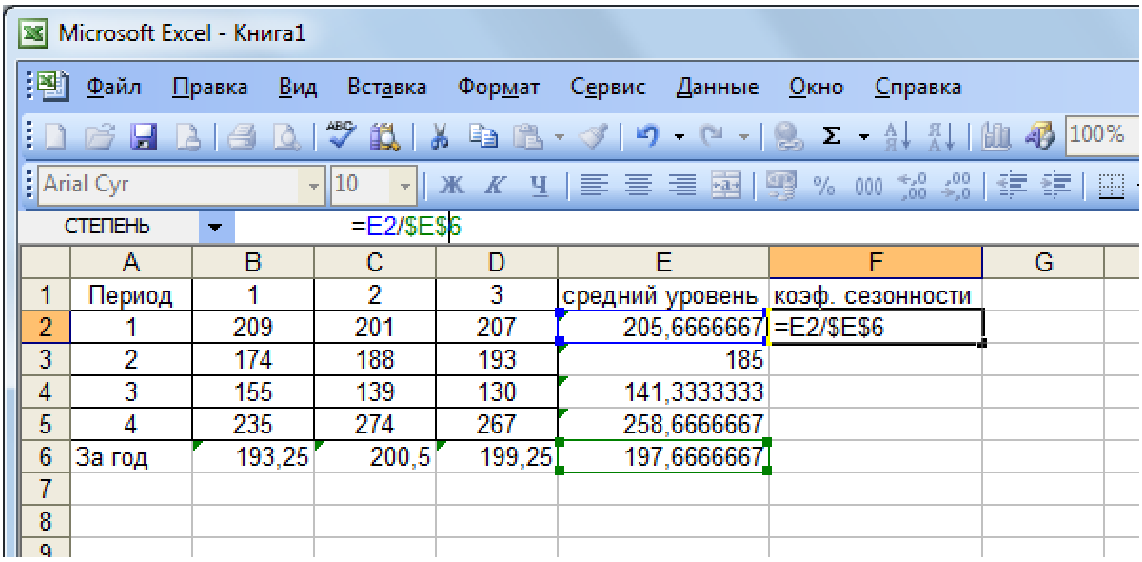
Средние уровни по годам не демонстрируют устойчивой динамики:



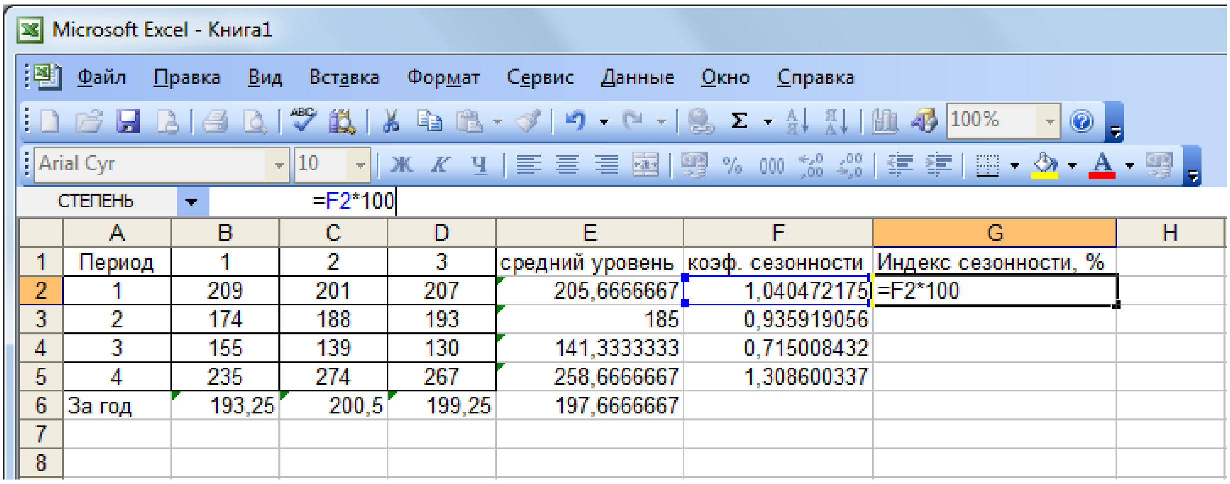
Следовательно можно использовать индексы сезонности, для чего рассчитываем средние уровни показателя за каждый квартал и среднегодовой уровень показателя:



Сопоставим средние уровни по кварталам со среднегодовым уровнем показателя, зафиксировав знаменатель для использования функции АВТОЗАПОЛНЕНИЕ:



Перейдем к процентному выражению:



Индексы сезонности демонстрируют наличие значимой сезонной волны.

Решить задачи:

Задача 7.1. Имеются следующие данные о ценах и объемах реализации товаров предприятия:

Вид продукции	Цена единицы продукции, руб.		Объем реализации, шт.	
	прошлый год	текущий год	прошлый год	текущий год
А	100	110	890	910
Б	85	89	756	745
В	65	72	1000	1030

Определить индивидуальные и агрегатные индексы цен, объема и товарной продукции, сделать выводы.При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 7.2. Имеются данные о динамике производства и себестоимости продукции предприятия:

Вид продукции	Выпуск продукции, тыс. руб.		Себестоимость единицы продукции, руб.	
	III квартал	IV квартал	III квартал	IV квартал
А	6	4	80	91
Б	10	6	20	30

Определите:
1. Индивидуальные индексы себестоимости и физического объема произведенной продукции.
2.Индексы на общие затраты:
а) затрат на выпуск продукции;
б) себестоимости продукции;
в) физического объема продукции.
Покажите взаимосвязь исчисленных индексов и сделайте выводы.

Задача 7.3. Имеются данные о товарообороте по 3 товарным группам:

Группа товаров	Розничный товарооборот в фактических ценах, тыс.р.		Изменение цен в отчетном периоде по сравнению с базисным, %
	Базисный период	Отчетный период	
1	2500	3000	+6
2	2000	2200	+12
3	3000	3180	-4

Определите по 3 группам товаров вместе:
1. Индексы товарооборота в фактических и сопоставимых ценах, индекс цен.
2. Абсолютное изменение товарооборота, в том числе вследствие изменения физического объема товарооборота и цен.
Сделайте выводы.При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 7.4.Имеются данные о реализации фруктовых соков в городе в течение 3-х лет по кварталам (млн. руб.)

Квартал	1-ый год	2-ой год	3-ий год
I	30	34	40
II	45	48	52
III	64	70	68
IV	36	34	38

1.Охарактеризуйте сезонность продажи фруктовых соков, рассчитав индексы сезонности.
2. Постройте график сезонной волны.
Сделайте выводы. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Занятие 8. Парная корреляция и парная линейная регрессия.

Практически для количественной оценки тесноты связи широко используют линейный *коэффициент корреляции.* Иногда его называют просто коэффициентом корреляции. Если заданы значения переменных X и У, то он вычисляется по формуле

$$r_{ух} = r_{ху} = \frac{\overline{XУ} - \overline{X} \cdot \overline{У}}{\sigma_x \cdot \sigma_y}.$$

Коэффициент корреляции принимает значения в интервале от -1 до + 1. Принято считать, что если $|r|< 0{,}30$, то связь слабая; при $|r|= (0{,}3\div0{,}7)$ – средняя; при $|r|> 0{,}70$ – сильная, или тесная. Когда $|r|= 1$ – связь функциональная. Если же r принимает значение около 0, то это дает основание говорить об отсутствии линейной связи между У и X. Однако в этом случае возможно нелинейное взаимодействие. что требует дополнительной проверки и других измерителей.

Для характеристики влияния изменений X на вариацию У служат методы регрессионного анализа. В случае парной линейной зависимости строится регрессионная модель

$$У_i = a_0 + a_1 \cdot X_i + \varepsilon_i,\ i = 1,\ldots,n.$$

где n – число наблюдений;
a_0, a_1 – неизвестные параметры уравнения;
e_i – ошибка случайной переменной У.

Уравнение регрессии записывается как

$$У_{iтеор} = a_0 + a_1X_i,$$

где $У_{iтеор}$ – рассчитанное выравненное значение результативного признака после подстановки в уравнение X.

