

Министерство образования и науки Республики Казахстан
Костанайский государственный университет им. А.Байтурсынова

Кафедра бухгалтерского учета и аудита

А.К. Курмангалиева

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ

Учебное пособие

Костанай, 2018

УДК 330:311(075.8)

ББК 65.051я73

Автор. **К 93** Курмангалиева Айжан Касымбековна к.э.н., доцент кафедры бухгалтерского учета и аудита.

Рецензенты:

Притула Р.А. к.э.н., профессор кафедры «Социальные науки и менеджмент», Костанайского инженерно-экономического университета имени М. Дулатова;

Майкопова Г.С. к.э.н., доцент кафедры финансов и банковского дела КГУ имени А. Байтурсынова

Кенжебекова Д.С. к.э.н., доцент, заведующая кафедрой финансов и банковского дела КГУ имени А. Байтурсынова

К 93 Курмангалиева А.К. Учебное пособие. Статистические методы в экономике.-Костанай: КГУ им. А. Байтурсынова, 2018г. -112 с.

ISBN 978-601-7955-08-3

В учебном пособии рассматриваются теоретические, методологические и практические вопросы, даны примеры решения типовых задач, а также предлагаются задачи для самостоятельного решения.

Учебное пособие предназначено для студентов всех специальностей экономического факультета, а также лиц занимающихся самообразованием в системе заочного обучения, включая дистанционное.

УДК 330:311(075.8)

ББК65.051я73

Утверждено и рекомендовано к изданию учебно-методическим советом Костанайского государственного университета им. А. Байтурсынова

21. _02_.2018 г., протокол №_1_

© Костанайский государственный
университет им. А. байтурсынова, 2018

Содержание

Введение.....	3
Тема 1. Статистическая сводка и группировка.....	5
Тема 2. Абсолютные и относительные статистические величины.....	15
Тема 3. Средние величины и показатели вариации.....	26
Тема 4. Выборочное наблюдение.....	43
Тема 5. Ряды динамики.....	54
Тема 6. Индексы.....	75
Тема 7. Статистическое изучение взаимосвязей.....	93
Список использованных источников.....	109
Приложение.....	111

Введение

Учебная дисциплина «Статистические методы в экономике» является обязательным компонентом в подготовке дипломированных специалистов по экономическим специальностям. Основное назначение данной дисциплины состоит в повышении экономико-математической подготовки студентов в области современных методов сбора, обработки и анализа статистической информации, достижении высокого и устойчивого уровня профессионализма. Современный специалист должен обладать глубокими знаниями, уметь проводить количественный анализ сложных экономических проблем, применять математические расчеты в решении экономических задач. Поэтому изучение данной дисциплины поможет сформировать у студентов целостный взгляд на место и роль статистической науки в современной экономике.

Цель преподавания курса «Статистические методы в экономике» – подготовка специалистов, владеющих современными методами сбора, обработки и анализа статистической информации, принятыми в отечественной и международной практике учета и статистики.

Задачи курса: овладение комплексом статистических методов наблюдения, сводки и группировки массовых данных; освоение системы статистических величин, характеризующих количественную сторону социально-экономических явлений и процессов; применение методов статистического анализа при исследовании различных сфер экономики.

Тема 1. Статистическая сводка и группировка

1.1 Методические указания и решения типовых задач

Один из этапов статистического исследования – сводка и группировка данных.

Сводка - упорядочивание и обобщение первичного материала, сводка его в группы и выдача на этой основе обобщающих характеристик совокупности.

Составными элементами сводки являются: программа сводки; подсчет групповых итогов; оформление конечных результатов сводки в виде таблиц и графиков.

Различают **простую сводку** (подсчет только общих итогов) и **статистическую группировку**, которая сводится к расчленению совокупности на группы по существенному для единиц совокупности признаку. Группировка позволяет получить такие результаты, по которым можно выявить состав совокупности, характерные черты и свойства типичных явлений, обнаружить закономерности и взаимосвязи.

Результаты сводки могут быть представлены в виде статистических рядов распределения.

Пусть из совокупности извлечена выборка, причем x_1 наблюдалось n_1 раз, x_2 n_2 раз, x_k – n_k раз и $\sum n_i = n$ - объем выборки. Наблюдаемые значения x_i называют **вариантами**, а последовательность вариантов, записанных в возрастающем порядке, - **вариационным рядом**. Числа наблюдений называют **частотами**, а их отношения к объему выборки $n_i/n = w_i$ - **относительными частотами**.

$$n = \sum_{i=1}^k n_i ; \quad w_i = \frac{n_i}{n} ; \quad n_i^{\text{нак}} \frac{n_i^{\text{нак}}}{n} . \quad (1)$$

Статистическим распределением выборки называют перечень вариантов и соответствующих им частот или относительных частот. Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот (в качестве частоты, соответствующей интервалу, принимают сумму частот, попавших в этот интервал).

Статистическим рядом распределения называют упорядоченное распределение единиц совокупности на группы по изучаемому признаку. В зависимости от признака ряды могут быть вариационными (количественными) и атрибутивными (качественными).

При построении вариационного ряда с равными интервалами определяют его число групп (n) и величину интервала (h). Число групп можно определить с помощью различных формул. Оптимальное число групп может быть определено по формуле Стерджесса:

$$n = 1 + 3,322 \cdot \lg N , \quad (2)$$

где n – число групп;

N - число единиц совокупности.

В зависимости от исследовательских целей можно использовать равные и неравные интервалы (в последнем случае – равномерно возрастающие или убывающие) открытые и закрытые.

Величина равного интервала рассчитывается по формуле:

$$i = \frac{x_{\max} - x_{\min}}{n} \quad (3)$$

где i - длина интервала;

$x_{\max}$ – максимальное значение признака в совокупности;

$x_{\min}$ – минимальное значение признака в совокупности.

При проведении анализа вариационных рядов с неравными интервалами необходимо использовать показатель плотности распределения признака. Он рассчитывается как отношение частоты или частости каждого интервала к его величине.

Для вариационного ряда можно рассчитать накопленные частоты. По накопленным частотам можно судить о том, какое число единиц в совокупности имеет значение признака не выше того значения, которое соответствует выбранной величине накопленной частоты.

Группировка - это процесс образования групп единиц совокупности однородных в каком-либо отношении, а также имеющих одинаковые или близкие значения группировочного признака.

Группировки бывают простые и комбинационные. Простая группировка образуется по одному признаку, комбинационная - по двум и более признакам.

Основные задачи метода группировок сводятся к выделению основных типов явлений, определению структуры совокупности, изучению взаимосвязи признаков. В зависимости от цели и задач исследования различают следующие виды группировок: *типологические, структурные, аналитические*.

Типологическим группировкам относят все группировки, которые характеризуют качественные особенности и различия между типами явлений. Типологические группировки широко применяются в экономических, социальных и других исследованиях.

Структурная группировка - выявляет состав однородной в качественном отношении совокупности по какому-либо признаку. Примером могут служить группировки предприятий по проценту выполнения плана, по числу рабочих и т.д.

Аналитическая группировка - применяется для исследования взаимосвязи между явлениями. Используя аналитические группировки, определяют факторные и результативные признаки изучаемых явлений.

Статистические данные должны быть представлены так, чтобы ими было удобно пользоваться. Существуют, по крайней мере три способа представления

данных: они могут быть включены в текст, в таблицы или выражены графически.

Статистические таблицы - это наиболее рациональная форма представления результатов статистической сводки и группировки.

При построении статистических таблиц следует четко разграничивать статистическое подлежащее и статистическое сказуемое. Статистическое подлежащее располагают, как правило, в строках, статистическое сказуемое - в графах таблицы.

В зависимости от строения подлежащего различают три вида таблиц: простые, групповые, комбинационные.

Простые (перечневые) таблицы в подлежащем содержат перечень рассматриваемых объектов.

Групповые таблицы в подлежащем содержат группировку единиц изучаемого объекта, образованную по какому-либо одному признаку.

Комбинационные таблицы в подлежащем содержат группировку единиц, образованную по двум и более признакам.

Подлежащее таблицы - это объект нашего изучения (название района, города, предприятия).

Сказуемое - это система показателей, которыми характеризуется объект изучения, т.е. подлежащее таблицы.

Правила построения и заполнения статистических таблиц. При построении и заполнении статистических таблиц рекомендуется придерживаться определенных правил:

1. Следует избегать построения сложных всеобъемлющих таблиц. Вместо одной сложной таблицы лучше составить две или несколько более простых, взаимосвязанных таблиц.

2. Заголовок таблицы, тексты подлежащего (боковика) и сказуемого (шапки) должны быть сформулированы четко и кратко. Обязательно должно быть указано место и время, к которым относятся данные, а также единица измерения.

3. Показатели подлежащего и сказуемого должны располагаться в определенной логической последовательности.

4. Если число показателей подлежащего и сказуемого таблицы велико, то строки и графы таблицы следует пронумеровать. При этом графы подлежащего обозначают буквами («А», «Б» и т.д.), а сказуемого - цифрами.

5. Таблица не должна иметь пустых незаполненных клеток. Если сведений нет, то проставляется многоточие (...) или пишется «нет сведений». Случаи отсутствия явления отмечаются дефисом (-). Если клетки поля данных не подлежат заполнению, то ставится крест (х).

6. Данные граф и строк должны приводиться с одинаковой степенью точности. Если числовые значения меньше принятой в таблице точности, то проставляется 0,0.

7. Таблицы могут иметь сноски, в которых приводятся примечания, дающие пояснения к показателям, а также источники данных.

8. Таблицы, как правило, должны быть замкнутыми, т.е. иметь итоги по группам, подгруппам («всего») и в целом по таблице («итого»).

Пример 1. Имеются данные о сумме активов и кредитных вложений 20 коммерческих банков

Таблица 1 - Сумма активов и кредитные вложения коммерческих банков

№ банка	Кредитные вложения, млрд. тнг.	Сумма активов, млрд. тнг.
1	311	518
2	658	1194
3	2496	3176
4	1319	1997
5	783	2941
6	1962	3066
7	1142	1865
8	382	602
9	853	1304
10	2439	4991
11	3900	6728
12	305	497
13	799	1732
14	914	2002
15	1039	2295
16	2822	5636
17	1589	2998
18	1012	1116
19	1350	2482
20	3500	6453

С целью изучения зависимости суммы активов и кредитных вложений коммерческих банков произведите группировку банков по кредитным вложениям (факторный признак), образовав 5 групп с равными интервалами.

По каждой группе и совокупности банков подсчитайте:

- число банков;
- кредитные вложения – всего и в среднем на один банк;
- сумму активов – всего и в среднем на один банк.

Результаты представьте в виде групповой таблицы.

Сделайте краткие выводы.

Решение. Определим величину интервала группировки банков по кредитным вложениям:

$$i = \frac{x_{\max} - x_{\min}}{5} = \frac{3900 - 305}{5} = 719 \text{ млрд. тнг.,}$$

где $x_{\max}$, $x_{\min}$ – максимальное и минимальное значения кредитных вложений.

Таблица 2 - Расчетная таблица

Группа	Величины кредитных вложений в группе, млрд. тнг.	№ банка	Кредитные вложения, млрд. тнг	Сумма активов, млрд. тнг.
1	305 - 1024	12	305	497
		1	311	518
		8	382	602
		2	658	1194
		5	783	2941
		13	799	1732
		9	853	1304
		14	914	2002
		18	1012	1116
Итого		9	6017	11906
2	1024 - 1743	15	1039	2295
		7	1142	1865
		4	1319	1997
		19	1350	2482
		17	1589	2998
Итого		5	6439	11637
3	1743 - 2462	6	1962	3066
		10	2439	4991
Итого		2	4401	8057
4	2462 - 3181	3	2496	3176
		16	2822	5636
Итого		2	5318	8812
5	3181 - 3900	20	3500	6453
		11	3900	6728
Итого		2	7400	13181
Всего		20	29575	53593

Определим интервалы групп (x_i, x_{i+1}) табл. 2:

1 группа: 305 – 1024 млрд. тнг.;

2 группа: 1024 – 1743 млрд. тнг.;

3 группа: 1743 – 2462 млрд. тнг.;

4 группа: 2462 – 3181 млрд. тнг.;

5 группа: 3181 – 3900 млрд. тнг.,

где $x_1 = x_{\min} = 305$ млрд. тнг.;

$$x_2 = x_1 + i = 305 + 719 = 1024 \text{ млрд. тнг.};$$

$$x_3 = x_2 + i = 1743 \text{ млрд. тнг.};$$

$$x_4 = x_3 + i = 2462 \text{ млрд. тнг.};$$

$$x_5 = x_4 + i = 3181 \text{ млрд. тнг.};$$

$$x_6 = x_5 + i = 3900 \text{ млрд. тнг.}$$

Далее упорядочим исходную таблицу по возрастанию кредитных вложений и выделим группы, в которые попадут банки:

На основе полученной таблицы определим требуемые показатели. Результаты представим в виде групповой таблицы таблицы 3:

Таблица 3 - Групповая таблица

Группа	Количество банков в группе, шт.	Величины кредитных вложений в группе, млрд. тнг.	Кредитные вложения, млрд. тнг		Сумма активов, млрд. тнг.	
			Всего	В среднем на один банк	Всего	В среднем на один банк
1	9	305 - 1024	6017	668,556	11906	1322,889
2	5	1024 - 1743	6439	1287,8	11637	2327,4
3	2	1743 - 2462	4401	2200,5	8057	4028,5
4	2	2462 - 3181	5318	2659	8812	4406
5	2	3181 - 3900	7400	3700	13181	6590,5

Вывод: По результатам группировки можно сформулировать следующее: около 50 % рассматриваемых банков имеют кредитные вложения до 1000 млрд. тнг. Сумма активов возрастает в зависимости от роста величины кредитных вложений.

Пример 2. Известны следующие данные о количественном составе 50 семей. Построить вариационный ряд распределения.

1	6	3	2	2	1	2	1	3	3
3	2	3	1	5	2	4	5	4	4
6	4	2	3	5	3	3	2	3	2
6	1	6	2	3	3	3	3	1	2
5	4	5	1	3	3	2	2	2	1

Решение. Для получения представление о распределении семей по числу членов, следует построить вариационный ряд.

Признак, по которому будет строиться ряд распределения, - дискретный (число членов в семье). Построим дискретный вариационный ряд: для этого выпишем все значения признака (число членов в семье) в порядке возрастания и подсчитаем число семей по каждой группе. Число членов семьи – это вариант варьирующего признака (x_i), число семей – частота вариантов (f_i), частость (w_i) табл.4.

Таблица 4 - Число членов семьи

Число членов семьи	Число семей	Число семей, %
(x_i)	(f_i)	(w_i)
1	8	16
2	13	26
3	15	30
4	5	10
5	5	10
6	4	8
Итого	50	100

Упорядоченные данные позволяют судить о распределении семей по числу членов.

Пример 3. Имеется следующее распределение заводов по стоимости основных фондов:

Стоимость основных фондов, млн. тенге	1-3	3-5	5-10	10-30	30-50	Всего
Распределение заводов, %	4	14	16	52	14	100

Используя метод вторичной группировки, образуйте следующие группы заводов по стоимости основных фондов: 1-5, 5-10, 10-20, 20-30, 30-40, свыше 40.

Решение. При формировании новых групп могут быть использованы два способа: *перегруппировки и укрупнения*. В нашем случае необходимо использовать оба способа.

Так, новая группа по стоимости основных фондов в пределах от 1 до 5 млн. тенге образуется путем *укрупнения* (объединения) первых двух групп первичной группировки. Вторая группа вторичной группы совпадает с третьей группой первичной группировки. Третья и четвертая группа вторичной группировки образуется путем деления пополам четвертой группы первичной группировки. Аналогичные действия необходимо произвести для образования пятой и шестой групп вторичной группировки. В результате мы получим следующую вторичную группировку:

Стоимость основных фондов, млн. тенге	1-5	5-10	10-20	20-30	30-40	Свыше 40
Распределение заводов, %	4+14	16	52/2	52/2	14/2	14/2

Пример 4. Постройте интервальный ряд и гистограмму распределения по следующим данным, имеющимся в отделении Народного банка, об остатках на текущих счетах отдельных организаций на конец месяца, тыс. тенге:

911	692	1396	1028	563	926	1295	866	1190	1173
478	782	695	878	793	770	1140	917	933	1244
956	728	844	819	1296	892	744	611	963	926
519	1093	756	1070	1165	1278	912	576	869	623
417	1367	911	1079	1057	845	656	1089	916	913

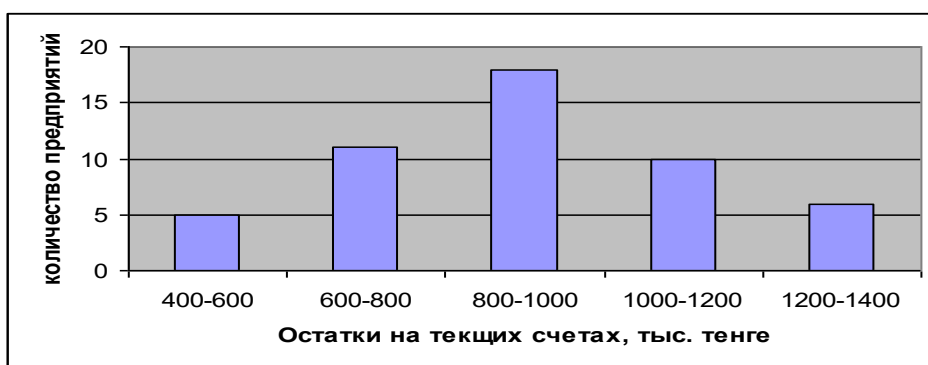
Ряд распределения постройте с равными интервалами. Необходимо образовать 5 групп.

Решение. Размах вариации (разность между максимальным и минимальным значением показателя) составляет $1396 - 417 = 979$ тыс. тенге. Исходя из числа групп интервал должен быть равен 194 тыс. тенге ($979/5$). В целях удобства интервал можно округлить до 200 тыс. тенге и образовать группы 400-600, 600-800, 800-1000, 1000-1200 и 1200-1400 тыс. тенге.

Сведем в таблицу количество предприятий, попадающих в каждую группу, а также подсчитаем процент предприятий, приходящихся на каждую группу:

Остатки на текущих счетах предприятий, тыс. тенге	Количество предприятий, единиц	В % к итогу,
400-600	5	10
600-800	11	22
800-1000	18	36
1000-1200	10	20
1200-1400	6	12
	50	100

Построим на основе этой таблицы гистограмму. Гистограмма распределения остатков на текущих счетах предприятий на конец месяца в Народного банка



Пример 5. При выборочном изучении покупательского спроса в обувном магазине была зарегистрирована продажа следующих размеров обуви:

34	33	34	34	33	33	31	33	31	32
32	34	33	32	34	31	33	31	34	31
33	31	32	31	32	32	34	34	32	34
34	33	34	34	31	33	31	34	31	32
32	32	33	32	32	34	33	33	34	33

Постройте ряд и полигон распределения и проанализируйте полученные результаты, сравнив их с типовой шкалой производства школьной обуви:

Типовая шкала производства школьной обуви

Размер обуви	34	33	32	31	Всего
В % к итогу	32	30	20	18	100

Решение. Ряд является атрибутивным, поскольку варьирующий признак является качественным. Количество значений варьирующего признака равно четырем, следовательно, нужно произвести сводку первичного материала в 4 группы:

Размер обуви	Количество проданных пар	В процентах к итогу, %
31	10	20
32	12	24
33	13	26
34	15	30
	50	100

На основе этой таблицы строится полигон распределения.

В сравнении с типовой шкалой производства школьной обуви результаты

Размер обуви	34	33	32	31	Всего
В % к итогу по типовой шкале производства	32	30	20	18	100
В % к итогу по фактической продаже	30	26	24	20	100
Расхождение по сравнению с типовой шкалой, процентных пунктов	.2	-4	+4	+2	

Как видно из таблицы, наибольшее отклонение наблюдается для размеров обуви 33 и 34, для которых отклонение составило по 4 процентных пункта, в одном случае в меньшую сторону (для 33 размера), в другом - в большую (для 32 размера).

1.2 Задачи и упражнения для самостоятельного решения

1. Имеется следующее распределение малых предприятий в промышленности по численности занятых:

Численность занятых, человек	до 5	6-10	11-15	16-35	36-50	Всего
Распределение предприятий, %	40	20	11	20	9	100

Используя метод вторичной группировки, образуйте следующие группы предприятий в промышленности по численности занятых: 1-5, 6-10, 11-20, 21-30, 31-40, 41-50.

2. Постройте интервальный ряд и гистограмму распределения роста детей в двух классах некоторой школы, см:

130	126	140	132	123	135	130	124	130	137
121	127	126	129	128	130	135	125	130	127
131	126	129	128	138	138	129	131	133	123
122	134	127	133	135	130	129	124	134	130
120	139	130	133	153	130	127	130	137	129

Ряд распределения построите с равными интервалами. Необходимо образовать 5 групп.

3. На предприятии работает 50 рабочих, имеющих следующие разряды:

6	5	6	6	5	5	3	5	4	4
4	6	5	4	6	3	5	3	6	3
5	3	4	3	4	4	6	6	4	6
6	5	6	6	3	5	3	6	3	4
4	4	5	4	4	6	5	5	6	5

Постройте ряд и полигон распределения и проанализируйте полученные результаты, сравнив их с требуемыми по технологии производства минимальными разрядами:

Разряд	3-й	4-й	5-й	6-й	Всего
в % к итогу	25	25	30	20	100

Контрольные вопросы

1. Охарактеризуйте содержание статистической сводки, ее составные элементы и способы организации.
2. Опишите роль группировки в статистике. Приведите классификацию признаков, положенных в основу группировки.
3. Объясните различие между номенклатурой и классификацией. Приведите положения, используемые при отборе группировочных признаков.
4. Охарактеризуйте основные виды и задачи группировок. Приведите примеры.
5. Приведите основные правила, применяемые при образовании групп по количественным признакам: количество групп и размер интервала.
6. Опишите отличие сложных группировок от простых, вторичных и первичных. Приведите их полную классификацию и назначение.
7. Охарактеризуйте содержание, построение и виды статистических рядов распределения.
8. Опишите различные виды графического представления рядов распределения.
9. Приведите назначение статистической таблицы, составные элементы и виды в зависимости от типов подлежащего и сказуемого таблицы.
10. Опишите правила построения и заполнения таблицы.

Тема 2. Абсолютные и относительные статистические величины

2.1 Методические указания и решения типовых задач

Первые этапы в проведении статистической работы заключаются в статистическом наблюдении, сводке и группировке статистических материалов. По итогам статистического наблюдения получают статистические материалы, где каждая единица изучаемой совокупности характеризуется рядом признаков. При статистической сводке и группировке собранные первичные материалы обобщают в сводные статистические таблицы, в которых *совокупности единиц представлены в целом и по группам*. Показатели, характеризующие их, называются *обобщающими показателями*.

Обобщение - важная задача статистики. Метод обобщающих показателей как результат обобщения является характерным, специфическим для статистики методом, как и методы группировок и массовых наблюдений. Обобщающие показатели в статистике могут быть **абсолютными, относительными и средними**.

Абсолютные показатели получают непосредственно в результате сводки (суммирования) первичного статистического материала. На их основе исчисляют относительные и средние величины, которые их дополняют.

Для характеристики того или иного явления часто применяют все три вида обобщающих показателей. Например, при изучении промышленности предприятия группируют по отраслям. В итоге получают сводную статистическую таблицу согласно представленному ниже макету.

Таблица 5 - Макет статистической таблицы для характеристики предприятий по основным отраслям промышленности

Отрасли промышленности	Число предприятий, единиц	Стоимость продукции, млн. тенге	Удельный вес		Средняя стоимость продукции на 1 предприятие, тыс. тенге
			по числу предприятий, %	по стоимости продукции. %	
Добывающая промышленность					
Обрабатывающая промышленность					
Производство электроэнергии, газа и воды					
Итого			100	100	

Число промышленных предприятий и стоимость произведенной ими продукции представляют обобщающие *абсолютные величины*. Удельный вес отдельных отраслей промышленности в общем количестве предприятий или стоимости продукции называются *относительными величинами*. Средняя

стоимость продукции на 1 предприятие является *средней величиной*. В целом все эти показатели дают разностороннюю характеристику промышленности.

Значение и виды абсолютных величин. Абсолютные статистические величины имеют большое значение в анализе любого явления и принятии решений. Они дают представление о масштабе этого явления и выражаются в определенных единицах измерения: килограммах, метрах, штуках, гектарах, кубометрах, литрах, тенге, долларах, евро и т.д.

Абсолютные величины характеризуют численность занятых в экономике, величину произведенной добавленной стоимости, объем внешнеторгового оборота и т.д. Как обобщающие показатели в статистике они всегда являются *суммарными величинами*. Их можно классифицировать по нескольким признакам:

1) по признаку характеристики совокупности в составе абсолютных величин — *показатели численности совокупности* (число предприятий, число рабочих, населения) и *показатели объемов признака* (заработная плата рабочих, продукция предприятий и т.д.). При этом один и тот же показатель может быть в одном случае показателем численности совокупности, в другом - показателем объема признака, например, численность рабочих на предприятии;

2) по признаку характеристики процесса развития - показатели, характеризующие либо состояние явления на определенный момент - *моментные показатели* (численность населения, рабочих, предприятий, машин и оборудования, скота и т.д.), либо результаты процессов за определенный период - *интервальные показатели* (производство продукции, затрат труда и материалов, прироста или уменьшения численности населения и т.д.);

3) по единице измерения, выражающей масштабы явления — *натуральные и стоимостные показатели*.

Натуральные показатели могут быть *простыми* (например, мощность электростанций выражается в киловаттах, а объем перевезенных грузов - в тоннах), и *составными* (количество выработанной электроэнергии выражается в киловатт-часах, а грузооборот - в тонно-километрах).

Применяют также *условные единицы измерения* (например, разные виды топлива пересчитывают в тонны условного топлива, для которого принята теплота сгорания на уровне 7000 ккал/кг, а разнообразные по емкости консервные банки переводят в условные банки с емкостью 353,4 куб. см).

Стоимостные показатели (объем произведенной или реализованной продукции, себестоимость продукции, затраты) отражаются как *в текущих ценах*, так и *в сопоставимых ценах*.

Расчетные абсолютные величины. Ряд абсолютных показателей получают не в результате сводки, а расчетным путем с учетом дополнительных данных (например, объемы производства с учетом не отчитавшихся предприятий, скрытой деятельности и неформального сектора). Широко используется метод расчета недостающего показателя на основе балансовых расчетов. Часто объем признака рассчитывают на основе среднего его значения и численности совокупности (например, путем перемножения среднего веса

мешка картофеля на количество мешков можно определить совокупный вес убранного картофеля).

Понятие об относительных величинах и формах их выражения. При анализе статистических материалов на первое место выдвигаются относительные и средние величины. Например, для сравнения уровня жизни в различных странах нельзя использовать только абсолютные показатели об объеме производства ВВП или количестве автомобилей в стране в целом. Необходимо вычислять показатели на душу населения или на 100 человек.

Таким образом, расчеты относительных величин представляют первый шаг на пути статистического анализа.

Относительные величины получают в результате сравнения двух показателей. Так, если Законом о республиканском бюджете определена величина доходной части бюджета в размере 500 млрд. тенге, а фактически доходы составили 525 млрд. тенге, то это означает, что Министерство финансов перевыполнило план по доходной части бюджета на 5%.

Знаменатель относительной величины называется *основанием*, или *базой сравнения*. В нашем случае основание - 300 млрд тенге. Если основание принять за единицу, то относительная величина выражается в форме *коэффициента* (1,05); она покажет, во сколько раз сравниваемая величина больше или меньше основания. Если основание принимается за 100, то относительная единица выражается в *процентах*, если за 1000, то в *промилле*. Для отдельных, редких явлений в статистике также используют относительные показатели в расчете на 10 000 или на 100 000.

Если из показателя, выраженного в процентах, вычесть 100, то можно сказать, на сколько процентов сравниваемая величина больше или меньше основания.

Промилле широко распространен в демографической статистике, где рождаемость, смертность и ряд других показателей вычисляются в расчете на 1000 населения.

Обычно сравниваются два абсолютных показателя, но можно сравнивать относительные и средние показатели. *Относительная величина* - это *обобщающий показатель, дающий числовую меру соотношения двух сопоставляемых статистических величин*.

При этом сравниваемые показатели должны быть методологически сопоставимы.

В зависимости от познавательной сущности различают относительные показатели динамики, планового задания, выполнения плана, структуры, координации, сравнения и интенсивности развития.

Показатели динамики характеризуют степень изменения явления во времени. При этом величину одного из прошлых периодов (моментов) времени, называемого базисным, принимают за 100 или единицу, а показатели последующих периодов выражаются в процентах или коэффициентах к базисному. Они называются также *темпами роста*, а за вычетом единицы или 100 - *темпами прироста*.

Показатели планового задания показывают соотношение величины планируемого показателя к фактически достигнутому уровню. При этом сравниваемые показатели относятся к разным периодам времени. Например, если индикативным планом было предусмотрено увеличение темпов прироста промышленного производства с 8 до 10%, а сокращение дефицита бюджета с 3% к ВВП до 2,5%, то это означает что план по увеличению темпов прироста промышленного производства составляет 125% ($10:8 * 100$), а по сокращению дефицита бюджета составляет 83,3% ($2,5 : 3,0 * 100$).

Показатели выполнения плана в отличие от предыдущего вида относительных величин, наоборот, отображают соотношение величины фактически достигнутого уровня по сравнению с запланированным уровнем. При этом, однако, показатели относятся к одному периоду (моменту) времени. Например, если для предыдущего примера фактический темп прироста промышленного производства составил не 10, а 12%, а сокращение дефицита бюджета составило не 0,5% (с 3% к ВВП до 2,5%), а 0,6% (т.е. фактически дефицит составил 2,4%), то это означает что план по росту промышленного производства перевыполнен на 20% ($12:10 \times 100$), а по сокращению дефицита бюджета также перевыполнен на 20% ($0,6 : 0,5 * 100$).

Показатели структуры представляют долю (или удельный вес) части в целом, выраженную в процентах.

Например, за переходный период доля сельского хозяйства в общем объеме ВВП сократилась с 34% (в 1990 г.) до 8% (в 2004 г.). Сокращение удельного веса при этом может не означать соответствующего уменьшения абсолютного размера данной части совокупности. Так, сокращение доли сельского хозяйства в ВВП более чем в 4 раза сопровождалось сокращением абсолютных объемов производства только на 39%. В то же время, несмотря на то что промышленность увеличила свою долю в ВВП почти в полтора раза (с 21 до 31 %), ее реальный объем, тем не менее, сократился на 14%.

Показатели координации характеризуют соотношение одноименных величин отдельных частей целого между собой, одна из которых принята за базу сравнения. Иными словами, если в показателе структуры в качестве базы сравнения принимался общий итог, то в данном случае в качестве базы сравнения может быть принята любая часть целого. Допустим, на предприятии из 1000 рабочих 900 являются производственными рабочими, а 100 — непроизводственными. Тогда с точки зрения структуры можно сказать, что 90% рабочих предприятия составляют производственные, а 10%-непроизводственные. С позиции же сравнения количества производственных и непроизводственных рабочих отметим, что производственных рабочих в 9 раз больше непроизводственных. Аналогично в демографической статистике: доля мужчин и женщин в общей численности населения составляет соответственно 48 и 52%, в то же время на 1000 женщин приходится только 923 мужчин.

Показатели сравнения характеризуют сравнение одноименных показателей, относящихся к различным объектам или территориям, но за один и тот же период (момент) времени. Тем самым, мы можем сравнить численность

населения или объемы промышленного производства различных регионов страны или провести межстрановые сравнения. Например, численность населения Российской Федерации примерно в десять раз превышает численность населения Казахстана, а сравнение по суммарному объему производимой за год добавленной стоимости показывает примерно пятнадцатикратное различие. Это свидетельствует о более высоком уровне производительности труда в России, нежели в Казахстане.

Показатели интенсивности развития отображают степень распространенности данного явления в определенной среде. Сравнивая совокупности родившихся или умерших с численностью населения, мы тем самым оцениваем степень, интенсивность развития данного явления. К другим показателям интенсивности относятся показатели, характеризующие рост материального и культурного уровня населения (ВВП на душу населения, уровень образования и др.). К ним относятся также показатели плотности населения, густоты железнодорожной сети и др.

Можно вычислять показатели производства на душу населения. Широко используются показатели интенсивности и для характеристики степени совершенства производства, оснащения его новой техникой, характеристики использования оборудования, эффективности капитальных вложений.

В отличие от других относительных величин показатели интенсивности развития не отвлеченные, а именованные числа: они всегда выражают количество единиц совокупности, которая стоит в числителе отношения, на единицу той совокупности, которая стоит в знаменателе.

Пример 1. Расход топлива на производственные нужды предприятия характеризуется в отчетном периоде следующими данными:

Вид топлива	Теплотворная способность, МДж/кГ	Расход, т	
		по плану	фактически
Дизельное топливо	41,9	1000	1050
Мазут	40,1	750	730
Уголь	26,4	500	555

Определить общее количество потребленного условного топлива (1 т.у.т. = 29,3 МДж/кГ) по плану и фактически, а также процент выполнения плана по общему расходу топлива.

Решение. Учитывая стандартную теплотворную способность 29,3 МДж/кГ, определяем количество потребленного условного топлива каждого вида по плану (X'_{li}) и фактически (X_{li}):

- дизельное топливо: $X'_{ldm} = 41,9/29,3 \cdot 1000 = 1430,034$ т.у.т.
- дизельное топливо: $X_{ldm} = 41,9/29,3 \cdot 1050 = 1501,536$ т.у.т.;
- мазут: $X'_{lm} = 40,1/29,3 \cdot 750 = 1026,451$ т.у.т.
- мазут: $X_{lm} = 40,1/29,3 \cdot 730 = 999,078$ т.у.т.;
- уголь: $X'_{ly} = 26,4/29,3 \cdot 500 = 450,512$ т.у.т.
- уголь: $X_{ly} = 26,4/29,3 \cdot 555 = 500,068$ т.у.т.

Суммируя количество потребленного условного топлива каждого вида, получим общее количество потребленного условного топлива:

- по плану $X'_I = \sum X'_{iI} = 2906,997$ т.у.т.;
- фактически $X_I = \sum X_{iI} = 3000,682$ т.у.т.

Для определения процента выполнения плана необходимо рассчитать *индекс выполнения плана*, то есть отношение значений по факту и плану отчетного периода:

$$i_{ВП} = \frac{X_I}{X'_I}, \quad (4)$$

Применяя формулу (4), имеем: $i_{ВП} = 3000,682/2906,997 = 1,032$, то есть план по общему расходу топлива перевыполнен на 3,2%.

Пример 2. Рассчитать индекс и темп изменения, если в марте произведено продукции 130 тонн, а в феврале 100 тонн.

Решение. *Индекс изменения (динамики)* характеризует изменение какого-либо явления во времени. Он представляет собой отношение значений одной и той же абсолютной величины в разные периоды времени. Данный индекс определяется по формуле (5):

$$i_d = \frac{X_1}{X_0}, \quad (5)$$

где подиндексы означают: 1 — отчетный или анализируемый период, 0 — прошлый или базисный период.

Критериальным значением индекса динамики (темпа роста) служит единица, то есть если $i_d > 1$, то имеет место рост явления во времени; если $i_d = 1$ — стабильность; если $i_d < 1$ — наблюдается спад явления. Применяя формулу (5), имеем: $i_d = 130/100 = 1,3$ (или 130%) > 1 — рост объема произведенной продукции.

Темп изменения (прироста) определяется по формуле (6):

$$T = i_d - 1. \quad (6)$$

Применяя формулу (6), имеем: $T = 1,3 - 1 = 0,3$ (или 30%), то есть объем произведенной продукции вырос в марте по сравнению с февралем на 30%.

Пример 3. Рассчитать индексы планового задания, выполнения плана и динамики, если выпуск продукции в отчетном году составил 100 млн. тенге, на следующий год планировалось 140 млн. тенге, а фактически получено 112 млн. тенге.

Решение. *Индекс планового задания* — это отношение значений одной и той же абсолютной величины по плану анализируемого периода и по факту базисного. Он определяется по формуле (7):

$$i_{пз} = \frac{X'_1}{X_0}, \quad (7)$$

где X'_1 — план анализируемого периода; X_0 — факт базисного периода.

Применяя формулу (7) имеем: $i_{пз} = 140/100 = 1,4$ (или 140%), то есть на следующий год планировалось выпустить продукции в размере 140% от объема предыдущего года.

Индекс выполнения плана определим, применяя формулу (4): $i_{вп} = 112/140 = 0,8$ (или 80%), то есть план по увеличению выпуска продукции выполнили лишь на 80% или недовыполнили на 20%.

Индекс динамики можно определить по формуле (5) или перемножая индексы планового задания и выполнения плана, то есть $i_d = \frac{X_1}{X_0} = i_{пз} i_{вп} = 1,12$.

Пример 4. Суммарные денежные доходы в 2015 г. составили 13522,5 млрд. тенге, из которых 8766,7 млрд. тенге составила оплата труда, 1748,4 млрд. тенге — социальные выплаты, 1541,7 млрд. тенге — доход от предпринимательской деятельности, 1201,5 млрд. тенге — доходы от собственности, остальное — прочие доходы. Рассчитать относительные величины структуры и координации, приняв за основу оплату труда. Построить секторную (круговую) диаграмму структуры доходов.

Решение. Индекс структуры (доля) — это отношение какой-либо части величины (совокупности) ко всему ее значению. Он определяется по формуле (8):

$$i_{ст} = d = \frac{f}{\sum f} \quad (8)$$

Применяя формулу (8) и округляя значения до 3-х знаков после запятой, имеем:

- доля оплаты труда $d_{от} = 8766,7/13522,5 = 0,648$ или 64,8%;
- доля социальных выплат $d_{св} = 1748,4/13522,5 = 0,129$ или 12,9%;
- доля доходов от предпринимательской деятельности $d_{пд} = 1541,7/13522,5 = 0,114$ или 11,4%;
- доля доходов от собственности $d_{дс} = 1201,5/13522,5 = 0,089$ или 8,9%.