Параметры a_0 и a_1 оцениваются с помощью процедур, наибольшее распространение из которых получил *метод наименьших квадратов.*

Можно воспользоваться формулами, вытекающими из метода наименьших квадратов, например:

$$a_1 = \frac{\sum(\overline{X_i} - \overline{X})(\overline{У_i} - \overline{У})}{\sum(X_i - X)^2} \text{ или } a_1 = r_{xy}\frac{\sigma_y}{\sigma_x},$$

$$a_0 = \overline{У} - a_1\overline{X}.$$

Получив оценки корреляции и регрессии, необходимо проверить их на соответствие истинным параметрам взаимосвязи.

Существующие программы для ЭВМ включают, как правило, несколько наиболее распространенных критериев. Для оценки значимости коэффициента парной корреляции рассчитывают стандартную ошибку коэффициента корреляции:

$$\sigma_{r^{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}.$$

В первом приближении нужно, чтобы $\sigma_{r^{xy}} < r_{xy}$. Значимость r_{xy} проверяется его сопоставлением с $\sigma_{r^{xy}}$, при этом получают

$$t_{расч} = r_{xy} \cdot \sqrt{\frac{n - 2}{1 - r_{xy}^2}},$$

где $t_{расч}$ – так называемое расчетное значение t-критерия.

Если $t_{расч}$ больше теоретического (табличного) значения критерия Стьюдента ($t_{табл}$) для заданного уровня вероятности и (n-2) степеней свободы, то можно утверждать, что r_{xy} значимо.

Подобным же образом на основе соответствующих формул рассчитывают стандартные ошибки параметров уравнения регрессии, а затем и t-критерии для каждого параметра. Важно опять-таки проверить, чтобы соблюдалось условие $t_{расч} > t_{табл}$. В противном случае доверять полученной оценке параметра нет оснований.

Вывод о правильности выбора вида взаимосвязи и характеристику значимости всего уравнения регрессии получают с помощью F-критерия, вычисляя его расчетное значение:

$$F_{расч} = \frac{R^2(n - m)}{(1 - R^2)(m - 1)},$$

где n – число наблюдений;
m – число параметров уравнения регрессии.

$F_{расч}$ также должно быть больше $F_{теор}$ при $v_1 = (m-1)$ и $v_2 = (n-m)$ степенях свободы. В противном случае следует пересмотреть форму уравнения, перечень переменных и т.д.

Так как регрессия была построена не по генеральной, а по выборочной совокупности, это означает, что полученные значения коэффициентов регрессии не детерминированы, а являются всего лишь оценками истинных коэффициентов и при другой выборке они могут получиться другими. Представление о том, каким же в принципе может быть истинное значение коэффициентов, дает доверительный интервал, который определяется через t-тест:

Пусть α - истинное значение коэффициента регрессии а, т.е. найденное нами а является оценкой α. Тогда доверительный интервал ищется по формуле:

$$a - c.o.(a) \cdot t_{крит} < \alpha < a + c.o.(a) \cdot t_{крит}$$

Здесь $t_{крит}$ - соответствующее значение t-теста, взятое из статистической таблицы распределения Стьюдента.

$c.o.(a)$ - оценка стандартного отклонения функции плотности вероятности (стандартная ошибка) коэффициента а.

Для коэффициента при х парной линейной регрессии стандартная ошибка определяется формулой:

$$c.o.(a_0) = \sqrt{\frac{1}{n-2} D_в(\varepsilon) \cdot \left(1 + \frac{\bar{x}^2}{D_д(x)}\right)}; \quad c.o.(a_1) = \sqrt{\frac{1}{n-2} \left(\frac{D_д(\varepsilon)}{D_д(x)}\right)};$$

Для коэффициентов множественной линейной регрессии при наличии двух факторов формула примет вид:

$$c.o.(a_1) = \sqrt{\frac{D_в(\varepsilon)}{(n-3) \cdot D_д(x)} \cdot \frac{1}{1 - \rho^2_{x1.x2}}}$$

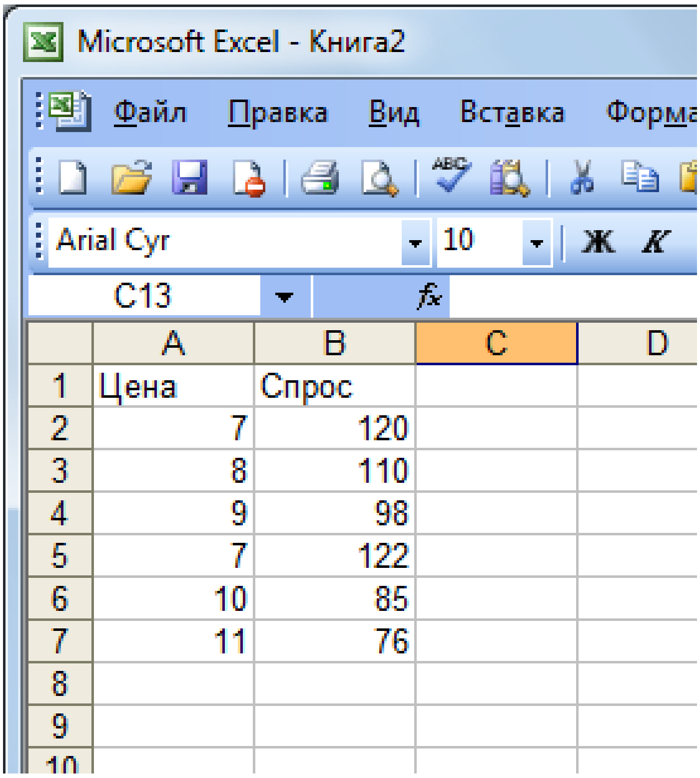
Если полученный интервал включает в себя ноль, это означает, что нельзя исключить отсутствие зависимости.

Рассмотрим пример.

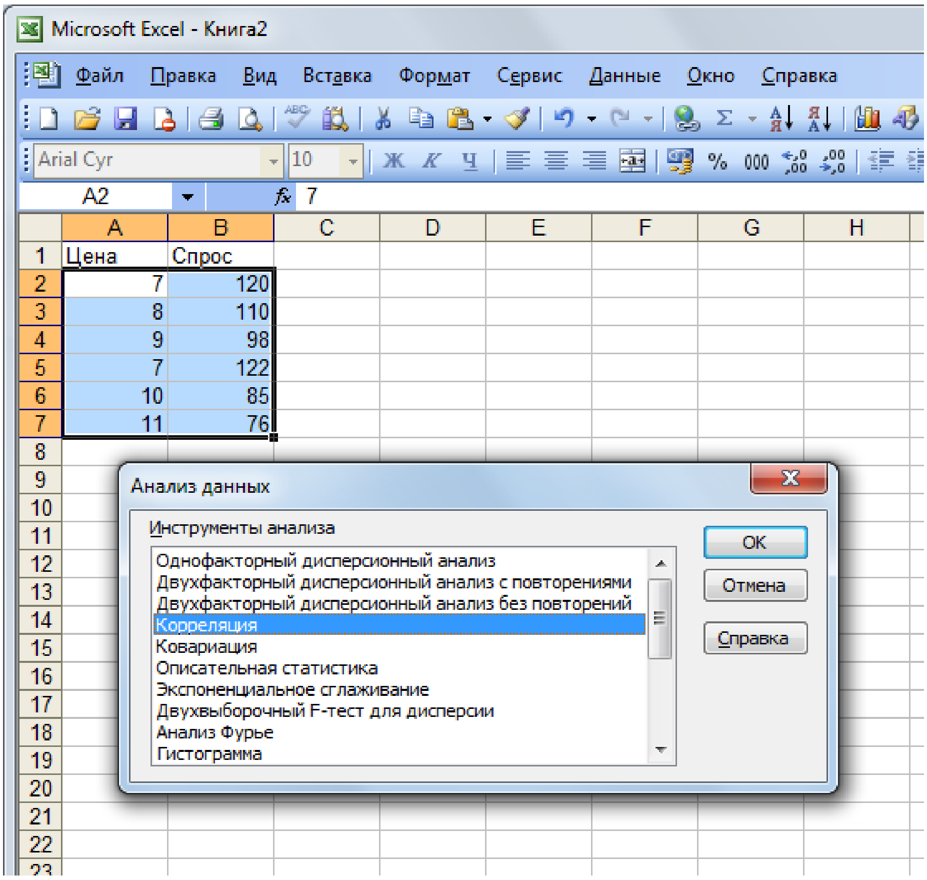
В течение 6 недель менеджер предприятия меняет цену на товар и отслеживает изменение спроса:

Неделя	1	2	3	4	5	6
Цена, у.е.	7	8	9	7	10	11
Объем продаж, шт.	120	110	98	122	85	76

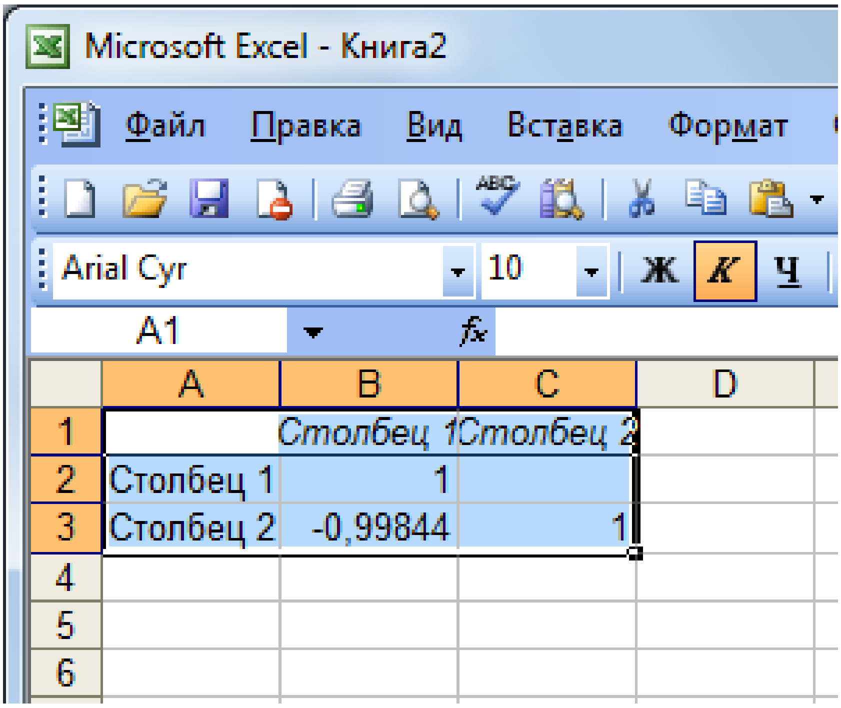
Данные необходимо задать в виде двух столбцов:



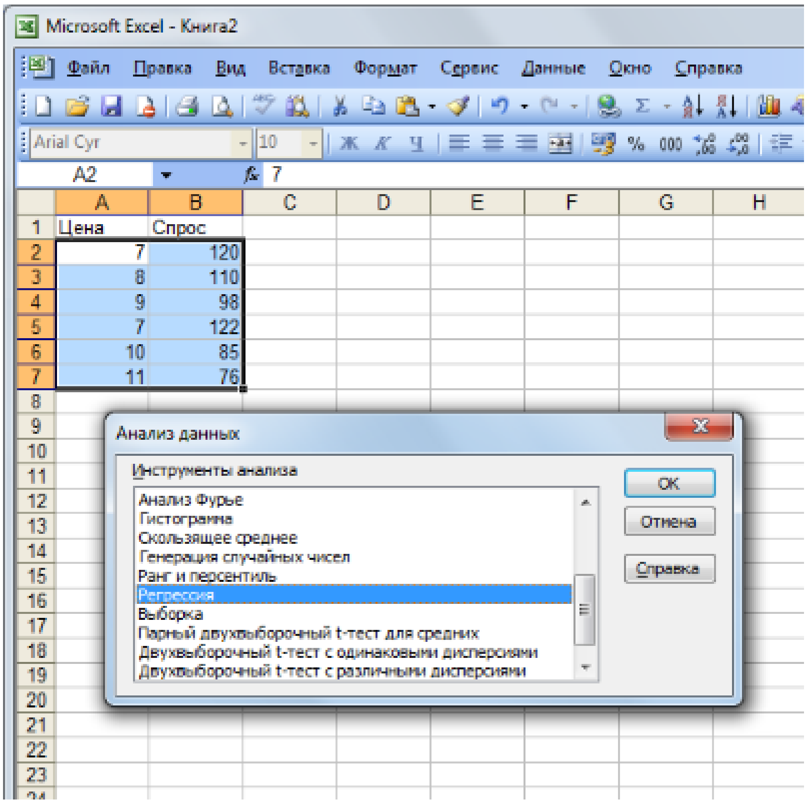
Сначала необходимо сделать предварительный вывод о наличии связи между показателями, используя надстройку Анализ данных:

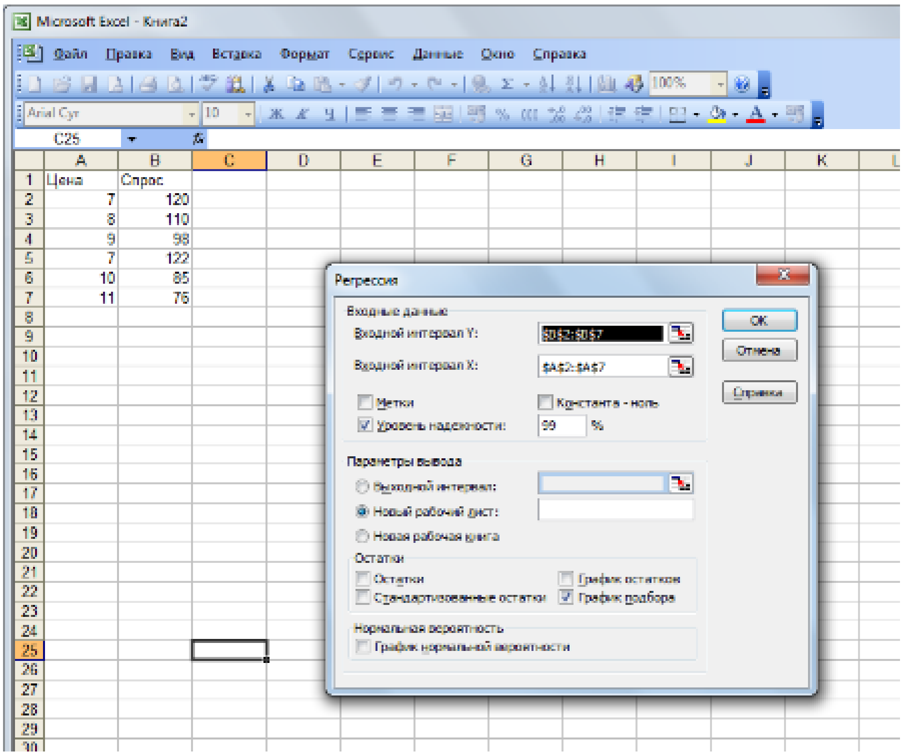


Получили коэффициент линейной корреляции -0,99844, что говорит о тесной обратно пропорциональной связи:

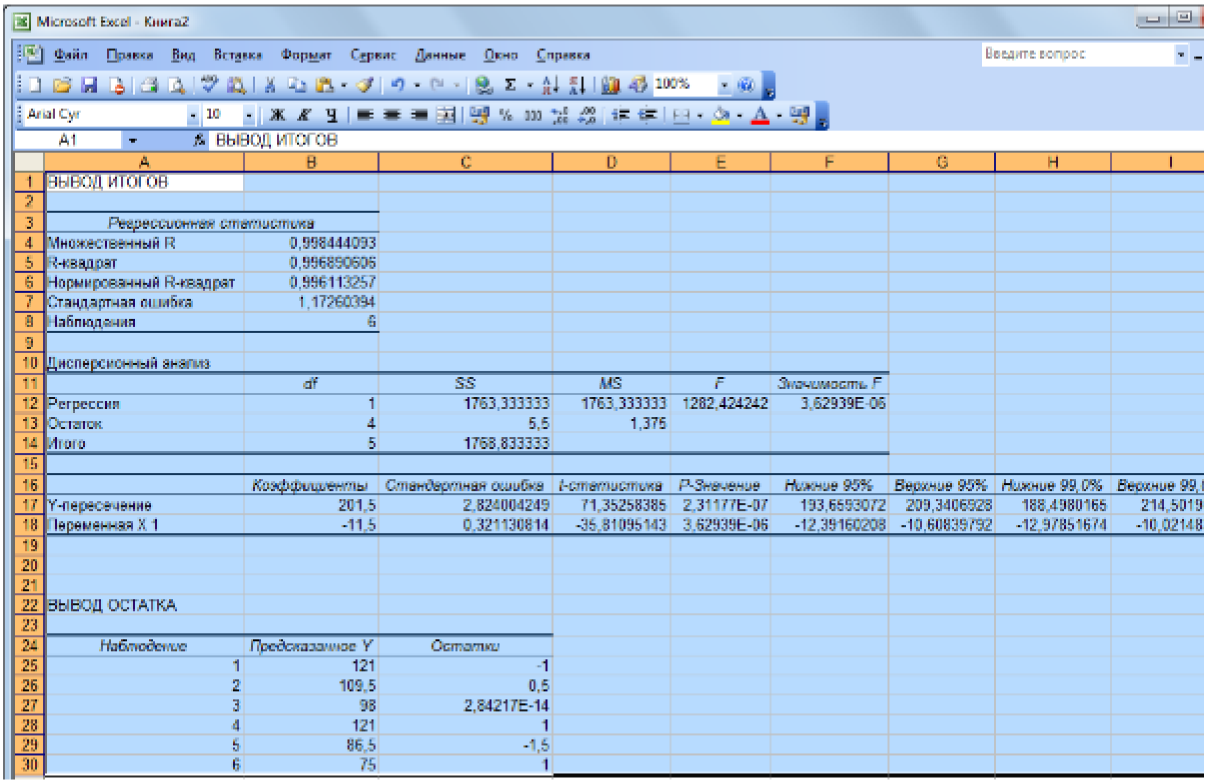


Используем приложение Регрессия:


Microsoft Excel - Книга2
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
Arial Cyr
10
A2
7
Цена
Спрос
Анализ данных
Инструменты анализа
Анализ Фурье
Гистограмма
Скользящее среднее
Генерация случайных чисел
Ранг и персентиль
Регрессия
Выборка
Парный двухвыборочный t-тест для средних
Двухвыборочный t-тест с одинаковыми дисперсиями
Двухвыборочный t-тест с различными дисперсиями
OK
Отмена
Справка


Microsoft Excel - Книга2
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
Цена
Спрос
Регрессия
Входные данные
Входной интервал Y:
$B$2:$B$7
Входной интервал X:
$A$2:$A$7
Метки
Константа - ноль
Уровень надежности:
99
%
Параметры вывода
Выходной интервал:
Новый рабочий лист:
Новая рабочая книга
Остатки
Остатки
График остатков
Стандартизованные остатки
График подбора
Нормальная вероятность
График нормальной вероятности
OK
Отмена
Справка

Получим:



ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,998444093
R-квадрат	0,996890606
Нормированный R-квадрат	0,996113257
Стандартная ошибка	1,17260394
Наблюдения	6

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	1	1763,333333	1763,333333	1282,424242	3,62939E-06
Остаток	4	5,5	1,375		
Итого	5	1768,833333			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 99,0%	Верхние 99,(
Y-пересечение	201,5	2,824004249	71,35258385	2,31177E-07	193,6593072	209,3406928	188,4980165	214,5019
Переменная X 1	-11,5	0,321130814	-35,81095143	3,62939E-06	-12,39160208	-10,60839792	-12,97851674	-10,02148

ВЫВОД ОСТАТКА

Наблюдение	Предсказанное Y	Остатки
1	121	-1
2	109,5	0,5
3	98	2,84217E-14
4	121	1
5	86,5	-1,5
6	75	1

Таким образом, получили уравнение зависимости спроса от цены:

$$Y = 201{,}5 - 11{,}5x$$

Это означает, что при увеличении цены на 1 у.е. спрос будет падать на 11,5 штук.

Коэффициент детерминации очень близок к 1 (0,996890606), что говорит о высоком качестве полученной регрессии, его значимость подтверждает очень высокое значение теста Фишера (1282,424242).

Отсутствие нулей в доверительных интервалах для параметров регрессии даже при 99% уровне значимости (от 188,5 до 214, 5 для свободного члена и от -12,98 до -10, 02 для коэффициента при x) также подтверждает статистическую значимость полученного уравнения регрессии.

Решить задачи:

Задача 8.1.За отчетный период работа предприятий торговли района характеризуется данными:

Предприятия	**Розничный товарооборот, тыс. руб.**	**Издержки обращения, тыс. руб.**
1	511	30,0
2	560	34,0
3	800	46,0

4	465	30,9
5	228	15,9
6	392	25,2
7	640	42,0
8	404	27,0
9	200	16,4
10	425	34,8
11	570	37,0
12	472	28,6
13	250	18,7
14	665	39,0
15	650	36,0
16	620	36,0
17	383	25,0
18	550	38,5
19	750	44,0
20	660	37,0
21	452	27,0
22	563	35,0

Определить тесноту связи с помощью коэффициента корреляции, сделать вывод о наличии зависимости, выполнить регрессионный анализ данных.При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 8.2.В табл. 1 приведены результаты обследования 20 предприятий по следующим показателям:

Y_1 – производительность труда;

Y_2 - рентабельность;

X_1 –среднегодовая численность ППП;

X_2 - среднегодовая стоимость ОПФ;

X_3 - фондоотдача;

X_4 – оборачиваемость нормируемых оборотных средств.

Рассчитать среднее арифметическое значение, дисперсию и среднее квадратическое отклонение для каждого показателя по индивидуальным значениям.

Таблица 1

Y_1	Y_2	X_1	X_2	X_3	X_4
9,26	13,26	26006	167,69	1,45	166,32
9,38	10,16	23935	186,10	1,30	92,88
12,11	13,72	22589	220,45	1,37	158,04
10,81	12,85	21220	169,30	1,65	93,96
9,35	10,63	7394	39,53	1,91	173,88
9,87	9,12	11586	40,41	1,68	162,30
8,17	25,83	26609	102,96	1,94	88,56
9,12	23,39	7801	37,02	1,89	101,16
5,88	14,68	11587	45,74	1,94	166,32
6,30	10,05	9475	40,07	2,06	140,76
6,22	13,99	10811	45,44	1,96	128,52
5,49	9,68	6371	41,08	1,02	177,84
6,50	10,03	26761	136,14	1,85	114,48
6,61	9,13	4210	42,39	0,88	93,24
4,32	5,37	3557	37,39	0,62	126,72
7,37	9,86	14148	101,78	1,09	91,80
7,02	12,62	9872	47,55	1,60	69,12
8,25	5,02	5975	32,61	1,53	66,24
8,15	21,18	16662	103,25	1,40	67,68
8,72	25,17	9166	38,95	2,22	50,40

Провести корреляционно-регрессионный анализ зависимости производительности труда и рентабельности от среднегодовой численности ППП, среднегодовой стоимости ОПФ, фондоотдачи, оборачиваемости нормируемых оборотных средств. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Занятие 9. Непараметрические методы оценки связи.

Если изучается взаимосвязь двух качественных признаков, каждый из которых состоит только из двух групп, применяются коэффициенты ассоциации и контингенции. Они вычисляются по следующей вспомогательной таблице:

a	b	a+b

c	d	c+d
a+c	b+d	a+b+c+d

Коэффициент ассоциации:

$$K_a = \frac{ad - bc}{ad + bc}.$$

Коэффициент контингенции:

$$K_k = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}}.$$

Связь считается подтвержденной, если $K_a > 0{,}5$ или $K_k > 0{,}3$.