Долю прочих доходов найдем, используя формулу (9), согласно которой сумма всех долей равна единице:

$$\sum d = 1. \quad (9)$$

Таким образом, доля прочих доходов $d_{проч} = 1 - 0,648 - 0,129 - 0,114 - 0,089 = 0,020$ или 2,0%.

Для иллюстрации структуры (составных частей) доходов построим секторную диаграмму:

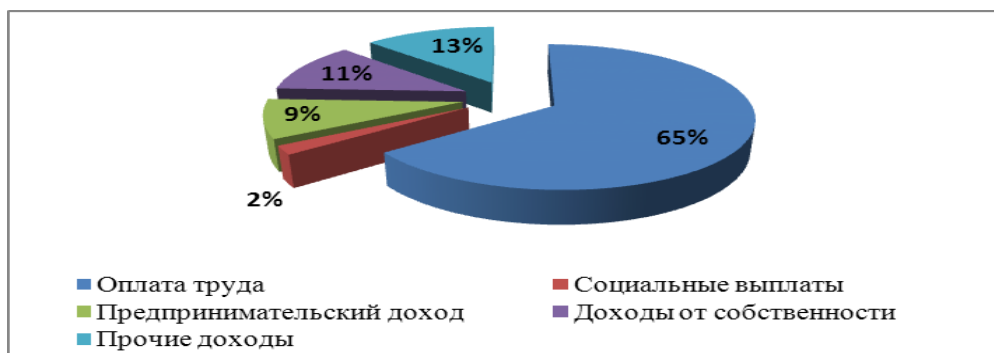


Рисунок 1. Структура денежных доходов населения РК в 2015 году.

Таким образом, очевидно, что наибольшую долю в суммарных денежных доходах составляет оплата труда (64,8%), на 2-м месте – социальные выплаты (12,9%), затем следуют предпринимательский доход (11,4%), доходы от собственности (8,9%), а прочие доходы составляют лишь 2%.

Индекс координации – это отношение какой-либо части величины к другой ее части, принятой за основу (базу сравнения). Он определяется по формуле (10):

$$i_K = \frac{f}{f_0}. \quad (10)$$

Применяя формулу (10) и принимая за основу оплату труда, имеем:

– индекс координации социальных выплат $i_{K_{CB}} = 1748,4/8766,7 \approx 0,129/0,648 = 0,199$;

– индекс координации предпринимательского дохода $i_{K_{ПД}} = 1541,7/8766,7 \approx 0,114/0,648 = 0,176$;

– индекс координации доходов от собственности $i_{K_{ДС}} = 1201,5/8766,7 \approx 0,089/0,648 = 0,137$;

– индекс координации прочих доходов $i_{K_{проч}} \approx 0,02/0,648 = 0,031$.

Таким образом, социальные выплаты составляют 19,9% от оплаты труда, предпринимательский доход – 17,6%, доходы от собственности – 13,7%, а прочие доходы – 3,1%.

Пример 5. Запасы воды в озере Байкал составляют 23000 км³, а в Ладожском озере 911 км³. Рассчитать относительные величины сравнения запасов воды этих озер.

Решение. *Индекс сравнения* – это отношение значений одной и той же величины в одном периоде или моменте времени, но для разных объектов или территорий. Он определяется по формуле (11):

$$i_C = \frac{X_A}{X_B}, \quad (11)$$

где A, B — признаки сравниваемых объектов или территорий.

Применяя формулу (11) и принимая за объекты A и B , соответственно, озера Байкал и Ладожское, найдем индекс сравнения: $i_c = 23000/911 = 25,25$, то есть запасов воды в озере Байкал в 25,25 раза больше, чем в Ладожском озере.

Меняя базу сравнения, найдем индекс сравнения Ладожского озера с Байкалом по той же формуле: $i_c = 911/23000 = 0,0396$ или 3,96%, то есть запасы воды в Ладожском озере составляют 3,96% запасов воды в озере Байкал.

Пример 6. Рассчитать относительную величину интенсивности валового внутреннего продукта (ВВП) в сумме 131,9 млрд. \$ на душу населения в Казахстане в 2016 году при численности населения в 17,7 млн. человек..

Решение. Показатель интенсивности – это отношение значений двух разнородных абсолютных величин для одного периода времени и одной территории или объекта. Он определяется по формуле (12):

$$i_{ин} = \frac{X}{Y}. \quad (12)$$

Применяя формулу (12) имеем: $i_{ин} = 131,9/0,0177 = 7451,97$ \$/чел в год

2.2 Задачи и упражнения для самостоятельного решения

1. Определить общее производство моющих средств в условных тоннах (условная жирность 40%) по плану и фактически, а также процент выполнения плана по следующим данным:

Вид продукта	Жирность, %	Физическая масса, т	
		по плану	фактически
Мыло хозяйственное	60	500	600
Мыло туалетное	80	1000	1500
Стиральный порошок	10	50000	40000

2. По данным о численности жителей двух крупнейших городов Казахстана (тыс. чел) определить индексы сравнения и динамики.

Города	2015 г.	2016 г.
Астана	872,6	972,7
Алматы	1702,7	1751,3

3. По плану на 2015 год намечалось увеличение товарооборота на 3%. В 2015 году плановое задание перевыполнили на 600 млн. тенге или на 2,5%. Определить фактический прирост товарооборота (в млн. тенге) в 2015 году по сравнению с 2014 годом.

4. По данным о товарообороте из предыдущей задачи, состоящего из реализации собственной продукции и продажи покупных товаров, определить относительные величины координации и структуры собственной и покупной продукции в 2014 и 2015 годах, если известно, что доля собственной продукции в 2014 году составила 65%, а в 2015 году она увеличилась на 10%.

5. Жилищный фонд и численность населения Казахстана следующие (на начало года). Охарактеризовать изменение обеспеченности населения жилой площадью с помощью относительных величин динамики и координации.

Год	2011	2012	2013	2014
Весь жилищный фонд, млн. м ²	283,9	303,4	336,1	339,1
Численность населения, млн. чел.	16,56	16,79	17,04	17,29

6. 1. В Казахстане в 2017 численность женщин составила 9,503931 млн. чел, а мужчин – 8,784293 млн. чел. Рассчитать относительные величины структуры и координации.

7. По плану объем продукции в отчетном году должен возрасти по сравнению с прошлым годом на 2,5%. План выпуска продукции перевыполнен на 3,0%. Определить фактический выпуск продукции в отчетном году, если известно, что объем продукции в прошлом году составил 25,3 млн. тенге

8. Определить общий объем фактически выпущенной продукции по следующим данным по трем филиалам предприятия, выпускающих однородную продукцию:

Номер филиала	Планируемый объем выпуска продукции, млн. тенге	Выполнение намеченного плана, %
1	500	104
2	750	92
3	250	116

9. По промышленному предприятию за отчетный год имеются следующие данные о выпуске продукции:

Наименование продукции	План на I квартал, тыс. т	Фактический выпуск, тыс. т			Отпускная цена за 1 т, у.е.
		январь	февраль	март	
Сталь арматурная	335	110	115	108	1700
Прокат листовой	255	75	90	100	2080

Определить процент выполнения квартального плана: 1) по выпуску каждого вида продукции; 2) в целом по выпуску всей продукции.

10. Определить процент выполнения плана по продажам условных школьных тетрадей (1 у.ш.т. – 12 листов) по каждому виду тетрадей и в целом по магазину по следующим данным:

Вид тетради	Цена, тенге/шт.	Объем продаж, тыс. шт.	
		по плану	фактически
Тетрадь общая 90 листов	20	50	40
Тетрадь общая 48 листов	13	200	350
Тетрадь общая 16 листов	9	700	500

11. В Казахстане на начало 2017 года численность населения составила 18,288 млн. чел., в течение года: родилось 0,353 млн. чел., умерло – 0,139 млн. чел., мигрировало из других государств 0,339 млн. чел., мигрировало за границу – 0,295 млн. чел. Охарактеризовать изменение численности населения в 2017 году с помощью относительных величин.

12. Определить общий объем фактически выпущенной условной консервной продукции (1 у.к.б. = 0,33 л) по следующим данным:

Вид продукции	Планируемый объем выпуска продукции, тыс. шт.	Выполнение плана, %
Томатная паста 1 л	500	85
Томатная паста 0,5 л	750	104
Томатная паста 0,2 л	250	130

Контрольные вопросы

1. Какие показатели называются обобщающими? Как их получают?
2. Назовите виды обобщающих показателей. Приведите примеры.
3. Что называется абсолютными статистическими величинами и каково их значение? Приведите примеры.
4. Назовите классификации абсолютных величин. Приведите примеры.
5. В каких единицах измерения выражаются абсолютные статистические величины? Приведите примеры.
6. Что называется относительной величиной? Какова роль относительной величины в статистике?
7. Какие виды относительных величин вы знаете? Приведите примеры.
8. Чем отличаются показатели интенсивности от других относительных величин?
9. В каких единицах измерения выражаются относительные статистические величины? Приведите примеры.
10. В чем заключается значение статистических графиков?
11. Каковы основные элементы графиков?
12. Опишите основные виды статистических графиков.
13. В чем различие картограмм и картодиаграмм?

Тема 3. Средние величины и показатели вариации

3.1 Методические указания и решения типовых задач

Значение средней величины в статистике. Средняя величина является самым распространенным обобщающим показателем в статистике. Это связано с тем, что с ее помощью можно охарактеризовать совокупность по количественно варьирующему признаку. Например, для сравнения заработной платы рабочих двух предприятий не может быть взята заработная плата двух конкретных рабочих, поскольку она выступает варьирующим показателем.

Также не может быть взята общая сумма заработной платы, выплаченной на предприятиях, так как она зависит от количества работающих. Если же мы разделим общую сумму заработной платы каждого предприятия на численность работающих, то сможем их сравнить и определить, на каком предприятии средняя заработная плата выше иными словами заработная плата изучаемой совокупности рабочих получает обобщенную характеристику в средней величине. В ней выражается то общее и типичное, что характерно для совокупности рабочих в отношении изучаемого признака. Она в одной величине показывает общую меру этого признака, имеющего различное значение у единиц совокупности.

Определение средней величины. Средней величиной в статистике называется обобщенная характеристика совокупности однотипных явлений по какому-либо количественно варьирующему признаку. Средняя величина показывает уровень этого признака, отнесенный к единице совокупности.

С помощью средней величины можно сравнивать между собой разные совокупности по варьирующим признакам (доходы на душу населения, урожайность сельскохозяйственных культур, себестоимость производства продукции на различных предприятиях).

Средняя величина всегда обобщает количественную вариацию при котором мы характеризуем изучаемую совокупность и который в равной степени присущ всем единицам совокупности. Значит, за всякой средней величиной всегда скрывается ряд распределения единиц совокупности по какому-то варьирующему признаку, т.е. вариационный ряд.

В этом отношении средняя величина принципиально отличается от относительных величин и, в частности от показателей интенсивности. Показатель интенсивности - отношение объемов двух разных совокупностей (например, производство ВВП на душу населения), в то время как средняя - обобщает характеристику элементов совокупности по одному из признаков (например, средняя заработная плата рабочего).

Средняя величина и закон больших чисел. В изменении средних показателей проявляется общая тенденция, под влиянием которой складывается процесс развития явлений в целом, в отдельных же индивидуальных случаях эта тенденция может и не обнаруживаться явно. Важно, чтобы средние

величины были основаны на массовом обобщении фактов. Только при этом условии они выявят общую тенденцию, лежащую в основе процесса в целом.

Во все более полном погашении отклонений, порождаемых случайными причинами, по мере увеличения числа наблюдений проявляется сущность закона больших чисел и его значение для средних величин. То есть закон больших чисел создает условия, чтобы в средней величине проявился типичный уровень варьирующего признака в конкретных условиях места и времени. Величина этого уровня определяется сущностью явления.

Виды средних величин. Средние величины, применяемые в статистике, относятся к классу степенных средних, общая формула которых имеет следующий вид:

$$\bar{x} = \sqrt[m]{\frac{\sum x^m}{n}}, \quad (12)$$

где $\bar{x}$ - степенная средняя;

x - меняющиеся величины признака (варианты);

n - число вариантов;

m - показатель степени средней;

$\sum$ - знак суммирования.

При различных значениях показателя степени средней (m) получаются различные виды средней величины:

$m = 1$ - средняя арифметическая;

$m = 2$ - средняя квадратическая;

$m = 3$ - средняя кубическая;

$m = -1$ - средняя гармоническая;

$m = 0$ - средняя геометрическая (после преобразований).

Следует иметь в виду, что различные виды средней величины имеют разные значения при использовании одних и тех же исходных статистических материалов. При этом, чем больше показатель степени средней, тем выше ее величина (правило мажорантности средних).

В статистике правильную характеристику совокупности в каждом отдельном случае дает только вполне определенный вид средних величин. Для определения этого вида средней величины используется критерий, определяющий свойства средней: средняя величина только тогда будет верной обобщающей характеристикой совокупности по варьирующему признаку, когда при замене всех вариантов средней величиной общий объем варьирующего признака остается неизменным. То есть правильный вид средней определяется тем, как образуется общий объем варьирующего признака.

Так, средняя арифметическая применяется тогда, когда объем варьирующего признака образуется как сумма отдельных вариантов, средняя квадратическая - когда объем варьирующего признака образуется как сумма квадратов, средняя гармоническая - как сумма обратных значений отдельных вариантов, средняя геометрическая - как произведение отдельных вариантов.

Кроме средних величин в статистике применяют описательные характеристики распределения варьирующего признака (структурные средние): моду (наиболее часто встречающаяся варианта) и медиану (серединная варианта).

Средняя величина – это обобщающий показатель совокупности, характеризующий уровень изучаемого явления или процесса. Средние величины могут быть *простыми* и *взвешенными*. Простая средняя рассчитывается при наличии двух и более статистических величин, расположенных в произвольном (несгруппированном) порядке, по общей формуле (1). Взвешенная средняя величина рассчитывается по сгруппированным статистическим величинам с использованием общей формулы (24).

$$\bar{X} = \sqrt[m]{\frac{\sum X_i^m}{N}}; \quad (1)$$

$$\bar{X} = \sqrt[m]{\frac{\sum X_i^m f_i}{\sum f_i}}. \quad (24)$$

При этом обозначено: X_i – значения отдельных статистических величин или середин группировочных интервалов; m - показатель степени, от значения которого зависят *виды* средних величин. Используя формулы (1) и (24) при разных показателях степени m , получаем частные формулы каждого вида (см. таблицу 6).

Таблица 6 - Виды степенных средних и их применение

m	Название средней	Формула расчета средних		Когда применяется
		простая	взвешенная	
1	Арифметическая	$\bar{X}_{ap} = \frac{\sum X_i}{N}$	$\bar{X}_{ap} = \frac{\sum X_i f_i}{\sum f_i}$	Чаще всего, кроме тех случаев, когда должны применяться другие виды средних
-1	Гармоническая	$\bar{X}_{gm} = \frac{N}{\sum \frac{1}{X_i}}$	$\bar{X}_{gm} = \frac{\sum f_i}{\sum \frac{f_i}{X_i}}$	Для осреднения величин с дробной размерностью при наличии дополнительных данных по числителю дробной размерности
0	Геометрическая	$\bar{X}_{geom} = \sqrt[N]{\prod_{i=1}^N X_i}$	$\bar{X}_{geom} = \sqrt[N]{\prod_{i=1}^N X_i^{f_i}}$	Для осреднения цепных индексов динамики
2	Квадратическая	$\bar{X}_{kv} = \sqrt{\frac{\sum X_i^2}{N}}$	$\bar{X}_{kv} = \sqrt{\frac{\sum X_i^2 f_i}{\sum f_i}}$	Для осреднения вариации признака (расчет средних отклонений)
3	Кубическая	$\bar{X}_{куб} = \sqrt[3]{\frac{\sum X_i^3}{N}}$	$\bar{X}_{куб} = \sqrt[3]{\frac{\sum X_i^3 f_i}{\sum f_i}}$	Для расчета индексов нищеты населения
1	Хронологическая	$\bar{X}_{xp} = \frac{\frac{X_1 + X_N}{2} + \sum_{i=2}^{N-1} X_i}{N-1}$	$\bar{X}_{xp} = \frac{\sum (X_i + X_{i+1}) f_i}{2 \sum f_i}$	Для осреднения моментных статистических величин

Выбор вида формулы средней величины зависит от содержания осредняемого признака и конкретных данных, по которым ее приходится вычислять. Показатель степени m в общей формуле средней величины оказывает существенное влияние на значение средней величины: по мере увеличения степени возрастает и средняя величина (правило мажорантности

средних величин), то есть $\bar{X}_{ГМ} < \bar{X}_{геом} < \bar{X}_{ар} < \bar{X}_{KB} < \bar{X}_{куб}$. Так, если $m \rightarrow +\infty$, то $\bar{X} \rightarrow X_{\max}$, а если $m \rightarrow -\infty$, то $\bar{X} \rightarrow X_{\min}$.

Структурные средние. Определение моды и медианы. Мода и медиана являются вспомогательными описательными характеристиками распределения варьирующего признака.

Модой называется величина признака (варианта), которая чаще всего встречается в данной совокупности. В вариационном ряду это будет варианта, имеющая наибольшую частоту.

Медиана - это варианта, расположенная в середине упорядоченного вариационного ряда. Медиана делит ряд пополам, по обе стороны от нее находится одинаковое количество единиц совокупности.

Мода используется для характеристики наиболее часто встречающегося признака в совокупности (наиболее распространенная должность в организации, наиболее распространенный размер обуви и т.д.). Иными словами мода характеризует типичность явления.

Медиана показывает количественную границу значения варьирующего признака, которую достигла половина членов совокупности. Например, средняя заработная плата наемных работников в целом по экономике Казахстана составляла 19754 тенге, в то же время половина работающих получали заработную плату не более 13 505 тенге, т.е. у половины занятых наемным трудом заработная плата была меньше средней не менее чем в полтора раза!

Два ряда распределения могут иметь заметно различающиеся средние величины некоторого признака и в то же время одинаковое медианное значение. Отсюда, медиана, как и мода, также характеризует *типичность* признака.

Кроме того, мода и медиана позволяют получить представление о структуре совокупности, поэтому их еще называют *структурными* средними.

Пример 1. За смену выработка рабочими однородной продукции характеризуется таким распределением:

Выработка, шт.	40	42	45	46	48	50
Число рабочих, человек	25	50	100	125	150	50

Исчислите среднюю выработку на одного рабочего за смену, моду и медиану.

Решение. В случае, когда имеются ряд дискретных значений признака и частоты их встречаемости, среднее значение признака вычисляется по формуле средней арифметической взвешенной:

$$\bar{x} = (\sum xf) / \sum f = (40 \cdot 25 + 42 \cdot 50 + 45 \cdot 100 + 46 \cdot 125 + 48 \cdot 150 + 50 \cdot 50) / (25 + 50 + 100 + 125 + 150 + 50) = (1000 + 2100 + 4500 + 5750 + 7200 + 2500) / 500 = 23050 / 500 = 46,1.$$

Мода в нашем случае равна 48, так как именно этой variante соответствует наибольшая частота встречаемости выработки (150 рабочих производят за смену такое количество деталей).

Номер медианной варианты равен $500/2 + 0,5 = 250,5$, т.е. медиана находится между 250-й и 251-й вариантой. Накапливая частоты, получим, что медиана находится в четвертой группе ($25 + 50 + 100 + 125 = 300$) и равна 46 деталям.

Пример 2. Два предприятия в отчетном периоде фактически произвели продукции на 10 млн. тенге каждое. При этом одно предприятие выполнило план производства на 112%, а второе - на 105%. Вычислите, как в среднем выполнен план производства продукции на этих двух предприятиях вместе.

Решение. В данном примере не приведены значения планируемых объемов производства, которые можно использовать в качестве весов для исчисления среднего значения выполнения планового задания по двум предприятиям по формуле средней арифметической взвешенной. Однако имеются значения фактически произведенных объемов производства, представляющих произведение планируемых объемов производства и темпов роста производства. Поэтому для определения среднего значения выполнения планового задания следует применить формулу средней гармонической:

$$\bar{x} = (\sum w / \sum w \cdot 1/x).$$

Для исчисления средней гармонической взвешенной нужно:

а) веса разделить на соответствующие варианты:

$$10/1,12 = 8,929, \quad 10/1,05 = 9,524.$$

б) сумму весов разделить на сумму частных от первых делений:

$$20 / (8,929 + 9,524) = 20/18,453 = 1,084.$$

Таким образом, план производства продукции на двух предприятиях выполнен в среднем на 108,4%.

Разделив веса на соответствующие варианты, получаем по существу величину планового объема продукции, а затем, разделив сумму фактического объема производства на сумму плановых объемов, получим процент выполнения плана в среднем.

Пример 3. По следующим данным исчислите среднюю заработную плату рабочих:

Группа рабочих	Средняя месячная заработная плата одного рабочего, тенге	Всего начислено заработной платы, тенге
А	28000	3 500 000
Б	31000	4 340 000
В	27000	2 538 000

Решение. В данном случае в качестве весов можно использовать показатель «Всего начислено заработной платы», представляющий собой не единицу совокупности, а произведение этих единиц на значение признака (т.е. $w = xf$).

Поэтому для исчисления средней должна быть применена формула средней гармонической:

$$\bar{x} = \frac{\text{начисленная зарплата}}{\text{численность рабочих}} = \frac{\sum w}{\sum w \frac{1}{x}}$$

Разделим веса на варианты (получая, тем самым, значения единиц совокупностей по каждой группе):

$$3\,500\,000/28000 = 125; 4\,340\,000/31000 = 140; 2\,538\,000/27000 = 94.$$

Далее, разделим сумму весов на сумму частных от первых делений:

$$(3\,500\,000 + 4\,340\,000 + 2\,538\,000) / (125 + 140 + 94) = 10\,378\,000 / 359 = 28908 \text{ тенге.}$$

То есть средняя заработная плата одного рабочего для данного ряда распределения составляет 28908 тенге.

Пример 4. Выполнение норм выработки рабочих характеризуется следующими данными:

Процент выполнения норм выработки	Число рабочих
90-100	10
100-110	160
110-120	100
120-130	60
130-140	20

На основе этих данных рассчитайте обычным способом и способом моментов: а) средний процент выполнения норм выработки всеми рабочими; б) моду и медиану.

Решение. Поскольку в данном случае мы имеем дело с интервальным вариационным рядом, то для вычисления средней необходимо перейти к дискретному ряду, т.е. по каждой группе вычислить среднее значение интервала и заменить его средним значением:

Процент выполнения норм выработки	Серединное значение интервалов	Число рабочих	Произведение вариантов на частоты
90-100	95	10	950
100-110	105	160	16800
110-120	115	100	11500
120-130	125	60	7500
130-140	135	20	2700
Итого		350	39450

Отсюда средний процент выполнения норм выработки всеми рабочими составит:

$$\bar{x} = (\sum xf) / \sum f = 39450 / 350 = 112,7\%.$$

Вычисление средней способом моментов предполагает: 1) вычитание из всех вариантов постоянного числа (варианты с наибольшей частотой или

варианты, находящейся в середине ряда, в данном случае 115); 2) деление вариант на постоянное число, равное ширине интервала (10):

Серединное значение интервалов (х)	$x_1 = (x - 105) / 10$	Число рабочих (f)	Произведение вариант на частоты (х/ f)
95	-2	10	-20
105	-1	160	-160
115	0	100	0
125	1	60	60
135	2	20	40
Итого		350	-80

Следовательно, средняя арифметическая новых вариант или момент первого ряда составит:

$$m_1 = (\sum xf) / \sum f = -80 / 350 = -0,2286.$$

Для определения средней арифметической величину момента первого порядка умножим на величину интервала, на который делили все варианты, и прибавим к полученному произведению величину варианты, которую вычитали:

$$\bar{x} = i m_1 + A = 10 * (-0,23) + 115 = 112,7\%.$$

Таким образом, мы получили то же самое значение средней арифметической, которое было рассчитано обычным способом.

Для вычисления моды в интервальном вариационном ряду сначала определяется модальный интервал, которому принадлежит наибольшее количество значений признака. В нашем случае это интервал 100-110%. Для определения значения модальной величины применим формулу:

$$M_o = x_{Mo} + i_{Mo} * (f_{Mo} - f_{Mo-1}) / [(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})] = 100 + 10 * (160 - 10) / [(160 - 10) + (160 - 100)] = 100 + 10 * 150 / 210 = 100 + 7,1 = 107,1\%.$$

То есть наиболее часто встречаются значения норм выработки рабочих в районе 107,1%.

Для исчисления медианы необходимо найти интервал (медианный интервал), кумулятивная частота которого превышает половину суммы частот (в нашем примере $350 / 2 = 175$). Им будет третий интервал, поскольку сумма частот в первых двух интервалах ($10 + 160 = 170$) еще не превышает значения 175, а с учетом третьего интервала ($170 + 100 = 270$) - превышает.

Искомое значение медианы определим по формуле:

$$Me = x_{Mo} + i_{Mo} * (\sum f / 2 - S_{Me-1}) / f_{Me} = 110 + 10 * (175 - 170) / 100 = 110 + 10 * 5 / 100 = 110 + 0,5 = 110,5\%.$$

Таким образом, половина рабочих имеют выработку не менее 110,5%.

Пример 5. Используя способ моментов, исчислите среднюю урожайности, моду и медиану по следующим данным:

Урожайность, ц/га	25	28	31	34	37	40	Итого
Площадь посева, в % к итогу	11	19	30	27	8	5	100

Решение. Для исчисления средней урожайности способом моментов вычтем из всех вариантов значение варианты с наибольшей частотой (31) и разделим на величину интервала (3):

(Урожайность - 31) / 3 (x_1)	-2	-1	0	1	2	3	Итого
Площадь посева, в % к итогу (f)	11	19	30	27	8	5	100
Произведение вариантов на частоты ($x_1 f$)	-22	-19	0	27	16	15	-41+58= 17

Отсюда, средняя арифметическая новых вариантов (момент первого порядка) составит:

$$m_1 = (\sum x * f) / \sum f = 17/100 = 0,17.$$

Для определения средней арифметической величину момента первого порядка умножим на величину интервала, на который делили все варианты (на 3), и прибавим к полученному произведению величину варианты, которую вычитали (31):

$$\bar{x} = i m_1 + A = 3 * 0,17 + 31 = 31,51 \text{ ц/га.}$$

Мода равна 31, так как этой variante соответствует наибольшая частота встречаемости (на 30% поле наблюдается урожайность в размере 31 ц/га).

Для исчисления медианы необходимо найти интервал (медианный интервал), кумулятивная частота которого превышает половину суммы частот. Им будет третий интервал, поскольку сумма частот в первых двух интервалах ($11 + 19 = 30$) еще не превышает значения 50, а с учетом третьего интервала ($30 + 30 = 60$) - превышает. Исходя из этого, можно сказать, что на 60% полей урожайность не превышает 31 ц/га и в то же время можно сказать, что на 40% полей урожайность составляет 34 ц/га и выше. Медиана же равна 31 ц/га.

Пример 6. Имеются следующие данные о возрастном составе студентов группы заочного отделения ВУЗа (лет): 19; 19; 19; 20; 20; 20; 20; 20; 20; 20; 20; 21; 21; 21; 22; 23; 23; 24; 25; 25; 25; 26; 27; 29.

Для анализа распределения студентов по возрасту требуется: 1) построить интервальный ряд распределения и его график; 2) рассчитать модальный, медианный и средний возраст, установить его типичность с помощью коэффициентов вариации; 3) проверить распределение на нормальность с помощью коэффициентов асимметрии и эксцесса.

Решение. Для построения интервального ряда из дискретного используется формула Стерджесса, с помощью которой определяется оптимальное количество интервалов (n):

$$n = 1 + 3,322 \lg N, \quad (15)$$

где N – число величин в дискретном ряде.

В нашей задаче $n = 1 + 3,322 \lg 25 = 1 + 3,322 * 1,398 = 5,64$. Так как число интервалов не может быть дробным, то округлим его до ближайшего целого числа, т.е. до 6.

После определения оптимального количества интервалов определяем размах интервала по формуле:

$$h = H / n, \quad (16)$$

где H – размах вариации, определяемый по формуле (37).

$$H = X_{\max} - X_{\min}, \quad (37)$$

где $X_{\max}$ и $X_{\min}$ — максимальное и минимальное значения в совокупности.

В нашей задаче $h = (29 - 19)/6 = 1,67$.

Интервальная группировка данных приведена в первом столбце таблицы **Error! Reference source not found.**, которая содержит также алгоритм и промежуточные расчеты.

Таблица 6 - Вспомогательные расчеты для решения задачи

X_i , лет	f_i	$X_{и}$	$X_{и}f_i$	$X_{и} - \bar{X}$	$ X_{и} - \bar{X} f_i$	$(X_{и} - \bar{X})^2$	$(X_{и} - \bar{X})^2 f_i$	$(X_{и} - \bar{X})^3 f_i$	$(X_{и} - \bar{X})^4 f_i$
до 20,67	12	19,833	237,996	-2,134	25,602	4,552	54,623	-116,539	248,638
20,67-22,33	4	21,5	86,000	-0,467	1,866	0,218	0,871	-0,406	0,189
22,33-24	3	23,167	69,501	1,200	3,601	1,441	4,323	5,190	6,231
24-25,67	3	24,833	74,499	2,866	8,599	8,217	24,650	70,659	202,543
25,67-27,33	2	26,5	53,000	4,533	9,067	20,552	41,105	186,348	844,806
более 27,33	1	28,167	28,167	6,200	6,200	38,446	38,446	238,383	1478,091
Итого	25	—	549,163	—	54,937	—	164,018	383,636	2780,498

На основе этой группировки строится график распределения возраста студентов (рис.2).

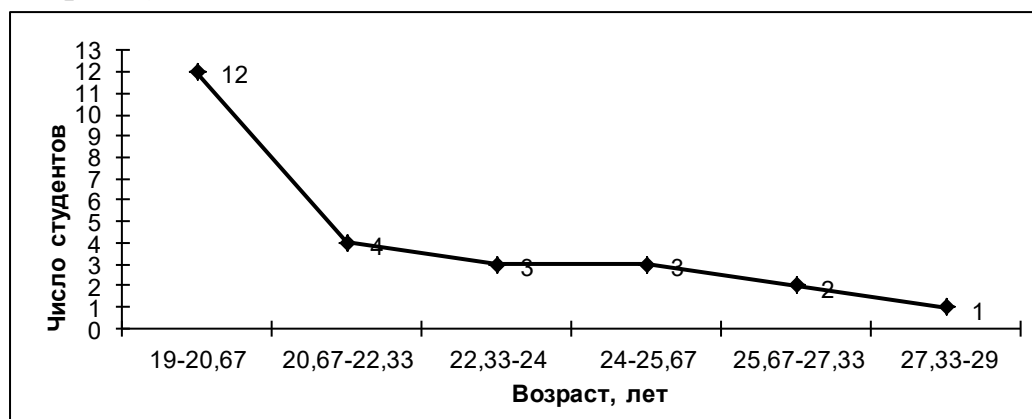


Рис.2. График распределения возраста студентов

Мода — это наиболее часто повторяющееся значение признака. Для интервального ряда с равными интервалами величина моды определяется по формуле (18):

$$Mo = X_{Mo} + h \frac{f_{Mo} - f_{Mo-1}}{2f_{Mo} - f_{Mo-1} - f_{Mo+1}}, \quad (18)$$

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

В нашей задаче чаще всего повторяется (12 раз) первый интервал возраста (до 20,67), значит, это и есть модальный интервал. Используя формулу (18), определяем точное значение модального возраста:

$$Mo = 19 + 1,667 \cdot (12 - 0) / (2 \cdot 12 - 4 - 0) = 20 \text{ (лет)}.$$

Медиана – это такое значение признака, которое приходится на середину ранжированного ряда. Таким образом, в ранжированном ряду распределения одна половина ряда имеет значения признака больше медианы, другая – меньше медианы. Для интервального ряда с равными интервалами величина медианы определяется так:

$$Me = X_{Me} + h \frac{0,5 \sum f - f'_{Me-1}}{f_{Me}}, \quad (49)$$

где X_{Me} – нижняя граница медианного интервала; h – его величина (размах); f'_{Me-1} – сумма наблюдений (или объема взвешивающего признака), накопленная до начала медианного интервала; f_{Me} – число наблюдений или объем взвешивающего признака в медианном интервале.

В нашей задаче второй интервал возраста (от 20,67 до 22,33) является медианным, так как на него приходится середина ряда распределения возраста. Используя формулу (49), определяем точное значение медианного возраста:

$$Me = 20,67 + 1,667 \cdot (12,5 - 12) / 4 = 20,878 \text{ (года)}.$$

В нашей задаче, применяя формулу **Error! Reference source not found.** и подставляя вместо X_i середины интервалов возраста X_{II} , определяем средний возраст студентов: $\bar{X}_{ap} = 549,163 / 25 = 21,967$ (года). Теперь осталось определить типичность или нетипичность найденной средней величины. Это осуществляется с помощью расчета показателей вариации. Чем ближе они к нулю, тем типичнее найденная средняя величина для изучаемой статистической совокупности. При этом критериальным значением коэффициента вариации служит 1/3.

Коэффициенты вариации рассчитываются как отношение среднего отклонения к средней величине. Поскольку среднее отклонение может определяться линейным и квадратическим способами, то соответствующими могут быть и коэффициенты вариации.

Среднее линейное отклонение определяется по формулам (50.1) и (20.26):

$$L = \frac{\sum |X_i - \bar{X}|}{N} - \text{простое}; \quad (50.1)$$

(20.26)

$$L = \frac{\sum |X_i - \bar{X}| f_i}{\sum f_i} - \text{взвешенное}$$

Среднее квадратическое отклонение определяется как корень квадратный из дисперсии, то есть по формуле (217):

$$\sigma = \sqrt{D}. \quad (217)$$

Дисперсия определяется по формулам простая (22.1) или (8):

$$D = \frac{\sum (X_i - \bar{X})^2}{N} - \text{простая (22.1);} \quad D = \frac{\sum (X_i - \bar{X})^2 f_i}{\sum f_i} - \text{взвешенная (8)}$$

В нашей задаче, применяя формулу (22.2), определим ее числитель и внесем в расчетную таблицу. В итоге получим среднее линейное отклонение: $L = 54,937/25 = 2,198$ (года).

Разделив это значение на средний возраст, получим *линейный коэффициент вариации*: $\lambda = \frac{L}{\bar{X}} = 2,198/21,967 = 0,100$.

По значению этого коэффициента для рассмотренной группы студентов делаем вывод о типичности среднего возраста, т.к. расчетное значение коэффициента вариации не превышает критерияльного ($0,100 < 0,333$).

Применяя формулу (8), получим в итоге дисперсию: $D = 164,018/25 = 6,561$. Извлечем из этого числа корень и получим в результате среднее квадратическое отклонение: $\sigma = \sqrt{D} = 2,561$ (года).

Разделив это значение на средний возраст, получим *квадратический коэффициент вариации*: $\nu = \frac{\sigma}{\bar{X}} = 2,561/21,967 = 0,117$.

По значению этого коэффициента для рассмотренной группы студентов можно сделать вывод о типичности среднего возраста, т.к. расчетное значение коэффициента вариации не превышает критерияльного ($0,117 < 0,333$).

В качестве *показателей асимметрии* используются: коэффициент асимметрии – нормированный момент третьего порядка (9) и коэффициент асимметрии Пирсона (10):

$$r_3 = \frac{\mu_3}{\sigma^3}, \quad (9)$$

$$As = \frac{\bar{X} - Mo}{\sigma}. \quad (10)$$

Если значение коэффициента асимметрии положительно, то в ряду преобладают варианты, которые больше средней (правосторонняя

скошенность), если отрицательно – левосторонняя скошенность. Если коэффициент асимметрии равен 0, то вариационный ряд симметричен.

В нашей задаче $\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i} = 383,636/25 = 15,345$; $\sigma^3 = 2,561^3 = 16,797$;

$r_3 = 15,345/16,797 = 0,914 > 0$, значит, распределение студентов по росту с правосторонней асимметрией. Это подтверждает и значение коэффициента асимметрии Пирсона: $As = (21,967 - 20)/2,561 = 0,768$.

Для характеристики *крутизны распределения* используется центральный момент 4-го порядка:

$$\mu_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}. \quad (11)$$

Для образования безразмерной характеристики определяется нормированный момент 4-го порядка $r_4 = \frac{\mu_4}{\sigma^4}$, который и характеризует крутизну (заостренность) графика распределения. При измерении асимметрии эталоном служит нормальное (симметричное) распределение, для которого $r_4 = 3$. Поэтому для оценки крутизны данного распределения в сравнении с нормальным вычисляется эксцесс распределения (12):

$$Ex = \frac{\mu_4}{\sigma^4} - 3. \quad (12)$$

Для приближенного определения эксцесса может быть использована формула Линдберга (13):

$$Ex = d_{\sigma/2} - 0,3829, \quad (13)$$

где $d_{\sigma/2}$ – доля количества вариантов, лежащих в интервале, равном половине σ (в ту и другую сторону от средней величины).

В нашей задаче числитель центрального момента 4-го порядка рассчитан в последнем столбце расчетной таблицы. В итоге по формуле (12) имеем: $Ex = (2780,498 / 25) / 2,561^4 - 3 = 111,220 / 43,017 - 3 = - 0,415$. Так как $Ex < 0$, то распределение низковоершинное. Это подтверждает и приблизительный расчет по формуле (13): в интервале $21,967 \pm 0,5 * 2,561$, то есть от 20,687 до 23,248 находится примерно 21,4% студентов. Таким образом, $Ex = 0,214 - 0,3829 = - 0,169$.