Если каждый из качественных признаков состоит более чем из двух групп, то используют комбинационное распределение единиц совокупности в форме так называемых *таблиц взаимной сопряженности.*

Рассмотрим методику анализа таблиц взаимной сопряженности на конкретном примере социальной мобильности как процесса преодоления замкнутости отдельных социальных и профессиональных групп населения. Ниже приведены данные о распределении выпускников средних школ по сферам занятости с выделением аналогичных общественных групп их родителей.

Занятия родителей	Число детей, занятых в				*Всего*
	Промышленности и строительстве	сельском хозяйстве	сфере обслуживания	сфере интеллектуального труда	
1. Промышленность и строительство	40	5	7	39	91
2. Сельское хозяйство	34	29	13	12	88
3. Сфера обслуживания	16	6	15	19	56
4. Сфера интеллектульного труда	24	5	9	72	110
Всего	114	45	44	142	345

Распределение частот по строкам и столбцам таблицы взаимной сопряженности позволяет выявить основные закономерности социальной мобильности: 42,9 % детей родителей группы 1 («Промышленность и строительство») заняты в сфере интеллектуального труда (39 из 91); 38,9 % детей. родители которых трудятся в сельском хозяйстве, работают в промышленности (34 из 88) и т.д.

Можно заметить и явную наследственность в передаче профессий. Так, из пришедших в сельское хозяйство 29 человек, или 64,4 %, являются детьми работников сельского хозяйства; более чем у 50 % в сфере интеллектуального труда родители относятся к той же социальной группе и т.д.

Однако важно получить обобщающий показатель, характеризующий тесноту связи между признаками и позволяющий сравнить проявление связи в разных совокупностях. Для этой цели исчисляют, например, *коэффициенты взаимной сопряженности* Пирсона (С) и Чупрова (К):

$$C = \sqrt{\frac{\phi^2}{1+\phi^2}}$$

$$K = \sqrt{\frac{\phi^2}{\sqrt{(K_1-1)(K_2-1)}}},$$

где f^2 – показатель средней квадратической сопряженности, определяемый путем вычитания единицы из суммы отношений квадратов частот каждой клетки корреляционной таблицы к произведению частот соответствующего столбца и строки:

$$\phi^2 = \sum_{ij} \frac{\phi_{ij}^2}{f_i f_j} - 1, \; f_i = \sum_j f_{ij}, \; f_j = \sum_i f_{ij};$$

K_1 и K_2 – число групп по каждому из признаков. Величина коэффициента взаимной сопряженности, отражающая тесноту связи между качественными признаками, колеблется в обычных для этих показателей пределах от 0 до 1.

В социально-экономических исследованиях нередко встречаются ситуации, когда признак не выражается количественно, однако единицы совокупности можно упорядочить. Такое упорядочение единиц совокупности по значению признака называется ранжированием. Примерами могут быть ранжирование студентов (учеников) по способностям, любой совокупности людей по уровню образования, профессии, по способности к творчеству и т.д.

При ранжировании каждой единице совокупности присваивается ранг, т.е. порядковый номер. При совпадении значения признака у различных единиц им присваивается объединенный средний порядковый номер. Например, если у 5-й и 6-й единиц совокупности значения признаков одинаковы, обе получат ранг, равный (5 + 6) / 2 = 5,5.

Измерение связи между ранжированными признаками производится с помощью ранговых коэффициентов корреляции Спирмена (r) и Кендэлла (t). Эти методы применимы не только для качественных, но и для количественных показателей, особенно при малом объеме совокупности, так как непараметрические методы ранговой корреляции не связаны ни с какими ограничениями относительно характера распределения признака.

Для расчета коэффициента Спирмена значениям признаков x и y присваивают определенный ранг (N_x и N_y) – порядковый номер в ранжированном ряду. Затем для каждой пары рангов находят их разность и осуществляется расчет коэффициента Спирмена по следующей формуле:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)},$$

где d – разность рангов признаков x и y;

n – число наблюдений.

Если у X и $У$ присутствуют связанные ранги, формула усложняется:

$$A = \sum \left(k_j(X)^3 - k_j(X) \right),$$
$$B = \sum \left(k_j(Y)^3 - k_j(Y) \right).$$

$k_j(X), k_j(Y)$ – мощности j-ой связки у X и $У$.

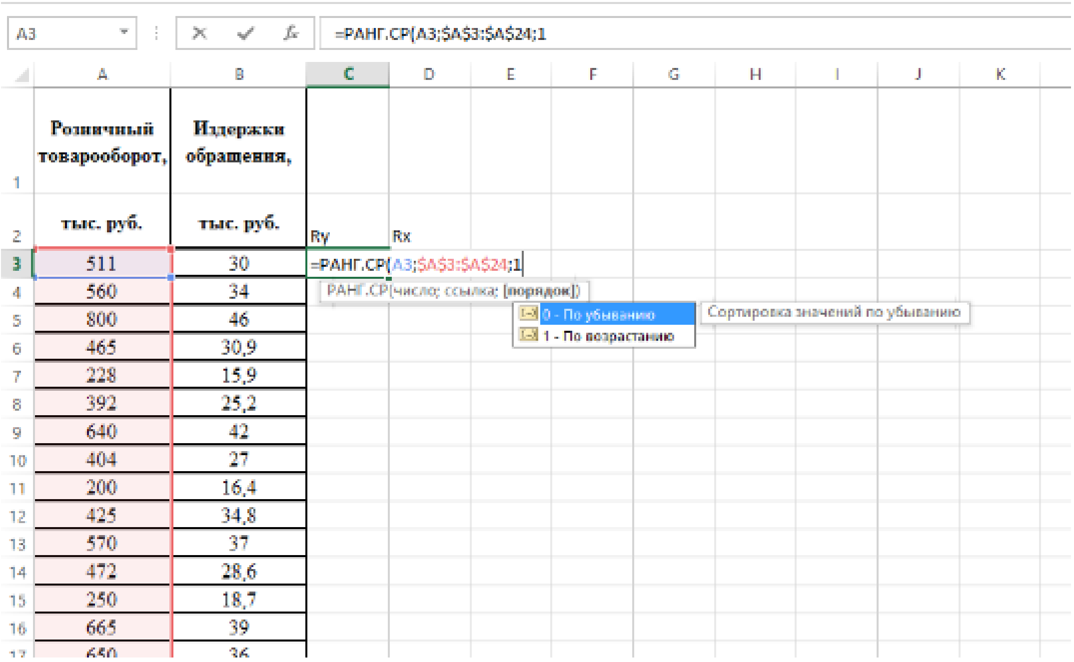
Рассмотрим пример.

Выше уже изучалась зависимость розничного товарооборота от издержек обращения:

Розничный товарооборот, тыс. руб.	**Издержки обращения, тыс. руб.**
511	30,0
560	34,0
800	46,0
465	30,9
228	15,9
392	25,2
640	42,0
404	27,0
200	16,4
425	34,8
570	37,0
472	28,6
250	18,7
665	39,0
650	36,0
620	36,0
383	25,0
550	38,5
750	44,0

660	37,0
452	27,0
563	35,0

Для ранжирования данных используем функцию РАНГ.СР (и АВТОЗАПОЛНЕНИЕ):



Определив также Rx, найдем квадраты разностей рангов:

D3 =(C3-D3)^2

	A	B	C	D	E
1	**Розничный товарооборот,**	**Издержки обращения,**			
2	**тыс. руб.**	**тыс. руб.**	Ry	Rx	d^2
3	511	30	11	9	=(C3-D3)^2
4	560	34	13	11	
5	800	46	22	22	
6	465	30,9	9	10	
7	228	15,9	2	1	
8	392	25,2	5	5	
9	640	42	17	20	
10	404	27	6	6,5	
11	200	16,4	1	2	
12	425	34,8	7	12	
13	570	37	15	16,5	
14	472	28,6	10	8	
15	250	18,7	3	3	
16	665	39	20	19	
17	650	36	18	14,5	