Пример 7 (продолжение примера 1). За смену выработка рабочими однородной продукции характеризуется таким распределением:

Выработка, шт.	40	42	45	46	48	50
Число рабочих, человек	25	50	100	125	150	50

Исчислите дисперсию, среднее квадратическое отклонение и коэффициент вариации. Рассчитайте коэффициент асимметрии.

Решение. При решении задачи к теме 6 нами уже была вычислена средняя выработка рабочими за смену. Она составила 46,1 штук.

Для исчисления дисперсии, среднего квадратического отклонения и коэффициента вариации проведем некоторые расчеты в таблице:

Выработка, шт. (x)	40	42	45	46	48	50
Число рабочих, человек (f)	25	50	100	125	150	50
$x - \bar{x}$	-6,1	-4,1	-1,1	-0,1	1,9	3,9
$(x - \bar{x}) / 2$	37,21	16,81	1,21	0,01	3,61	15,21
$(x - \bar{x}) / 2 f$	930,25	840,5	121	1,25	541,5	760,5

Затем вычислим дисперсию, среднее квадратическое отклонение и коэффициент вариации:

$$\sigma^2 = (\sum (x - \bar{x})^2 f) / \sum f = 3195 / 500 = 6,39;$$

$$\sigma = (639)^{1/2} = 2,53; v = \sigma / \bar{x} * 100 = 2,53 / 46,1 * 100 = 5,5\%.$$

Таким образом, среднее квадратическое отклонение равняется 2,53 штуки, а коэффициент вариации - 5,5%.

Зная среднюю (46,1), моду (48) и среднеквадратическое отклонение (2,53) можно определить коэффициент асимметрии:

$$K_A = (\bar{x} - Mo) / \sigma = (46,1 - 48) / 2,53 = -0,75.$$

Так как знак коэффициента отрицательный, то это характеризует левостороннюю асимметрию. Абсолютная величина коэффициента говорит о наличии сильной асимметрии.

Пример 8 (продолжение примера 4). Выполнение норм выработки рабочих характеризуется следующими данными:

Процент выполнения норм выработки	Число рабочих
90-100	10
100-110	160
110-120	100
120-130	60
130-140	20

На основе этих показателей рассчитайте обычным способом и способом моментов: а) среднее квадратичное отклонение; б) коэффициент вариации; в) коэффициент асимметрии.

Решение. Поскольку в данном случае мы имеем дело с интервальным вариационным рядом, то для вычисления средней и дисперсии необходимо перейти к дискретному ряду, т.е. по каждой группе вычислить срединное значение интервала и заменить интервал этим значением.

Средний процент выполнения норм выработки был вычислен ранее и составил, $x = (\sum xf) / \sum f = 39450 / 350 = 112,7\%$.

Процент выполнения норм выработки	Серединное значение интервалов (x)	Число рабочих (f)	$x - \bar{x}$	$(x - \bar{x}) / 2$	$(x - \bar{x}) / 2 f$
90-100	95	10	-17,7	313,29	3132,9
100-110	105	160	-7,7	52,29	9486,4
110-120	115	100	2,3	5,29	529
120-130	125	60	12,3	151,29	9077,4
130-140	135	20	22,3	497,29	9945,8
Итого		350			32171,5

Подставив полученные данные в формулы, вычислим дисперсию, среднее квадратическое отклонение и коэффициент вариации:

$$\sigma^2 = (\sum (x - \bar{x})^2 f) / \sum f = 32171,5 / 350 = 91,9;$$

$$\sigma = (91,9)^{1/2} = 9,6; \quad v = \sigma / \bar{x} * 100 = 9,6 / 112,7 * 100 = 8,5\%.$$

Таким образом, среднее квадратическое отклонение равняется 9,6 процентных пункта, а коэффициент вариации - 8,5%.

Зная среднюю (112,7), моду (107,1) и среднеквадратическое отклонение (9,6) можно определить коэффициент асимметрии:

$$K_A = (\bar{x} - Mo) / \sigma = (112,7 - 107,1) / 9,6 = +0,58.$$

Так как знак коэффициента положительный, то это говорит о правосторонней асимметрии. Абсолютная величина коэффициента говорит о наличии сильной асимметрии.

Вычисление средней способом моментов предполагает: 1) вычитание из всех вариантов постоянного числа (варианты с наибольшей частотой или варианты, находящейся в середине ряда распределения, в данном случае 115); 2) деление вариантов на постоянное число, равное ширине интервала (10). Далее новые варианты возводятся в квадрат и умножаются на частоты, чтобы получить момент второго порядка, на основе которого затем вычисляется дисперсия.

Серединное значение интервалов (x)	Число рабочих (f)	$x_l = (x - 115) / 10$	x_l^2	$x_l^2 f$
95	10	-2	4	40
105	160	-1	1	160
115	100	0	0	0
125	60	1	1	60
135	20	2	4	80
Итого	350			340

Момент второго порядка вычисляется по формуле:

$$m_2 = (\sum x_l^2 f) / \sum f = 340 / 350 = 0,9714.$$

Дисперсия, исчисленная по способу моментов, равна квадрату величины интервала (10^2), умноженному на разность момента второго порядка (0,9714) и квадрата момента первого порядка, полученного ранее (0,2286²):

$$\sigma^2 = i^2 (m_2 - m_1^2) = 10^2 * (0,9714 - 0,2286^2) = 100 * (0,9714 - 0,0523) = 91,91.$$

Таким образом, мы получили тот же итог, что и в результате прямого исчисления дисперсии, но путем более простых вычислений.

Пример 9 (продолжение примера 5). Используя способ моментов, вычислите среднее квадратическое отклонение и коэффициент асимметрии по следующим данным:

Урожайность, ц/га	25	28	31	34	37	40	Итого
Площадь посева, в % к итогу	11	19	30	27	8	5	100

Решение. Для исчисления среднего квадратического отклонения способом моментов Необходимо вычесть из всех вариантов значение варианты с наибольшей частотой (31) и разделить на величину интервала (3). Далее новые варианты возводятся в квадрат и умножаются на частоты.

Урожайность, ц/га	25	28	31	34	37	40	Итого
Площадь посева, к % к итогу (f)	11	19	30	27	8	5	100
(Урожайность — 31) / 3 (x_l)	-2	-1	0	1	2	3	
XI^2	4	1	0	1	4	9	
xI^2f	44	19	0	27	32	45	167

Момент второго порядка вычисляется по формуле:

$$m_2 = (\sum x^2 \cdot f) / \sum f = 167 / 100 = 1,67.$$

Дисперсия, исчисленная по способу моментов, равна квадрату величины интервала (3^2), умноженному на разность момента второго порядка (1,51) и квадрата момента первого порядка, полученного ранее (0,17²):

$$\sigma^2 = i^2 (m_2 - m_1^2) = 3^2 * (1,67 - 0,17^2) = 9 * (1,67 - 0,0289) = 9 * 1,6411 = 14,77.$$

Отсюда среднеквадратическое отклонение составит:

$$\sigma = (14,77)^{1/2} = 3,84 \text{ ц/га.}$$

Зная среднюю (31,5 ц/га), моду (31 ц/га) и среднеквадратическое отклонение 3.65), можно определить коэффициент асимметрии:

$$K_A = (\bar{x} - Mo) / \sigma = (31,5 - 31) / 3,84 = + 0,133.$$

Так как знак коэффициента положительный, то это говорит о правосторонней асимметрии. Абсолютная величина коэффициента говорит о небольшой асимметрии.

3.2 Задачи и упражнения для самостоятельного решения

1. По имеющимся в следующей таблице данным по группе из 20 студентов заочного отделения необходимо:

- 1) построить интервальный ряд распределения признака и его график;

2) рассчитать модальное, медианное и среднее значение, установить его типичность с помощью коэффициентов вариации;

3) проверить распределение на нормальность с помощью коэффициентов асимметрии и эксцесса.

№ п/п	Вариант									
	1	2	3	4	5	6	7	8	9	10
	Рост, см	Вес, кг	Доход, у.е./мес.	IQ (тест Айзенка)	Тет- радь, листов	Воз- раст, лет	Соот- ношение «рост/вес»	Стаж работы, мес.	Кол-во друзей, чел.	Время решения контрольной, час.
1	159	45	430	95	24	20	3,533	26	5	8,5
2	160	61	640	115	32	25	2,623	63	7	6,2
3	161	56	610	111	24	28	2,875	94	10	6,8
4	162	48	330	97	24	19	3,375	16	4	12,0
5	162	54	420	105	60	23	3,000	49	2	7,5
6	164	58	290	98	16	20	2,828	14	6	10,0
7	166	51	480	109	90	26	3,255	78	9	7,2
8	169	62	610	120	24	19	2,726	10	5	4,2
9	170	70	840	122	48	30	2,429	130	10	3,5
10	170	72	330	92	24	20	2,361	20	3	9,5
11	171	73	560	110	16	28	2,342	86	8	7,8
12	171	64	450	102	48	21	2,672	29	4	8,0
13	172	73	350	108	32	26	2,356	75	7	6,0
14	174	68	310	100	48	21	2,559	22	4	4,8
15	176	81	380	104	64	20	2,173	32	1	8,6
16	176	84	340	104	48	19	2,095	21	5	10,0
17	178	76	660	128	90	27	2,342	96	8	4,5
18	181	90	450	106	48	26	2,011	70	9	12,5
19	183	68	540	105	32	23	2,691	59	6	10,5
20	192	95	750	117	60	27	2,021	98	4	6,5

2. Коэффициенты рождаемости возрастам матери на 1000 женщин соответствующего возраста в 2016 г. в Костанайской области составляли:

Возраст, лет	15-19	20-24	25-29	30-34	35-39	40-44	45-45
коэффициент рождаемости, промилле	24,60	105,15	98,63	73,46	42,71	9,56	0,72

а) Определите средний возраст женщины, родившей в 2016 г. ребенка, моду и медиану. б) Исчислите дисперсию среднего возраста женщин, родивших

в 2016 году ребенка, среднее квадратическое отклонение и коэффициент вариации. Рассчитайте коэффициент асимметрии.

3.Используя способ моментов, определите среднее квадратичное отклонение от среднего возраста докторов наук и коэффициент асимметрии по данным переписи населения 1999 г. в Республике Казахстан:

Возраст, лет	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Количество докторов наук	2	40	110	178	269	339	376	383

4. Распределение населения Костанайской области по возрасту в 2016 г. составляли:

Возраст, лет	Всего населения, чел	Доля населения, %
0-14	168680	19,17
15-19	58080	6,60
20-24	81543	9,27
25-29	79826	9,07
30-34	67323	7,65
35-39	59719	6,79
40-44	59557	6,77
45-49	61375	6,97
50-54	67537	7,67
55-59	50981	5,79
60-64	40123	4,56
65. и старше	85372	9,70

На основе этих данных исчислите обычным способом и способом моментов: а) среднее квадратичное отклонение; б) коэффициент вариации; в) коэффициент асимметрии. Примите во внимание, что среднее значение возраста населения в возрасте 65 лет и старше равно 75 годам.

Контрольные вопросы

1. Дайте определение средней величины. Каково значение средних величин в статистике?
2. Какое отношение к средним величинам имеет закон больших чисел?
3. Какие виды средних величин применяются в статистике?
4. Что называется модой и медианой?
5. Как рассчитываются мода и медиана дискретном вариационном ряду?
6. Как определяется мода и медиана в интервальном вариационном ряду?
7. Укажите основные свойства дисперсии.
8. Что называется коэффициентом детерминации и эмпирическим корреляционным отношением?
9. Что означает выравнивание фактического распределения по кривой нормального распределения?
10. Что такое критерий согласия, асимметрия распределения и эксцесс?

Тема 4. Выборочное наблюдение

4.1 Методические указания и решения типовых задач

Выборочное наблюдение - такое несплошное наблюдение, при котором статистическому обследованию (наблюдению) подвергаются не все единицы изучаемой совокупности, а лишь часть, отобранная в определенном порядке.

Цель выборочного наблюдения - по отобранной части единиц дать характеристику всей совокупности единиц.

По способу - бесповторный и повторный.

Метод отбора определяет конкретный механизм выборки единиц из генеральной совокупности и подразделяется на: собственно случайный; механический; типический; серийный; комбинированный.

Для отобранных единиц рассчитываются обобщенные показатели (средние или относительные) и в дальнейшем результаты выборочного исследования распространяются на всю генеральную совокупность.

Основной задачей при выборочном исследовании является определение ошибок выборки. Принято различать среднюю и предельную ошибки выборки.

Предельной ошибкой выборочного наблюдения называется разность между величиной средней в генеральной совокупности и ее величиной, вычисленной по результатам выборочного наблюдения: $\Delta_{\bar{x}} = |\bar{x} - \tilde{x}|$

В курсах математической статистики доказано, что величина предельной ошибки выборки не должна превышать соотношения: $\Delta_{\bar{x}} \leq t \times \mu$

Значения, связывающие коэффициент доверия t с вероятностью того, что разность между выборочной и генеральной средними не превысит значения средней ошибки выборки μ :

$$\begin{array}{lll} t = 1 \rightarrow F(t) = 0,683 & t = 2 \rightarrow F(t) = 0,954 & t = 3 \rightarrow F(t) = 0,997 \\ t = 1,5 \rightarrow F(t) = 0,866 & t = 2,5 \rightarrow F(t) = 0,988 & t = 3,5 \rightarrow F(t) = 0,999 \end{array}$$

Таблица 7 - Основные характеристики генеральной и выборочной совокупностей

Характеристика	Генеральная совокупность	Выборочная совокупность
Объем совокупности (численность единиц)	N	n
Численность единиц, обладающих исследуемым признаком	M	m
Доля единиц, обладающих исследуемым признаком	$p = M / N$	$w = m / n$
Средний размер признака	$\bar{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
Дисперсия признака	$\sigma_{\bar{x}}^2 = \frac{\sum (x_i - \bar{x})^2}{N}$	$\sigma_{\tilde{x}}^2 = \frac{\sum (x_i - \tilde{x})^2}{n}$
Дисперсия доли	$\sigma_p^2 = pq$	$\sigma_w^2 = W(1 - W)$

Таблица 8 - Формулы расчета средних ошибок выборочной доли и выборочной средней

Метод отбора выборки	Средняя ошибка	
	выборочной доли	выборочной средней
Механический и собственно–случайный повторный	$\mu_w = \sqrt{\frac{w(1-w)}{n}}$	$\mu_x = \sqrt{\frac{\sigma_x^2}{n}}$
Механический и собственно–случайный бесповторный	$\mu_w = \sqrt{\frac{\sigma_w^2}{n} \left(1 - \frac{n}{N}\right)}$	$\mu_x = \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)}$
Серийный при бесповторном отборе серий	$\mu_w = \sqrt{\frac{\delta_w^2}{r} \left(\frac{R-r}{R-1}\right)}$	$\mu_x = \sqrt{\frac{\delta_x^2}{r} \left(\frac{R-r}{R-1}\right)}$
Типический при повторном случайном отборе внутри групп	$\mu_w = \sqrt{\frac{\overline{\sigma_w^2}}{n}}$	$\mu_x = \sqrt{\frac{\overline{\sigma_x^2}}{n}}$
Типический при бесповторном случайном отборе внутри групп	$\mu_w = \sqrt{\frac{\overline{\sigma_w^2}}{n} \left(1 - \frac{n}{N}\right)}$	$\mu_x = \sqrt{\frac{\overline{\sigma_x^2}}{n} \left(1 - \frac{n}{N}\right)}$

Таблица 9 - Формулы расчета объема выборки

Вид выборочного наблюдения	Повторный отбор	Бесповторный отбор
Собственно случайная выборка		
а) при определении среднего размера признака	$n = \frac{t^2 \cdot \sigma_x^2}{\Delta_x^2}$	$n = \frac{t^2 \cdot \sigma_x^2 \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \sigma_x^2}$
б) при определении доли признака	$n = \frac{t^2 \cdot W(1-W)}{\Delta_W^2}$	$n = \frac{t^2 \cdot W(1-W) \cdot N}{\Delta_W^2 \cdot N + t^2 \cdot W(1-W)}$
Механическая выборка	То же	То же
Типичная выборка:		
а) при определении среднего размера признака	$n = \frac{t^2 \cdot \overline{\sigma_x^2}}{\Delta_x^2}$	$n = \frac{t^2 \cdot \overline{\sigma_x^2} \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \overline{\sigma_x^2}}$
б) при определении доли признака	$n = \frac{t^2 \cdot W(1-W)}{\Delta_W^2}$	$n = \frac{t^2 \cdot W(1-W) \cdot N}{\Delta_W^2 \cdot N + t^2 \cdot W(1-W)}$
Серийная выборка:		
а) при определении среднего размера признака	$r = \frac{t^2 \cdot \delta_x^2}{\Delta_x^2}$	$r = \frac{t^2 \cdot \delta_x^2 \cdot R}{\Delta_x^2 \cdot R + t^2 \cdot \delta_x^2}$
б) при определении доли признака	$r = \frac{t^2 \cdot W_r(1-W_r)}{\Delta_W^2}$	$r = \frac{t^2 \cdot W_r(1-W_r) \cdot R}{\Delta_W^2 \cdot R + t^2 \cdot W_r(1-W_r)}$

Пример 1. Результаты выборочного исследования жилищных условий жителей города

Общая (полезная) площадь жилищ, приходящаяся на 1 человека, м ²	До 5	5—10	10—15	15—20	20—25	25—30	30 и более
Число жителей	8	95	204	270	210	130	83

С вероятностью 0,954 определить пределы, в которых находится средний размер общей площади в генеральной совокупности.

Решение. Промежуточные расчеты

Общая (полезная) площадь жилищ, приходящаяся на 1 человека, м ²	Число жителей, f	Середина интервала, x'	$x' \cdot f$	$x'^2 \cdot f$
До 5,0	8	2,5	20,0	50,0
5,0—10,0	95	7,5	712,5	5343,75
10,0—15,0	204	12,5	2550,0	31875,0
15,0—20,0	270	17,5	4725,0	82687,5
20,0—25,0	210	22,5	4725,0	106321,5
25,0—30,0	130	27,5	3575,0	98312,5
30,0 и более	83	32,5	2697,5	87668,75
Итого	1000		19 005,0	412259,0

Выборочная средняя:

$$\bar{x} = 19005,0 / 1000 = 19,0 \text{ м}^2$$

Рассчитываем дисперсию:

$$\sigma^2 = 51,25$$

Рассчитываем среднеквадратическое отклонение:

$$\sigma = \sqrt{51,25} = 7,16 \text{ м}^2$$

Определяем среднюю ошибку выборки:

$$\mu_{\bar{x}} = \frac{7,16}{\sqrt{1000}} = 0,23 \text{ м}^2$$

Рассчитываем предельную ошибку выборки с вероятностью 0,954 (коэффициент доверия $t = 2$):

$$\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}} = 2 \times 0,23 = 0,46 \text{ м}^2$$

Определяем границы изменения генеральной средней:

$$\bar{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \bar{x} + \Delta_{\bar{x}}$$

$$18,54 \leq \bar{x} \leq 19,46$$

Вывод. На основании проведенного выборочного исследования с вероятностью 0,954 можно утверждать, что средний размер общей (полезной)

площади, приходящейся на одного человека, в целом по городу находится в пределах от 18,5 до 19,5 м².

Пример 2. Для определения средней длины детали следует провести исследование методом случайного повторного отбора. Какое количество деталей необходимо отобрать, чтобы ошибка выборки не превышала 3 мм с вероятностью 0,997 при среднем квадратическом отклонении 6 мм? Ошибка и среднее квадратическое отклонение заданы, исходя из технических условий.

При $P = 0,997 \rightarrow t = 3$. Тогда $n = (3^2 \times 6^2) / 3^2 = 36$ деталей.

Пример 3. Методом собственно-случайного (или механического) бесповторного отбора из общей численности работников предприятия (5 тыс.чел.) было отобрано 500 работников. Установлено, что 20% работников в выборке старше 60 лет.

Определить с вероятностью 0,683 пределы, в которых находится доля работников предприятия в возрасте старше 60 лет.

Решение. Средняя ошибка выборочной доли работников старше 60 лет определяется следующим образом (см.табл.1.2)

$$\mu_w = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0,20(1-0,20)}{500} \left(1 - \frac{500}{5000}\right)} = 0,017.$$

С вероятностью 0,683 предельная ошибка выборочной доли работников старше 60-ти лет составит

$$\Delta_w = \mu_w \tau = 0,017 \cdot 1 = 0,017 \text{ (1,7\%).}$$

Верхняя граница генеральной доли

$$p = w + \Delta_w = 0,20 + 0,017 = 0,217 \text{ (21,7\%).}$$

Нижняя граница генеральной доли

$$p = w - \Delta_w = 0,20 - 0,017 = 0,183 \text{ (18,3\%).}$$

Вывод: С вероятностью 0,683 можно утверждать, что доля работников в возрасте старше 60 лет на предприятии колеблется от 18,3% до 21,7%.

$$18,3\% \leq p \leq 21,7\%.$$

Пример 4. При обследовании 100 изделий, отобранных из партии методом механического (или собственно-случайного) повторного отбора, 10 изделий оказались дефектными. Определить с вероятностью 0,866 пределы, в которых находится доля дефектных изделий в партии.

Решение. Для дефектной продукции в выборочной совокупности

$$w = \frac{m}{n} = \frac{10}{100} = 0,1.$$

Средняя ошибка выборочной доли дефектных изделий равна

$$\mu_w = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0,1(1-0,1)}{100}} = 0,03.$$

Предельная ошибка выборочной доли с вероятностью 0.866

$$\Delta_w = \mu_w \tau = 0,03 \cdot 1,5 = 0,045 (4,5\%). \quad 5,5\% \leq p \leq 14,5\%.$$

Вывод: С вероятностью 0,866 можно утверждать, что доля дефектной продукции в партии колеблется от 5,5% до 14,5%.

Пример 5. В трех районах 30 тыс. семей. В первом районе - 15 тыс.; во втором - 12 тыс. и в третьем - 3 тыс. семей. Для определения числа детей в семье была проведена 10%-я типическая выборка с отбором единиц пропорционально численности единиц типических групп. Внутри групп применялся метод случайного бесповторного отбора. Результаты выборочного обследования семей в трех районах представлены:

Номер района	Число семей в районе	Среднее число детей в семье	Среднее квадратическое отклонение
1	15000	1,3	1,2
2	12000	1,8	2,5
3	3000	0,8	0,5

С вероятностью 0,997 определить предел, в котором находится среднее число детей в семье в трех районах.

Решение. Объем выборки в каждой типической группе (районе) n_j

$$n_j = n \frac{N_j}{N},$$

где N_j - число семей в j - м районе;

Число семей, выбранных для обследования в каждом районе при условии, что численность выборочной совокупности n по трем районам равна 3000 семей

$$n_1 = n \frac{N_1}{N} = 3000 \frac{15000}{30000} = 1500 \text{ семей;}$$

$$n_2 = 3000 \frac{12000}{30000} = 1200 \text{ семей;}$$

$$n_3 = 3000 \frac{3000}{30000} = 300 \text{ семей.}$$

Среднее число детей в семье по трем районам в выборочной совокупности (выборочная средняя) с учетом численности отобранных групп

$$\bar{x} = \frac{\sum \tilde{x}_j n_j}{\sum n_j} = \frac{1,3 \cdot 1500 + 1,8 \cdot 1200 + 0,8 \cdot 300}{3000} = 1,45 \text{ чел.}$$

Средняя из групповых дисперсий (внутригрупповая дисперсия)

$$\overline{\sigma_x^2} = \frac{\sum \sigma_j^2 n_j}{\sum n_j} = \frac{1,2^2 \cdot 1500 + 2,5^2 \cdot 1200 + 0,5^2 \cdot 300}{3000} = 3,245.$$

Средняя ошибка выборочной средней при типической выборке (средняя ошибка среднего числа детей в семье)

$$\mu_x = \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{3,245}{3000} \left(1 - \frac{3000}{30000}\right)} = 0,031 \text{ чел.}$$

Предельная ошибка средней с вероятностью 0,997 составит

$$\Delta_x = \mu_x \tau = 0,031 \cdot 3 = 0,093 \cong 0,1 \text{ чел.}$$

Вывод: С вероятностью 0,997 можно утверждать, что в трех районах среднее число детей в семье находится в пределах $1,35 \leq \bar{x} \leq 1,55$.

Пример 6. На предприятии в порядке случайной бесповторной выборки было опрошено 100 рабочих из 1000 и получены следующие данные об их доходе за месяц:

Доход, у.е.	до 300	300-500	500-700	700-1000	более 1000
Число рабочих	8	28	44	17	3

С вероятностью 0,950 определить:

- 1) среднемесячный размер дохода работников данного предприятия;
- 2) долю рабочих предприятия, имеющих месячный доход более 700 у.е.;
- 3) необходимую численность выборки при определении среднемесячного дохода работников предприятия, чтобы не ошибиться более чем на 50 у.е.;
- 4) необходимую численность выборки при определении доли рабочих с размером месячного дохода более 700 у.е., чтобы при этом не ошибиться более чем на 5%.

Решение. Выборочный метод (выборка) используется, когда применение сплошного наблюдения физически невозможно из-за огромного массива данных или экономической нецелесообразности. Учитывая, что на основе выборочного обследования нельзя точно оценить изучаемый параметр (например, среднее значение – $\bar{X}$ или долю какого-то признака – p) генеральной совокупности, необходимо найти пределы, в которых он находится. Для этого необходимо определить изучаемый параметр по данным выборки (выборочную среднюю – $\tilde{X}$ и/или выборочную долю – w) и его дисперсию (D_e). Для этого построим вспомогательную таблицу 10.

Таблица 10 - Вспомогательные расчеты для решения задачи

X_i	f_i	$X_{и}$	$X_{и}f_i$	$(X_{и} - \tilde{X})^2$	$(X_{и} - \tilde{X})^2 f_i$
до 300	8	200	1600	137641	1101128
300 - 500	28	400	11200	29241	818748
500 - 700	44	600	26400	841	37004
700 - 1000	17	850	14450	77841	1323297
более 1000	3	1150	3450	335241	1005723
Итого	100		57100		4285900

По формуле **Error! Reference source not found.** получим средний доход в выборке: $\tilde{X} = 57100/100 = 571$ (у.е.). Применив формулу (8) и рассчитав ее

числитель в последнем столбце таблицы, получим дисперсию среднего выборочного дохода: $D_{\epsilon} = 4285900/100 = 42859$.

Затем необходимо определить *предельную ошибку выборки* по формуле (14):

$$\Delta = t\mu, \quad (14)$$

где t – коэффициент доверия, зависящий от вероятности, с которой определяется предельная ошибка выборки; μ – средняя ошибка выборки, определяемая для повторной выборки по формуле (30), а для бесповторной – по формуле (15):

$$\mu = \sqrt{\frac{D_{\epsilon}}{n}}, \quad (30)$$

$$\mu = \sqrt{\frac{D_{\epsilon}}{n} \left(1 - \frac{n}{N}\right)}, \quad (15)$$

где n – численность выборки; N – численность генеральной совокупности.

В нашей задаче выборка бесповторная, значит, применяя формулу (15), получим среднюю ошибку выборки при определении среднего возраста в генеральной совокупности: $\mu = \sqrt{\frac{42859}{100} \left(1 - \frac{100}{1000}\right)} = 19,640$ (у.е.).

Для определения средней ошибки выборки при определении доли рабочих с доходами более 700 у.е. в генеральной совокупности необходимо определить дисперсию этой доли. Дисперсия доли альтернативного признака w (признак, который может принимать только два взаимоисключающих значения – например, больше или меньше определенного значения) определяется по формуле (16):

$$D_{\epsilon}^{альт} = w(1 - w). \quad (16)$$

В нашей задаче долю альтернативного признака (рабочие с доходами более 700 у.е.) найдем как отношение числа таких рабочих к общему числу рабочих в выборке: $w = 20/100 = 0,2$ или 20%. Теперь определим дисперсию этой доли по формуле (16): $D_{\epsilon}^{альт} = 0,2 * (1 - 0,2) = 0,16$.

Теперь можно рассчитать среднюю ошибку выборки по формуле (15): $\mu = \sqrt{\frac{0,16}{100} \left(1 - \frac{100}{1000}\right)} = 0,038$ или 3,8%.

Значения вероятности β и коэффициента доверия t имеются в математических таблицах нормального закона распределения вероятностей (если в выборке более 30 единиц), из которых в статистике широко применяются сочетания, приведенные в таблице 11:

Таблица 11- Значения интеграла вероятностей Лапласа

β	0,683	0,866	0,950	0,954	0,988	0,997
t	1	1,5	1,96	2	2,5	3

В нашей задаче $\beta = 0,950$, значит $t = 1,96$ (то есть предельная ошибка выборки в 1,96 раза больше средней). Предельная ошибка выборки по формуле (14) будет равна: $\Delta = 1,96 * 19,64 = 38,494$ (у.е.) при определении среднего дохода; $\Delta = 1,96 * 0,038 = 0,075$ или 7,5% при определении доли рабочих с доходами более 700 у.е.

После расчета предельной ошибки находят *доверительный интервал* обобщающей характеристики генеральной совокупности по формуле (17) – для средней величины и по формуле (18) – для доли альтернативного признака:

$$(\tilde{X} - \Delta) \leq \bar{X} \leq (\tilde{X} + \Delta) \quad (17) \quad (w - \Delta) \leq p \leq (w + \Delta) \quad (18)$$

В нашей задаче по формуле (17): $571 - 38,494 \leq \bar{X} \leq 571 + 38,494$ или $532,506$ у.е. $\leq \bar{X} \leq 609,494$ у.е., то есть средний доход всех рабочих предприятия с вероятностью 95% будет лежать в пределах от 532,5 до 609,5 у.е.

Аналогично определяем доверительный интервал для доли по формуле (18): $0,2 - 0,075 \leq p \leq 0,2 + 0,075$ или $0,125 \leq p \leq 0,275$, то есть доля рабочих с доходами более 700 у.е. на всем предприятии с вероятностью 95% будет лежать в пределах от 12,5% до 27,5%.

При разработке программы выборочного наблюдения очень часто задается конкретное значение предельной ошибки (Δ) и уровень вероятности (β). Неизвестной остается минимальная численность выборки (n), обеспечивающая заданную точность. Ее можно получить, если подставить формулу (30) или (15) в формулу (14) и выразить из них n . В результате получатся формулы для вычисления необходимой численности повторной (19) и бесповторной (20) выборок.

$$n_{повт} = \frac{D_{\epsilon} t^2}{\Delta^2}; \quad (19) \quad n_{б/повт} = \frac{D_{\epsilon} t^2}{\Delta^2 + D_{\epsilon} t^2 / N}. \quad (20)$$

В нашей задаче выборка бесповторная, значит, воспользуемся формулой (20), в которую подставим уже рассчитанные дисперсии среднего выборочного дохода рабочих ($D_{\epsilon} = 42859$) и доли рабочих с доходами более 700 у.е. ($D_{\epsilon} = 0,16$):

$$n_{б/повт} = \frac{42859 * 1,96^2}{50^2 + 42859 * 1,96^2 / 1000} = 62 \text{ (чел.)}, \quad n_{б/повт} = \frac{0,16 * 1,96^2}{0,05^2 + 0,16 * 1,96^2 / 1000} = 197 \text{ (чел.)}.$$

Таким образом, необходимо включить в выборку не менее 62 рабочих при определении среднего месячного дохода работников предприятия, чтобы не ошибиться более чем на 50 у.е., и не менее 197 рабочих при определении доли рабочих с размером месячного дохода более 700 у.е., чтобы при этом не ошибиться более чем на 5%.

4.2 Задачи и упражнения для самостоятельного решения

1. В районе А проживает 2000 семей. Предполагается определить средний размер семьи в районе по выборке, взятой методом механического (или собственно-случайного) бесповторного отбора. При этом с вероятностью 0,997 ошибка среднего размера семьи в выборке (выборочной средней) не должна превышать 0,8 человека при среднем квадратическом отклонении в размере семьи 2 человека.

Определить необходимую численность выборки для определения среднего размера семьи в районе.

2. Для определения средней длины детали необходимо провести выборочное обследование методом случайного (или механического) повторного отбора.

Определить, какое количество деталей необходимо отобрать (численность выборки), чтобы ошибка выборки (ошибка выборочной средней) не превышала 2 мм с вероятностью 0,988 при среднем квадратическом отклонении 8 мм.

3. В городе А имеется 10 тыс. семей. С использованием метода выборочных наблюдений предполагается определить долю семей с числом детей три и более. Определить численность выборки, чтобы при механическом (или собственно-случайном) отборе с вероятностью 0,954 ошибка выборки (доли семей с числом детей три и более) не превышала 0,02, если на основе предыдущих обследований известно, что дисперсия равна 0,2.

4. Для изучения оснащения 500 предприятий основными производственными фондами было проведено 10%-е выборочное обследование методом собственно-случайного (или механического) отбора, в результате которого получены следующие данные о распределении предприятий по стоимости основных производственных фондов:

Среднегодовая стоимость основных производственных фондов, млн тнг	До 20	20-40	40-60	Свыше 60	Итого:
Число предприятий	5	12	23	10	50

Определить:

- с вероятностью 0,997 предельную ошибку выборочной средней и границы, в которых будет находиться среднегодовая стоимость основных производственных фондов всех предприятий генеральной совокупности;
- с вероятностью 0,954 предельную ошибку выборки при определении доли и границы, в которых будет находиться удельный вес предприятий со стоимостью основных производственных фондов свыше 40 млн тнг.;
- объемы выборочной совокупности при условии, что:

– предельная ошибка выборки при определении среднегодовой стоимости основных производственных фондов с вероятностью 0,997 была бы не более 5 млн тнг.;

– предельная ошибка доли предприятий со стоимостью основных производственных фондов свыше 40 млн тнг. с вероятностью 0,954 была бы не более 15%.

5. Для определения средней заработной платы продавцов была проведена 20 %-ная типическая выборка с отбором единиц пропорционально численности типических групп (внутри типов применялся метод случайного бесповторного отбора). Результаты выборки представлены в таблице:

Тип магазина	Средняя заработная плата, тнг	Среднее квадратическое отклонение, тнг	Число продавцов, чел
1	10000	500	150
2	11000	2000	350
3	15000	1000	500

С вероятностью 0,997 определите пределы, в которых находится средняя заработная плата всех продавцов магазина.

6. На предприятиях города работает 15000 рабочих определенной профессии. Необходимо провести обследование их месячной заработной платы и определить среднюю ее величину у одного рабочего данной профессии. Провести обследование всех рабочих не представляется возможным. По схеме случайного бесповторного отбора были получены данные о размере заработной платы 150 челе век (табл.1). Необходимо при заданном уровне доверия (вероятности) определить границы, в которых находится средний уровень заработной платы в генеральной совокупности.

По условию задачи известно, что отбор - случайный бесповторный; $N = 15000$; $n = 150$. Рассмотрим два варианта допустимой вероятности ошибки выборочного наблюдения - 0,683 и 0,997.

Результаты выборочного наблюдения уровня заработной платы

Результаты наблюдения		
уровень заработной платы, тыс. тенге.	численность работников	средний уровень заработной платы в группе
	f	$\tilde{x}$
До 9,0	8	8,5
9,0-10,0	22	9,5
10,0-10,5	65	10,25
10,5—11,0	40	10,75
Свыше 11,0	15	11,25
Итого	150	

7. Для изучения вкладов населения в коммерческом банке города была проведена 5%-я случайная выборка лицевых счетов, в результате которой получено следующее распределение клиентов по размеру вкладов:

Размер вклада, у.е.	Число вкладчиков, чел.									
	Вариант									
	1	2	3	4	5	6	7	8	9	10
до 5000	10	80	100	50	60	30	90	20	70	40
5 000 – 15 000	40	60	150	30	40	110	75	65	90	80
15 000 – 30 000	25	35	70	90	120	90	130	140	60	95
30 000 – 50 000	30	45	40	5	80	30	60	75	20	115
свыше 50 000	15	10	30	25	50	15	25	5	10	5

С вероятностью 0,954 определить:

- 1) средний размер вклада во всем банке;
- 2) долю вкладчиков во всем банке с размером вклада свыше 15000 у.е.;
- 3) необходимую численность выборки при определении среднего размера вклада, чтобы не ошибиться более чем на 500 у.е.;
- 4) необходимую численность выборки при определении доли вкладчиков во всем банке с размером вклада свыше 30 000 у.е., чтобы не ошибиться более чем на 10%.

Контрольные вопросы

1. Что понимается под выборочным наблюдением?
2. В чем состоит главная цель выборочного наблюдения?
3. Как называется статистическая совокупность, из которой производится отбор единиц при организации выборочного наблюдения?
4. Как называется абсолютная разница между средними, определенными по генеральной и выборочной совокупностям?
5. Что означает коэффициент доверия в зависимости для определения предельной ошибки выборочного наблюдения?
6. Виды формирования выборочной совокупности.
7. Методы формирования выборочной совокупности.
8. Способы отбора единиц при формировании выборочной совокупности.
9. Что называется малой выборкой?
10. Какой закон распределения используется в малых выборках?
11. Почему при выборочном наблюдении неизбежны ошибки и как они классифицируются?
12. Каковы условия правильного отбора единиц совокупности при выборочном наблюдении?
13. Как производятся собственно случайный, механический, типический и серийный отборы?
14. В чем различие повторной и бесповторной выборки?
15. Что представляет собой средняя ошибка выборки (для средней и доли)?

Тема 5. Ряды динамики

5.1 Методические указания и решения типовых задач

Динамические ряды – это временная последовательность каких-либо показателей, где показатель периода времени – года (месяцы), а показатель уровня ряда – числа.

Характеристики динамических рядов – это показатели, которые характеризуют изменения явления во времени.