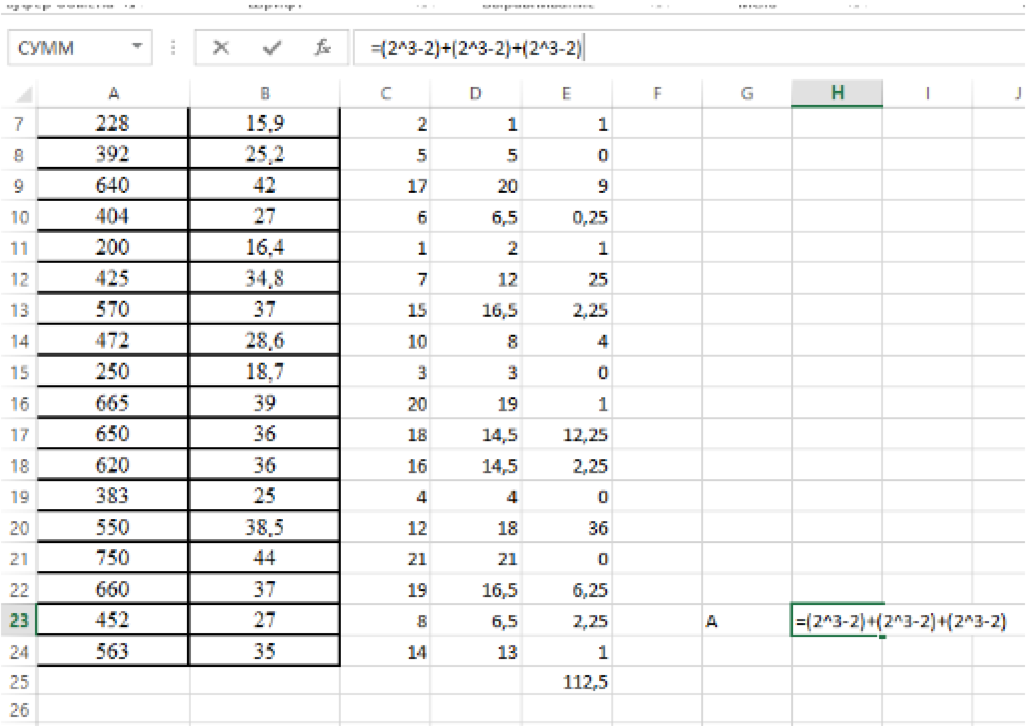
Далее просуммировав разности, определим сам коэффициент Спирмена:

СУММ =СУММ(E3:E24)

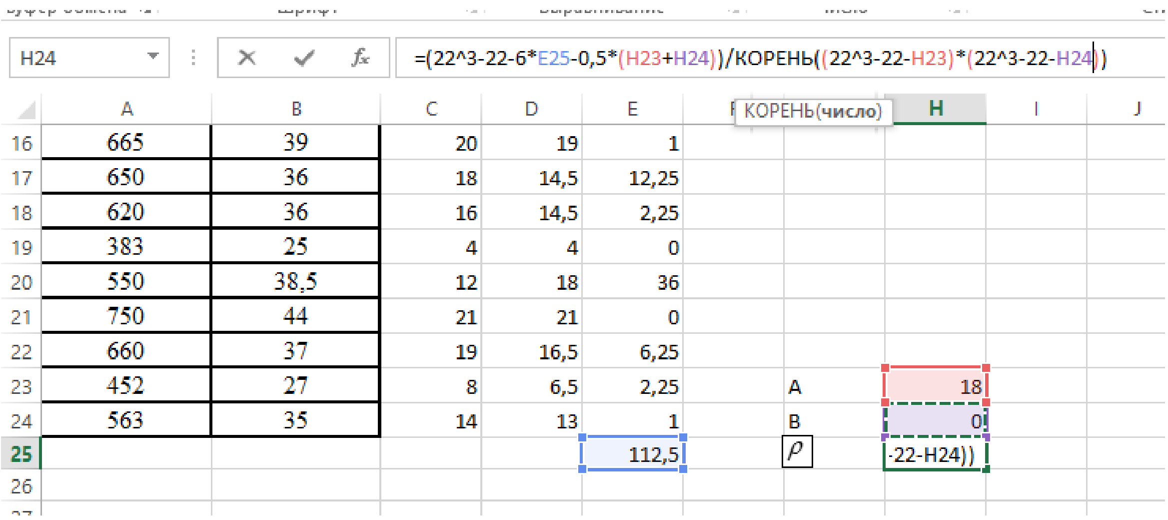
СУММ(число1; [число2]; ...)

	A	B	C	D	E
6	465	30,9	9	10	1
7	228	15,9	2	1	1
8	392	25,2	5	5	0
9	640	42	17	20	9
10	404	27	6	6,5	0,25
11	200	16,4	1	2	1
12	425	34,8	7	12	25
13	570	37	15	16,5	2,25
14	472	28,6	10	8	4
15	250	18,7	3	3	0
16	665	39	20	19	1
17	650	36	18	14,5	12,25
18	620	36	16	14,5	2,25
19	383	25	4	4	0
20	550	38,5	12	18	36
21	750	44	21	21	0
22	660	37	19	16,5	6,25
23	452	27	8	6,5	2,25
24	563	35	14	13	1
25					IM(E3:E24)

По Y связанных рангов нет, по X три пары связанных рангов:


СУММ =(2^3-2)+(2^3-2)+(2^3-2)

	A	B	C	D	E	F	G	H
7	228	15,9	2	1	1			
8	392	25,2	5	5	0			
9	640	42	17	20	9			
10	404	27	6	6,5	0,25			
11	200	16,4	1	2	1			
12	425	34,8	7	12	25			
13	570	37	15	16,5	2,25			
14	472	28,6	10	8	4			
15	250	18,7	3	3	0			
16	665	39	20	19	1			
17	650	36	18	14,5	12,25			
18	620	36	16	14,5	2,25			
19	383	25	4	4	0			
20	550	38,5	12	18	36			
21	750	44	21	21	0			
22	660	37	19	16,5	6,25			
23	452	27	8	6,5	2,25		A	=(2^3-2)+(2^3-2)+(2^3-2)
24	563	35	14	13	1			
25					112,5			
26								


H24 =(22^3-22-6*E25-0,5*(H23+H24))/КОРЕНЬ((22^3-22-H23)*(22^3-22-H24))

КОРЕНЬ(число)

	A	B	C	D	E	F	G	H
16	665	39	20	19	1			
17	650	36	18	14,5	12,25			
18	620	36	16	14,5	2,25			
19	383	25	4	4	0			
20	550	38,5	12	18	36			
21	750	44	21	21	0			
22	660	37	19	16,5	6,25			
23	452	27	8	6,5	2,25		A	18
24	563	35	14	13	1		B	0
25					112,5		ρ	-22-H24))
26								

Полученное значение (0,936) говорит о сильной связи.

Для расчета коэффициента Кендэла значениям признаков x и y присваивают определенный ранг (N_x и N_y), причем ранги факторного признака N_x располагают строго в порядке возрастания и параллельно записывают каждому значению N_x соответствующее значение N_y. Для каждого N_y последовательно определяют число следующих за ним рангов, превышающих его значение, и число рангов, меньших по значению. Первые учитываются как баллы со знаком «+» (положительные баллы), вторые – как баллы со знаком «-» (отрицательные баллы).

Расчет коэффициента Кендэла производится по следующей формуле:

$$\tau = \frac{2S}{n(n-1)},$$

где $S = P + Q$ – сумма положительных и отрицательных баллов;

P – сумма баллов со знаком «+»;

Q - сумма баллов со знаком «-»;

n - число наблюдений.