Определение статистических характеристик динамического ряда основано на абсолютном и относительном сравнении уровней ряда ($y_2 - y_1$ – абсолютное сравнение, y_2/y_1 – относительное сравнение).

При нахождении характеристик могут использоваться два способа:

- цепной способ, т.е. когда данный уровень сравнивается с предыдущим;
- базисный способ, т.е. когда каждый данный уровень сравнивается с одним и тем же начальным уровнем, принятым за базу сравнения.

К статистическим характеристикам динамического ряда относят: темп роста и прироста, абсолютный прирост, базисные и цепные, абсолютное содержание 1% прироста, средний уровень ряда, средний абсолютный прирост, средний темп роста и прироста.

Абсолютный прирост, Δ_i (или скорость роста), определяется как разность между двумя уровнями динамического ряда

$$\Delta_i = y_i - y_{баз.} \text{ (базисный способ)} \quad \Delta_i = y_i - y_{i-1} \text{ (цепной способ)}$$

Коэффициент роста, K_i , определяется как отношение двух сравниваемых уровней

$$K_i = \frac{y_i}{y_{баз.}} \text{ (базисный способ)} \quad K_i = \frac{y_i}{y_{i-1}} \text{ (цепной способ)}$$

Темп роста, T_p , то же, что и коэффициента роста, но выраженный в процентах

$$T_p = K \cdot 100 \%$$

Темп прироста, отношение абсолютного прироста к базисному или предшествующему уровню.

$$T_n = \frac{y_i - y_{баз.}}{y_{баз.}} \cdot 100\% \text{ (базисный способ)}$$

$$T_n = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%$$

Абсолютное значение 1% прироста, A_i , рассчитывают как отношение абсолютного прироста к темпу прироста (в %) за тот же период времени.

$$A_i = \frac{\Delta_i}{T_n}$$

Средний уровень ряда – величина абсолютная, т.е. имеет определенные единицы измерения, определенную размерность.

Средний уровень ряда ($\bar{y}$) характеризует среднюю величину показателя за данный период. Средний уровень ряда рассчитывается как средняя величина из уровней ряда, причем по-разному для интервальных и моментных рядов.

В интервальных рядах по средней арифметической:

$$\bar{y}_{ap} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{\sum y_i}{n}$$

В моментных рядах по средней хронологической:

$$\bar{y}_{xp_{np}} = \frac{y_0/2 + y_2 + \dots + y_{n-1} + y_n/2}{n-1}$$

где $n-1$ – количество изменений за данный период.

$$\bar{y}_{xp_{взвеш.}} = \frac{(y_1 + y_2) \times t_1 + (y_2 + y_3) \times t_2 \dots + (y_{n-1} + y_n) \times t_{n-1}}{2(t_1 + t_2 + \dots + t_{n-1})}$$

где $y_1, y_2, \dots, y_n$ – соответствующий уровень ряда,
 $t_1, t_2, \dots, t_{n-1}$ – соответствующий период времени.

Средний абсолютный прирост ($\bar{\Delta y}$) – это средняя из абсолютных приростов за равные промежутки времени:

$$\bar{\Delta y} = \frac{\Delta y_1 + \Delta y_2 + \dots + \Delta y_n}{n-1} = \frac{\sum_{i=1}^n \Delta y_i}{n-1} = \frac{y_n - y_0}{n}$$

где $\Delta y_1, \Delta y_2, \dots, \Delta y_n$ – соответствующий абсолютный прирост,

$n-1$ – количество изменений за данный период,

y_n – последний уровень ряда,

y_0 – начальный, базисный уровень ряда.

Средний темп роста ($\bar{T}_p$) – это средняя из темпов роста за данный период, которая показывает, во сколько раз в среднем (за год, месяц) изменяется явление.

Средний темп роста определяется всегда по средней геометрической.

Средний темп роста можно определить исходя из цепных коэффициентов (темпов) роста:

$$\bar{T}_p = \sqrt[n-1]{T_{p_{цепн1}} \times T_{p_{цепн2}} \times \dots \times T_{p_{цепн n}}}$$

или абсолютных уровней ряда (базисного темпа роста):

$$\bar{T}_p = \sqrt[n-1]{T_{p_{баз}}},$$

где $T_{p_{цепн1}}, T_{p_{цепн2}}, \dots, T_{p_{цепнn}}$ - соответствующие цепные темпы роста (y_i / y_{i-1}),
 $T_{p_{баз}}$ - базисный темп роста за весь период (y_n / y_0),
 $n-1$ – количество изменений за данный период.

Средний темп роста обычно выражается в коэффициентах, но может быть и в процентах.

Средний темп прироста ($\overline{T_{np}}$) – характеризует темп прироста в среднем за период и определяется на основе среднего темпа роста:

$$\overline{T_{np}} = \overline{T_p} - 1$$

$$\overline{T_{np}}(\%) = \overline{T_p}(\%) - 100\%$$

где $\overline{T_p}$ - средний темп роста (в коэффициентах или в процентах).

Средний темп прироста показывает, на сколько процентов изменился уровень ряда в среднем за данный период.

Средний темп прироста выражается в коэффициентах или в процентах.

Вычисление данных показателей является первым этапом анализа динамических рядов и позволяет выявить скорость и интенсивность развития явления, представленного данным рядом

Пример 1. Производство картофеля в регионе характеризуется следующими данными, млн. тонн:

Годы	Производство картофеля, млн. тонн
1990	84
1995	78
1996	83
1997	85
1998	82
1999	86
2000	89

Для анализа производства картофеля в регионе за 1995-2000 гг. вычислите:

1. Абсолютные приросты, темпы роста и темпы прироста по годам и к 1995 году. Полученные показатели представьте в таблице.
2. Среднегодовое производство картофеля.
3. Среднегодовой темп роста и прироста производства картофеля за 1995-2000 гг. и за 1990-1995 гг.
4. Постройте график производства картофеля в регионе за 1990-2000гг. Сделайте выводы.

Решение. Определим показатели, характеризующие рост производства картофеля: абсолютные приросты, темпы роста и прироста (по годам и к

базисному 1995 году). Для расчета используем следующие формулы (15-20).
Результаты приведены в таблице.

Годы	Производство картофеля, млн. тонн.	Абсолютный прирост, млрд. тнг.		Темпы роста, %		Темпы прироста, %	
		по годам	к базисному году	по годам	к базисному году	по годам	к базисному году
1990	84	-	-	-	-	-	-
1995	78	-6	-6	92,86	92,86	-7,14	-7,14
1996	83	5	-1	106,41	98,81	6,41	-1,19
1997	85	2	1	102,41	101,19	2,41	1,19
1998	82	-3	-2	96,47	97,62	-3,53	-2,38
1999	86	4	2	104,88	102,38	4,88	2,38
2000	89	3	5	103,49	105,95	3,49	5,95

Среднегодовое производство картофеля определим по формуле средней арифметической взвешенной:

$$\bar{y} = \frac{\frac{84+78}{2} \cdot 6 + 83 + 85 + 82 + 86 + 89}{11} = \frac{911}{11} = 82,818 \text{ млн. тонн.}$$

Среднегодовой темп роста ряда динамики определяется по формуле среднего геометрического:

$$\overline{T_D} = \sqrt[n-1]{\frac{y_n}{y_1}}.$$

Подставив исходные данные, получим среднегодовой темп роста производства картофеля:

$$\text{за 1995-2000 гг.: } \overline{T_{1995-2000}} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5]{\frac{89}{78}} = 1,0267, \text{ или } 102,67\%;$$

$$\text{за 1990-1995 гг.: } \overline{T_{1990-1995}} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5]{\frac{78}{84}} = 0,9853, \text{ или } 98,53\%.$$

Среднегодовой абсолютный прирост определяется по формуле:

$$\overline{T_{iD}} = \overline{T_D} - 1$$

Подставив рассчитанные, получим среднегодовой темп роста производства картофеля:

$$\text{за 1995-2000 гг.: } \overline{T_{iD_{1995-2000}}} = 1,0267 - 1 = 0,0267, \text{ или } 2,67\%;$$

$$\text{за 1990-1995 гг.: } \overline{T_{iD_{1990-1995}}} = 0,9853 - 1 = -0,0147, \text{ или } -1,47\%.$$

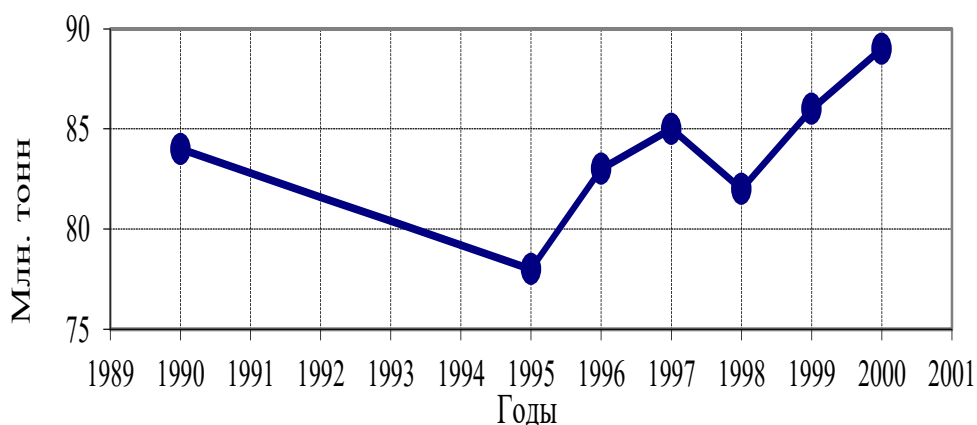


Рисунок 1. Производство картофеля в регионе за 1990-2000 гг.

Вывод. Анализ графика и полученных расчетных данных свидетельствует о том, что: производство картофеля убывало с 1990 г. по 1995 г. включительно, а затем стало расти (за исключением временного спада в 1998 г.); темп прироста в 2000 г. к 1990 г. составил лишь 5,95%.

Пример 2. Динамические ряды абсолютных величин основных технико-экономических показателей

Показатели	Ед. изм.	Года				
		1990	1991	1992	1993	1994
Объем валовой (товарной, реализуемой) продукции	млн. тнг.	69,5	69,3	68,4	68,5	64,3
Численность промышленно-производственного персонала	чел.	495	520	550	630	625
Стоимость основных производственных фондов	млн. тнг.	61,0	50,6	58,6	56,4	54,3
Фондовооруженность	млн. тнг./ чел.	0,123	0,097	0,107	0,090	0,087
Фондоотдача	млн. тнг.	1,139	1,37	1,167	1,215	1,184
Фондоемкость	тнг./тнг.	0,878	0,73	0,857	0,823	0,844

Вычислить:

- Абсолютные величины основных технико-экономических показателей;
- Статистические характеристики динамического ряда технико-экономического показателя (по каждому показателю).
- Средние характеристики динамических рядов технико-экономических показателей

Решение. Полученный первичный материал о работе данного производства обрабатывается и записывается в таблицу 2 в виде динамических рядов абсолютных величин.

- Динамические ряды абсолютных величин основных технико-экономических показателей.

Фондовооруженность рассчитывается по формуле: $\Phi_v = \frac{s}{T}$

где s - стоимость основных производственных фондов, млн. тнг.

T – численность промышленно-производственного персонала, чел.

$$\begin{aligned}\Phi_{\phi}(90) &= \frac{61}{495} = 0,123 \text{ (млн. тнг./чел.)} & \Phi_{\phi}(93) &= \frac{56,4}{630} = 0,09 \text{ (млн. тнг./чел.)} \\ \Phi_{\phi}(91) &= \frac{50,6}{520} = 0,097 \text{ (млн. тнг./чел.)} & \Phi_{\phi}(94) &= \frac{54,3}{625} = 0,087 \text{ (млн. тнг./чел.)} \\ \Phi_{\phi}(92) &= \frac{58,6}{550} = 0,107 \text{ (млн. тнг./чел.)}\end{aligned}$$

Показатели	Ед. изм.	Года				
		1990	1991	1992	1993	1994
Объем валовой (товарной, реализуемой) продукции	млн. тнг.	69.5	69.3	68.4	68.5	64.3
Численность промышленно-производственного персонала	чел.	495	520	550	630	625
Стоимость основных производственных фондов	млн. тнг.	61.0	50.6	58.6	56.4	54.3

Фондоотдача рассчитывается по формуле: $\Phi_o = \frac{q}{s}$

где q – объем валовой продукции, млн. тнг.

s – стоимость основных производственных фондов, млн. тнг.

$$\begin{aligned}\Phi_o(90) &= \frac{69,5}{61} = 1,139 \text{ (тнг./тнг.)} & \Phi_o(93) &= \frac{68,5}{56,4} = 1,215 \text{ (тнг./тнг.)} \\ \Phi_o(91) &= \frac{69,3}{50,6} = 1,37 \text{ (тнг./тнг.)} & \Phi_o(94) &= \frac{64,3}{54,3} = 1,184 \text{ (тнг./тнг.)} \\ \Phi_o(92) &= \frac{68,4}{58,6} = 1,167 \text{ (тнг./тнг.)}\end{aligned}$$

Статистические характеристики динамического ряда технико-экономического показателя.

Показатели	Ед. изм.	Года				
		1990	1991	1992	1993	1994
Фондовооруженность	Млн. тнг./чел	0,123	0,097	0,107	0,090	0,087
Фондоотдача		1,139	1,37	1,167	1,215	1,184
Фондоемкость		0,878	0,73	0,857	0,823	0,844

На основании имеющихся абсолютных значений основных технико-экономических показателей рассчитаем характеристики динамических рядов.

Фондоёмкость рассчитывается по формуле: $\Phi_e = \frac{1}{\Phi_o}$

где Φ_o – фондоотдача, тнг./тнг.

$$\begin{aligned}\Phi_e(90) &= \frac{1}{1,139} = 0,878 \text{ (тнг./тнг.)} & \Phi_e(93) &= \frac{1}{1,215} = 0,823 \text{ (тнг./тнг.)} \\ \Phi_e(91) &= \frac{1}{1,37} = 0,73 \text{ (тнг./тнг.)} & \Phi_e(94) &= \frac{1}{1,184} = 0,844 \text{ (тнг./тнг.)} \\ \Phi_e(92) &= \frac{1}{1,167} = 0,857 \text{ (тнг./тнг.)}\end{aligned}$$

Расчет абсолютного прироста

Рассмотрим расчет абсолютного прироста на примере показателя объема валовой продукции.

$$\Delta y_{i_{цепн}} = y_i - y_{i-1}$$

$$\Delta y_{91_{цепн}} = 69,3 - 69,5 = -0,2 \text{ (млн. тнг.)}$$

$$\Delta y_{92_{цепн}} = 68,4 - 69,3 = -0,9 \text{ (млн. тнг.)}$$

$$\Delta y_{93_{цепн}} = 68,5 - 68,4 = 0,1 \text{ (млн. тнг.)}$$

$$\Delta y_{94_{цепн}} = 64,3 - 68,5 = -4,2 \text{ (млн. тнг.)}$$

$$\Delta y_{i_{баз}} = y_i - y_1$$

$$\Delta y_{91_{баз}} = 69,3 - 69,5 = -0,2 \text{ (млн. тнг.)}$$

$$\Delta y_{92_{баз}} = 68,4 - 69,5 = -1,1 \text{ (млн. тнг.)}$$

$$\Delta y_{93_{баз}} = 68,5 - 69,5 = -1 \text{ (млн. тнг.)}$$

$$\Delta y_{94_{баз}} = 64,3 - 69,5 = -5,2 \text{ (млн. тнг.)}$$

Проверка: Проверку осуществляют на основе взаимосвязи показателей.

$$\Delta y_{94_{баз}} = \Delta y_{91_{цепн}} + \Delta y_{92_{цепн}} + \Delta y_{93_{цепн}} + \Delta y_{94_{цепн}}$$

$$\Delta y_{94_{баз}} = -0,2 - 0,9 + 0,1 - 4,2 = -5,2 \text{ (млн. тнг.)}$$

Расчет темпа роста

Рассмотрим расчет темпа роста на примере показателя объема валовой продукции.

$$T_{P_{баз}} = \frac{y_i}{y_1}$$

$$T_{P_{баз}} \left(\frac{91}{90} \right) = \frac{69,3}{69,5} = 0,997 ; 99,7\%$$

$$T_{P_{баз}} \left(\frac{92}{90} \right) = \frac{68,4}{69,5} = 0,984 ; 98,4\%$$

$$T_{P_{цепн}} = \frac{y_i}{y_{i-1}}$$

$$T_{P_{цепн}} \left(\frac{91}{90} \right) = \frac{69,3}{69,5} = 0,997 ; 99,7\%$$

$$T_{P_{цепн}} \left(\frac{92}{91} \right) = \frac{68,4}{69,3} = 0,987 ; 98,7\%$$

$$T_{P_{баз}} \left(\frac{93}{90} \right) = \frac{68,5}{69,5} = 0,986 ; 98,6\%$$

$$T_{P_{баз}} \left(\frac{94}{90} \right) = \frac{64,3}{69,5} = 0,925 ; 92,5\%$$

$$T_{P_{цепн}} \left(\frac{93}{92} \right) = \frac{68,5}{68,4} = 1,0015 ; 100,15\%$$

$$T_{P_{цепн}} \left(\frac{94}{93} \right) = \frac{64,3}{68,5} = 0,939 ; 93,9\%$$

Проверка: Проверку осуществляют на основе взаимосвязи показателей.

$$T_{P_{\text{баз}}}\left(\frac{94}{90}\right) = T_{P_{\text{ценн}}}\left(\frac{91}{90}\right) \times T_{P_{\text{ценн}}}\left(\frac{92}{91}\right) \times T_{P_{\text{ценн}}}\left(\frac{93}{92}\right) \times T_{P_{\text{ценн}}}\left(\frac{94}{93}\right)$$

$$T_{P_{\text{баз}}}\left(\frac{94}{90}\right) = 0,997 \times 0,987 \times 1,0015 \times 0,939 = 0,925 ; 92,5\%$$

Расчет темп прироста

Рассмотрим расчет темпа прироста на примере показателя объема валовой продукции.

$$T_{np_{\text{ценн}}} = \frac{\Delta y_i}{y_{i-1}}$$

$$T_{np_{\text{ценн}}}\left(\frac{91}{90}\right) = \frac{-0,2}{69,5} = -0,003 ; -0,3\%$$

$$T_{np_{\text{ценн}}}\left(\frac{93}{92}\right) = \frac{0,1}{68,4} = 0,0015 ; 0,15\%$$

$$T_{np_{\text{ценн}}}\left(\frac{92}{91}\right) = \frac{-0,9}{69,3} = -0,013 ; -1,3\%$$

$$T_{np_{\text{ценн}}}\left(\frac{94}{93}\right) = \frac{-4,2}{68,5} = -0,061 ; 6,1\%$$