Если исследуемые данные повторяются (имеют одинаковые ранги), то в расчетах используется скорректированный коэффициент корреляции Кендэла:

$$\tau = \frac{S}{\sqrt{\left[\frac{n(n-1)}{2} - U_x\right]\left[\frac{n(n-1)}{2} - U_y\right]}}$$

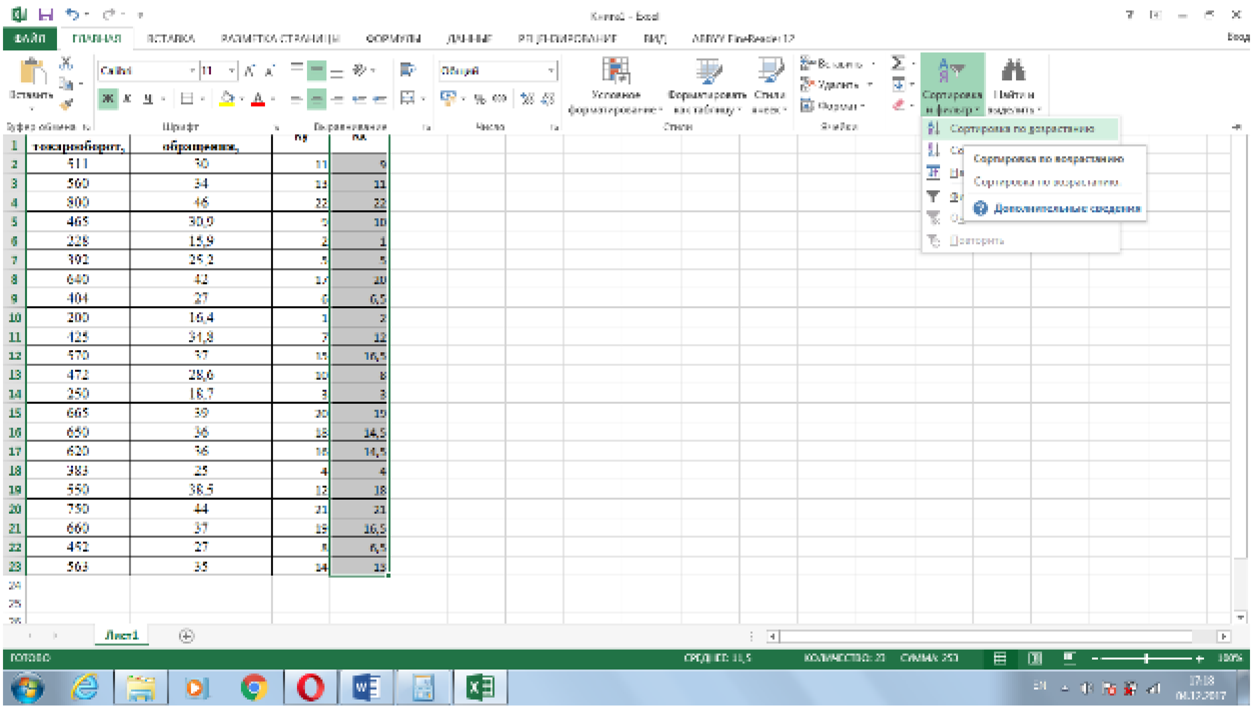
$$U_x = \frac{\sum t(t-1)}{2}$$

$$U_y = \frac{\sum t(t-1)}{2}$$

t - число связанных рангов в ряду X и Y соответственно.

Рассмотрим расчет коэффициента Кендэла на том же примере:

С помощью функции СОРТИРОВКА расположим ранги признака по возрастанию:



	товарооборот,	обращения,	ry	rx
2	511	30	11	9
3	560	34	13	11
4	800	46	22	22
5	465	30,9	9	10
6	228	15,9	2	1
7	392	25,2	5	5
8	640	42	17	20
9	404	27	6	6,5
10	200	16,4	1	2
11	425	31,8	7	12
12	570	37	15	16,5
13	472	28,6	10	8
14	250	18,7	3	3
15	665	39	20	19
16	650	36	18	14,5
17	620	36	16	14,5
18	383	25	4	4
19	550	38,5	12	18
20	750	44	21	21
21	660	37	19	16,5
22	452	27	8	6,5
23	563	35	14	13

И определим положительные и отрицательные баллы (можно использовать функцию СЧЕТЕСЛИ):

СКОС | =СЧЁТЕСЛИМН(D3:D23;">2")

СЧЁТЕСЛИМН(диапазон_условия1; условие1; ...)

	A	B	C	D	P	Q
1	Розничный товарооборот,	Издержки обращения,	Rx	Ry		
2	228	15,9	1	2	=СЧЁТЕСЛ	
3	200	16,4	2	1	20	
4	250	18,7	3	3	19	
5	383	25	4	4	18	
6	392	25,2	5	5	17	
7	404	27	6,5	6	16	
8	452	27	6,5	8	14	
9	472	28,6	8	10	14	
10	511	30	9	11	13	
11	465	30,9	10	9	12	
12	560	34	11	13	11	
13	425	34,8	12	7	10	
14	563	35	13	14	9	
15	650	36	14,5	18	8	
16	620	36	14,5	16	7	
17	570	37	16,5	15	6	
18	660	37	16,5	19	5	
19	550	38,5	18	12	4	
20	665	39	19	20	3	
21	640	42	20	17	2	
22	750	44	21	21	1	
23	800	46	22	22	0	
24						
25						

Лист1

ФАЙЛ | ГЛАВНАЯ | ВСТАВКА | РАЗМЕТКА СТРАНИЦЫ | ФОРМУЛЫ | ДАННЫЕ | РЕЦЕНЗИ

Вставить | Calibri 11 | Общий

Буфер обмена | Шрифт | Выравнивание | Число

СУММ | =СУММ(H3:H24)

СУММ(число1; [число2]; ...)

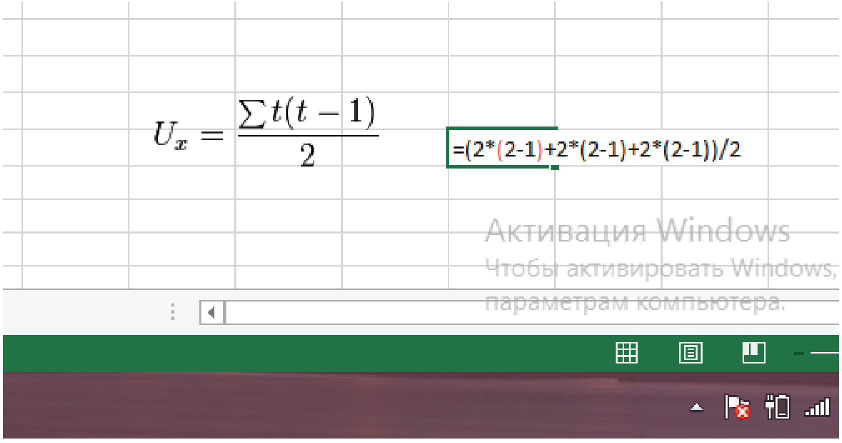
	A	B	C	D	E	F	G	H	I
6	465	30,9	9	10		4	4	18	0
7	228	15,9	2	1		5	5	17	0
8	392	25,2	5	5		6,5	6	16	0
9	640	42	17	20		6,5	8	14	1
10	404	27	6	6,5		8	10	12	2
11	200	16,4	1	2		9	11	11	2
12	425	34,8	7	12		10	9	11	1
13	570	37	15	16,5		11	13	9	2
14	472	28,6	10	8		12	7	10	0
15	250	18,7	3	3		13	14	8	1
16	665	39	20	19		14,5	16	6	2
17	650	36	18	14,5		14,5	18	4	3
18	620	36	16	14,5		16,5	15	5	1
19	383	25	4	4		16,5	19	3	2
20	550	38,5	12	18		18	12	4	0
21	750	44	21	21		19	20	2	1
22	660	37	19	16,5		20	17	2	0
23	452	27	8	6,5		21	21	1	0
24	563	35	14	13		22	22	0	0
25								=СУММ(H	19
26									

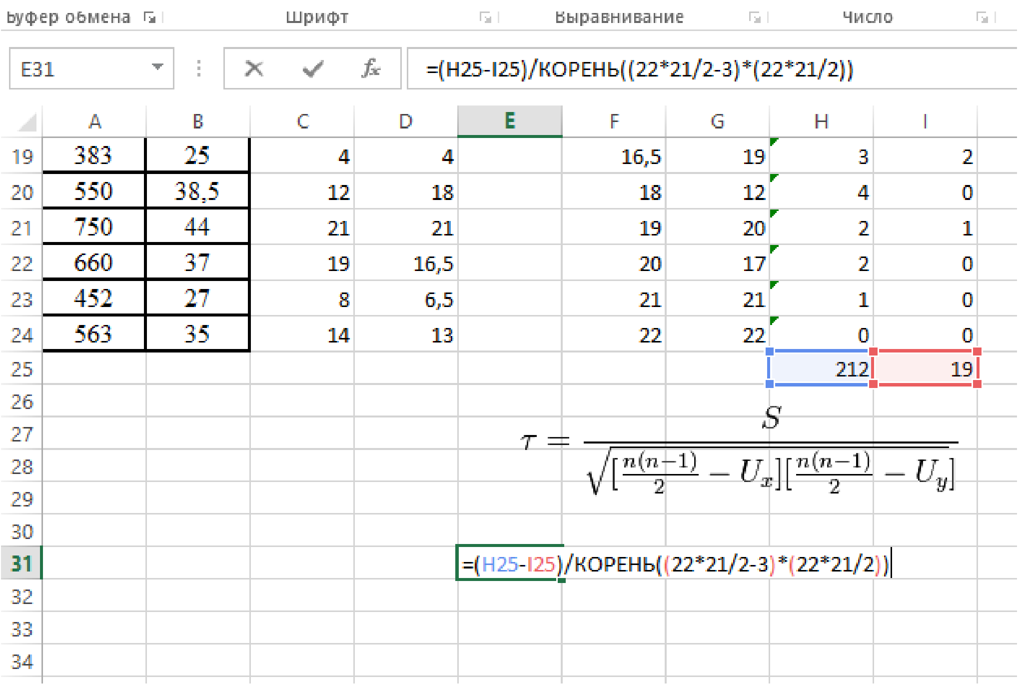
Лист1 | Лист2 | Лист3 | Лист4 | Лист5

ПРАВКА

И сам коэффициент:

Так как присутствуют три пары связанных рангов в X:





	A	B	C	D	E	F	G	H	I
19	383	25	4	4		16,5	19	3	2
20	550	38,5	12	18		18	12	4	0
21	750	44	21	21		19	20	2	1
22	660	37	19	16,5		20	17	2	0
23	452	27	8	6,5		21	21	1	0
24	563	35	14	13		22	22	0	0
25								212	19

Его высокое значение (0,84) также подтверждает наличие связи.

Если число ранжируемых признаков больше двух, то для измерения тесноты связи между ними можно использовать коэффициент конкордации (множественный коэффициент ранговой корреляции), предложенный М.Кендэлом и Б.Смитом:

$$W = \frac{12S}{m^2(n^3 - n)},$$

где S - сумма квадратов отклонений суммы рангов по m факторам от их средней арифметической:

$$S = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} R_{ij} - \overline{\sum R}\right)^2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} R_{ij}\right)^2 - \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij}\right)^2}{n};$$

R_{ij} - ранг i –го фактора у j –й единицы:

$$\sum R = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij}}{n};$$

m – число признаков (факторов);

n – число наблюдений.

Если ранги по отдельным признакам будут повторяться, т.е. будут присутствовать связанные ранги, то коэффициент конкордации рассчитывается с учетом числа повторяющихся (связанных) рангов по каждому фактору:

$$W = \frac{12S}{m^2(n^3 - n) - m\sum_{1}^{m}(t^3 - t)},$$

где t – число одинаковых рангов по каждому признаку.

Значения коэффициента конкордации колеблется в промежутке от 0 до 1.

Если W = 0, то мнения абсолютно несогласованы.

Если W =1, то мнения абсолютно согласованы.

Если W ≥ 0,8, то качество проведенной экспертизы высокое.

Если 0,4<W<0,8, то качество проведенной экспертизы удовлетворительное.

Если W<0,4, то качество проведенной экспертизы неудовлетворительное.

Пример.

Объект Экспертизы	Оценка эксперта				
	1	2	3	4	5
1	4	6	4	4	3
2	3	3	2	3	4
3	2	2	1	2	2
4	6	5	6	5	6

5	1	1	3	1	1
6	5	4	5	6	5
7	7	7	7	7	7

Сумма рангов по каждому объекту экспертизы определяется суммированием результатов опроса экспертов. Например, по первому объекту: 4+6+4+4+3=21. Подобные расчеты следует провести по каждому объекту экспертизы.

Далее определяем среднее арифметическое по проведенным расчетам суммы рангов. В нашем случае среднее суммы рангов равно 20. Полученную сумму квадратов отклонений от среднего (S) подставляем в формулу и определяем коэффициент конкордации. Расчеты сведем в таблицу:

Объект экспертизы	Сумма рангов	Отклонение от среднего	Квадрат отклонения
1	21	1	1
2	15	-5	25
3	9	-11	121
4	28	8	64
5	7	-13	169
6	25	5	25
7	35	15	225
	å		630

W = 0,9

Мнения экспертов согласованы. Значит, результаты опроса можно использовать для дальнейшего исследования.

Решить задачи:

Задача 9.1.Распределение предприятий по источникам средств для их покупки характеризуется следующими данными:

Источник средств	Зарождающийся бизнес	Зрелый бизнес	Итого
Банковский кредит	31	32	63
Собственные средства	38	15	53
Итого	69	47	116

Вычислить коэффициенты ассоциации и контингенции, сделать выводы. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 9.2. Распределение основных категорий потенциальных мигрантов по уровню образования характеризуется следующими данными:

Образование	Основные категории потенциальных мигрантов				Итого
	руководители	специалисты	служащие	рабочие	
Высшее	55	48	12	7	122
Неполное высшее	5	3	3	5	16
Среднее специальное	36	44	51	39	170
Среднее общее	4	4	33	39	80
Неполное среднее	0	1	1	10	12
Итого	100	100	100	100	400

Рассчитать коэффициенты сопряженности Пирсона и Чупрова, сделать выводы. При выполнении данного задания воспользоваться программным пакетом «MicrosoftExcel».

Задача 9.3. по данным задачи 8.2исследовать взаимосвязь показателей с помощью коэффициентов Кендэла и Спирмена, определить коэффициенты конкордации для Y_1 и Y_2.

Объекты исследования, оборудование, материалы и наглядные пособия

Объектом исследования является статистическая информация.

Необходимое оборудование: персональный компьютер с операционной системой Microsoft Windows версии XP, программа MicrosoftExcel.

Материалы: теоретические данные и задания по каждой теме практического и лабораторного занятия.

Наглядные пособия: примеры выполнения заданий с использованием программы MicrosoftExcel по темам.

Список рекомендованной литературы

Основная литература

1. Васильева, Э. К. Статистика [Электронный ресурс] : учебник для студентов вузов, обучающихся по специальностям экономики и управления (080100) / Э. К. Васильева, В. С. Лялин. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2017. — 398 с. — 978-5-238-01192-9. — Режим доступа: http://www.iprbookshop.ru/71058.html

2. Гусаров, В. М. Статистика [Электронный ресурс] : учебное пособие для студентов вузов, обучающихся по экономическим специальностям / В. М. Гусаров, Е. И. Кузнецова. — 2-е изд. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2017. — 479 с. — 978-5-238-01226-1. — Режим доступа: http://www.iprbookshop.ru/71166.html

Дополнительная литература

1 Статистика [Электронный ресурс] : учебное пособие / А. М. Восковых, Т. А. Журкина, С. Л. Закупнев [и др.] ; под ред. И. М. Сурков. — Электрон. текстовые данные. — Воронеж : Воронежский Государственный Аграрный Университет им. Императора Петра Первого, 2017. — 244 с. — 2227-8397. — Режим доступа: http://www.iprbookshop.ru/72755.html

2. Бондаренко, Л. Д. Статистика. Часть 1 [Электронный ресурс] : курс лекций / Л. Д. Бондаренко. — Электрон. текстовые данные. — Новосибирск : Новосибирский государственный архитектурно-строительный университет (Сибстрин), ЭБС АСВ, 2017. — 82 с. — 978-5-7795-0831-5. — Режим доступа: http://www.iprbookshop.ru/85868.html