Проверка: В качестве проверки рассчитаем этот показатель по формуле:

$$T_{np_{\text{ценн}}} = T_{P_{\text{ценн}}} - 1$$

$$T_{np_{\text{ценн}}}\left(\frac{91}{90}\right) = 0,997 - 1 = -0,003 ; -0,3\%$$

$$T_{np_{\text{ценн}}}\left(\frac{92}{91}\right) = 0,987 - 1 = -0,013 ; -1,3\%$$

$$T_{np_{\text{ценн}}}\left(\frac{93}{92}\right) = 1,0015 - 1 = 0,0015 ; 0,15\%$$

$$T_{np_{\text{ценн}}}\left(\frac{94}{93}\right) = 0,939 - 1 = -0,061 ; -6,1\%$$

$$T_{np_{\text{баз}}} = \frac{\Delta y_i}{y_1}$$

$$T_{np_{\text{баз}}}\left(\frac{91}{90}\right) = \frac{-0,2}{69,5} = -0,003 ; -0,3\%$$

$$T_{np_{\text{баз}}}\left(\frac{93}{90}\right) = \frac{-1}{69,5} = -0,014 ; -1,4\%$$

$$T_{np_{\text{баз}}}\left(\frac{92}{90}\right) = \frac{-1,1}{69,5} = -0,016 ; -1,6\%$$

$$T_{np_{\text{баз}}}\left(\frac{94}{90}\right) = \frac{-5,2}{69,5} = -0,075 ; -7,5\%$$

Проверка: В качестве проверки рассчитаем этот показатель по формуле:

$$T_{np_{\text{баз}}} = T_{P_{\text{баз}}} - 1$$

$$T_{np_{\text{баз}}}\left(\frac{91}{90}\right) = 0,997 - 1 = -0,003 ; -0,3\%$$

$$T_{np_{\text{баз}}}\left(\frac{92}{90}\right) = 0,984 - 1 = -0,016 ; -1,6\%$$

$$T_{np\bar{баз}}\left(\frac{93}{90}\right) = 0,986 - 1 = -0,014; -1,4\%$$

$$T_{np\bar{баз}}\left(\frac{94}{90}\right) = 0,925 - 1 = -0,075; -7,5\%$$

Расчет абсолютного содержания одного процента прироста

Рассмотрим расчет абсолютного содержания одного процента прироста на примере показателя объема валовой продукции.

$$\dot{A} = \frac{\Delta y_{i, \bar{баз}}}{T_{i, \bar{баз}}, \%}$$

$$\dot{A}_{91} = \frac{-0,2}{-0,3\%} = 0,695 (\text{млн. тнг./ \%})$$

$$\dot{A}_{93} = \frac{-0,1}{-0,15\%} = 0,684 (\text{млн. тнг./ \%})$$

$$\dot{A}_{92} = \frac{-0,9}{-1,3\%} = 0,693 (\text{млн. тнг./ \%})$$

$$\dot{A}_{94} = \frac{-4,2}{-0,1\%} = 0,685 (\text{млн. тнг./ \%})$$

Проверка: В качестве проверки рассчитаем этот показатель по формуле:

$$\dot{A}_i = \frac{y_{i-1}}{100\%}$$

$$\dot{A}_{91} = \frac{69,5}{100\%} = 0,695 (\text{млн. тнг./ \%})$$

$$\dot{A}_{93} = \frac{68,4}{100\%} = 0,684 (\text{млн. тнг./ \%})$$

$$\dot{A}_{92} = \frac{69,3}{100\%} = 0,693 (\text{млн. тнг./ \%})$$

$$\dot{A}_{94} = \frac{68,5}{100\%} = 0,685 (\text{млн. тнг./ \%})$$

Статистические характеристики динамического ряда технико-экономического показателя объема валовой продукции

Показатели	Ед. изм.	Года				
		1990	1991	1992	1993	1994
Объем валовой (товарной, реализуемой) продукции	млн. тнг.	69,5	69,3	68,4	68,5	64,3
1. Базисный темп роста	%	100	99,7	98,4	98,6	92,5
2. Базисный темп прироста	%	0	-0,3	-1,6	-1,4	-7,5
3. Цепной темп роста	%	100	99,7	98,7	100,1	93,9
4. Цепной темп прироста	%	0	-0,3	-1,3	0,15	-6,1
5. Абсолютный прирост к базисному периоду	млн. тнг.	-	-0,2	-1,1	-1	-5,2
6. Абсолютный прирост к предыдущему году	млн. тнг.	-	-0,2	-0,9	0,1	-4,2
7. Абсолютное содержание 1% прироста	млн. тнг./%	-	0,695	0,693	0,684	0,685

Расчет среднего уровня ряда (на примере показателя объема валовой продукции)

$$\bar{y} = \frac{65,4/2 + 69,3 + 68,4 + 68,5 + 64,3/2}{4} = 68,275 (\text{млн. тнг.})$$

Расчет среднего абсолютного прироста (на примере показателя объема валовой продукции)

$$\overline{\Delta y} = \frac{0 - 0,2 - 0,9 + 0,1 - 4,2}{4} = -1,3 \text{ (млн. тнг.)}$$

Проверка: В качестве проверки рассчитаем этот показатель по формуле:

$$\overline{\Delta y} = \frac{\Delta y_n - \Delta y_1}{n}$$

$$\overline{\Delta y} = \frac{64,3 - 69,5}{4} = -1,3 \text{ (млн. тнг.)}$$

Расчет среднего темпа роста (на примере показателя объема валовой продукции)

$$\overline{T_p} = \sqrt[4]{0,997 \times 0,987 \times 1,0015 \times 0,935} = 0,98; 98\%$$

Проверка: В качестве проверки рассчитаем этот показатель по формуле:

$$\overline{T_p} = \sqrt[n]{T_{P_{баз}}}$$

$$\overline{T_p} = \sqrt[4]{0,925} = 0,98; 98\%$$

Расчет среднего темпа прироста

$$\overline{T_{np}} = 0,98 - 1 = -0,02; -2\%$$

Средние характеристики динамических рядов технико-экономических показателей

Показатели	Средний уровень	Характеристики за период 1990-1994 гг.		
		Средний темп роста	Средний темп прироста	Средний абсолютный прирост
Объем валовой продукции, млн. тнг.	68,275	0,98	-0,02	-1,3
Численность ППП, чел.	565	1,06	0,06	33
Стоимость ОПФ, млн. тнг.	55,813	0,97	-0,03	-1,675
Фондовооруженность, млн. тнг./чел.	0,0996	0,92	-0,08	-0,009
Фондоотдача, тнг./тнг.	1,228	1,01	0,01	0,011
Фондоемкость, тнг./тнг.	0,818	0,99	-0,01	-0,009

Прогнозирование развития динамических рядов

Следует установить недостающий уровень ряда на основании:

- среднего абсолютного прироста;
- среднего темпа роста;
- теоретической кривой, выражающей основные черты развития явления.

Выравнивание рядов динамики может использоваться не только для выявления основных тенденций развития, но и для нахождения неизвестных

значений показателя как внутри рассматриваемого периода, так и за его пределами.

Нахождение неизвестного показателя внутри рассматриваемого периода называется интерполяцией, а за его пределами в прошлом или будущем – экстраполяцией.

Для экстраполяции и интерполяций можно использовать различные методы, но все они основаны на том, что тенденция, выявленная на данном промежутке времени, сохраняется как за его пределами, так и внутри.

При экстраполяции и интерполяции используются следующие простейшие методы:

1. Основан на определении среднего абсолютного прироста. Если абсолютные приросты по годам примерно одинаковы, то находится средний абсолютный прирост, а затем он прибавляется (отнимается) к последнему известному значению столько раз, на какой период осуществляется экстраполяция или интерполяция.

2. Основан на определении среднего темпа роста. Если по годам уровни ряда изменяются примерно в одно и то же число раз, то определяют средний темп роста, а затем последнее известное значение показателя умножают (делят) на средний темп роста столько раз, на какой период осуществляется экстраполяции или интерполяция.

3. Основан на использовании аналитической формулы. Если известна теоретическая кривая, характеризующая изменение показателя за данный период, то в аналитическое выражение этой кривой подставляется условное значение времени и находится неизвестное значение показателя.

Рассмотрим эти способы на примере определения уровня показателя фондоотдачи в 1989г. и 1995г.

а) на основе определения ср. абс. прироста

$$y_{89} = y_{94} - \bar{\Delta y} \times 5,$$

$$y_{89} = 1,184 - 0,011 \times 5 = 1,129 \text{ (ТНГ./ТНГ.)}$$

$$y_{95} = y_{94} + \bar{\Delta y} \times 1$$

$$y_{95} = 1,184 + 0,011 = 1,195 \text{ (ТНГ./ТНГ.)}$$

б) на основе определения ср. темпа роста

$$y_{89} = y_{94} \times (\bar{T}_p)^{-5}$$

$$y_{89} = \frac{1,184}{1,01^5} = 1,127 \text{ (ТНГ./ТНГ.)}$$

$$y_{95} = y_{94} \times \bar{T}_p$$

$$y_{95} = 1,184 \times 1,01 = 1,196 \text{ (ТНГ./ТНГ.)}$$

в) на основе использования аналитической формулы

$$y_t = a_0 + a_1 \times t$$

Год	1990	1991	1992	1993	1994
t	-2	-1	0	1	2

Для нахождения параметров искомой прямой решим систему из двух уравнений, удовлетворяющих методу наименьших квадратов.

$$\begin{cases} n \times a_0 + a_1 \times \sum t = \sum y \\ a_0 \times \sum t + a_1 \times \sum t^2 = \sum y \times t \end{cases}$$

$$\begin{cases} 5 \times a_0 + a_1 \times 0 = 1,139 + 1,37 + 1,167 + 1,215 + 1,184 \\ a_0 \times 0 + a_1 \times 10 = 1,139 \times (-2) + 1,37 \times (-1) + 1,167 \times 0 + 1,215 \times 1 + 1,184 \times 2 \end{cases}$$

$$\begin{cases} 5 \times a_0 = 6,075 \\ a_1 \times 10 = -0,065 \end{cases}$$

$$\begin{cases} a_0 = 1,215 \\ a_1 = -0,0065 \end{cases}$$

Расчет анализа ряда динамики

Годы	y	t	y*t	t^2	y(t)
1990	1,139	-2	-2,278	4	1,228
1991	1,37	-1	-1,37	1	1,222
1992	1,167	0	0	0	1,215
1993	1,215	1	1,215	1	1,209
1994	1,184	2	2,368	4	1,202
1995		3	-0,065	10	1,196

$$y_t = 1,215 - 0,0065 \times t$$

$$y_{89} = 1,215 - 0,0065 \times (-3) = 1,23 \text{ (ТНГ./ТНГ.)}$$

$$y_{95} = 1,215 - 0,0065 \times 3 = 1,196 \text{ (ТНГ./ТНГ.)}$$

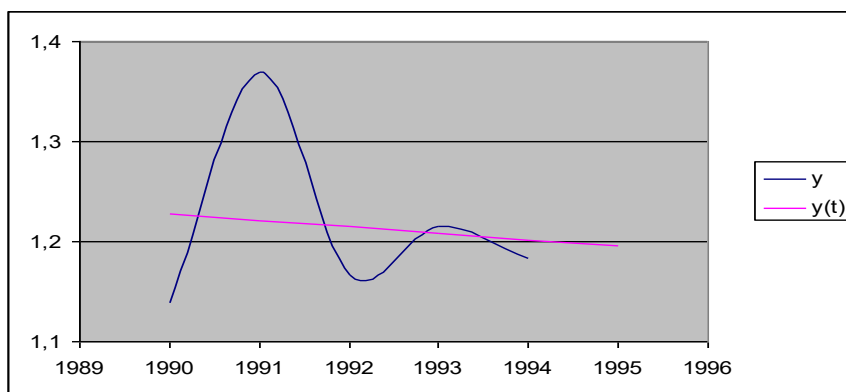


Рисунок 2. Исходные и теоретические значения показателя фондоотдачи

Проверка: В качестве проверки рассчитаем известный нам показатель ряда всеми тремя способами (например, уровень показателя фондоотдачи в 1990г.)

а) $y_{90} = 1,184 - 0,011 \times 4 = 1,14 \text{ (ТНГ./ТНГ.)}$

б) $y_{90} = \frac{1,184}{1,014} = 1,138 \text{ (ТНГ./ТНГ.)}$

в) $y_{90} = 1,215 - 0,0065 \times (-2) = 1,23 \text{ (ТНГ./ТНГ.)}$

Вывод: В данном случае различие данных по 1995г., полученных на основе трех вышеописанных методов, не существенно. По 1989г. значения, найденные с помощью двух первых методов (на основе среднего абсолютного прироста и среднего темпа роста) более схожи. В третьем случае (на основе теоретической кривой) в 1989г. заметно, небольшое различие. То же самое наблюдается и при проведении проверки. Это объясняется тем, что значения данного динамического ряда по годам изменяются неравномерно.

Пример 3. Смертность от болезней системы кровообращения в России за период 1995-2004 гг. характеризуется следующим рядом динамики.

Год	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Умершие, тыс. чел.	1163,5	1113,7	1100,3	1094,1	1187,8	1231,4	1253,1	1308,1	1330,5	1287,7

Вычислить: абсолютные, относительные, средние изменения и их темпы базисным и цепным способами. Проверить ряд на наличие в нем линейного тренда, на основе которого рассчитать интервальный прогноз на 2005 год с вероятностью 95%.

Решение. Любое изменение уровней ряда динамики определяется *базисным* (сравнение с первым уровнем) и *цепным* (сравнение с предыдущим уровнем) способами. Оно может быть *абсолютным* (разность уровней ряда) и *относительным* (соотношение уровней).

Базисное абсолютное изменение представляет собой разность конкретного и первого уровней ряда (21), а *цепное абсолютное изменение* представляет собой разность конкретного и предыдущего уровней ряда (22).

$$\Delta Y^B = Y_i - Y_1. \quad (21)$$

$$\Delta Y^C = Y_i - Y_{i-1}. \quad (22)$$

По знаку абсолютного изменения делается вывод о характере развития явления: при $\Delta Y > 0$ — рост, при $\Delta Y < 0$ — спад, при $\Delta Y = 0$ — стабильность.

В нашей задаче эти изменения определены в 3-м и 4-м столбцах таблицы **Error! Reference source not found..** Для проверки правильности расчетов применяется правило, согласно которому сумма цепных абсолютных изменений равняется последнему базисному. В нашей задаче это правило выполняется: $\sum \Delta Y^C = 124,2$ и $\Delta Y_{2004}^B = 124,2$.

Базисное относительное изменение представляет собой соотношение конкретного и первого уровней ряда (23), а *цепное относительное изменение* представляет собой соотношение конкретного и предыдущего уровней ряда (24).

$$i^B = \frac{Y_i}{Y_1}. \quad (23)$$

$$i^C = \frac{Y_i}{Y_{i-1}}. \quad (24)$$

Относительные изменения уровней — это по существу индексы динамики, критериальным значением которых служит 1. Если они больше ее, имеет место рост явления, меньше ее — спад, а при равенстве единице наблюдается стабильность явления.

В нашей задаче эти изменения определены в 5-м и 6-м столбцах таблицы 12.

Вычитая единицу из относительных изменений, получают *темпы изменения* уровней, критериальным значением которого служит 0. При положительном темпе изменения имеет место рост явления, при отрицательном — спад, а при нулевом темпе изменения наблюдается стабильность явления. В нашей задаче темпы изменения определены в 7-м и 9-м столбцах таблицы 12, а в 8-м и 10-м сделан вывод о характере развития изучаемого явления. Для проверки правильности расчетов применяется правило, согласно которому произведение цепных относительных изменений равняется последнему базисному. В нашей задаче это правило выполняется: $\prod i^{\text{Ц}} = 1,107$ и $i_{2004}^{\text{Б}} = 1,107$.

Таблица 12 - Вспомогательные расчеты для решения задачи

Год	Y	$\Delta Y^{\text{Б}}$	$\Delta Y^{\text{Ц}}$	$i^{\text{Б}}$	$i^{\text{Ц}}$	$T^{\text{Б}}$	хар-р	$T^{\text{Ц}}$	хар-р
1995	1163,5								
1996	1113,7	-49,8	-49,8	0,96	0,96	-0,04	спад	-0,04	спад
1997	1100,3	-63,2	-13,4	0,95	0,99	-0,05	спад	-0,01	спад
1998	1094,1	-69,4	-6,2	0,94	0,99	-0,06	спад	-0,01	спад
1999	1187,8	24,3	93,7	1,02	1,09	0,021	рост	0,086	рост
2000	1231,4	67,9	43,6	1,06	1,04	0,058	рост	0,037	рост
2001	1253,1	89,6	21,7	1,08	1,02	0,077	рост	0,018	рост
2002	1308,1	145	55	1,12	1,04	0,124	рост	0,044	рост
2003	1330,5	167	22,4	1,14	1,02	0,144	рост	0,017	рост
2004	1287,7	124	-42,8	1,11	0,97	0,107	рост	-0,03	спад
Итого	12070		124		1,11				

Обобщенной характеристикой ряда динамики является *средний уровень ряда* $\bar{y}$. Способ расчета $\bar{y}$ зависит от того, моментный ряд или интервальный (см. рис.3):



Рис.3. Методы расчета среднего уровня ряда динамики.

В нашей задаче ряд динамики интервальный, значит, применяем формулу средней арифметической простой **Error! Reference source not found.**: $\bar{y} =$

$12070,2 / 10 = 1207,02$ (тыс. чел.). То есть за период 1995-2004 в России в среднем за год от болезней системы кровообращения умирало 1207,02 тыс. чел.

Кроме среднего уровня в рядах динамики рассчитываются и другие средние показатели – среднее изменение уровней ряда (базисным и цепным способами), средний темп изменения.

Базисное среднее абсолютное изменение – это частное от деления последнего базисного абсолютного изменения на количество изменений уровней (25). *Цепное среднее абсолютное изменение* уровней ряда – это частное от деления суммы всех цепных абсолютных изменений на количество изменений (26).

$$\Delta \bar{Y}^B = \frac{\Delta Y_n^B}{n-1}. \quad (25)$$

$$\Delta \bar{Y}^C = \frac{\sum \Delta Y^C}{n-1}. \quad (26)$$

По знаку средних абсолютных изменений также судят о характере изменения явления в среднем: рост, спад или стабильность. Из правила контроля базисных и цепных абсолютных изменений следует, что базисное и цепное среднее изменение должны быть равными. В нашей задаче $\Delta \bar{Y} = 124,2/9 = 13,8$, то есть ежегодно в среднем смертность от болезней системы кровообращения растет на 13,8 тыс. чел.

Наряду со средним абсолютным изменением рассчитывается и среднее относительное. *Базисное среднее относительное изменение* определяется по формуле (27), а *цепное среднее относительное изменение* – по формуле (28):

$$\bar{i}^B = \sqrt[n-1]{i_n^B} = \sqrt[n-1]{Y_n/Y_1}. \quad (27)$$

$$\bar{i}^C = \sqrt[n-1]{\prod i_n^C}. \quad (28)$$

Естественно, базисное и цепное среднее относительное изменения должны быть одинаковыми и сравнением их с критериальным значением 1 делается вывод о характере изменения явления в среднем: рост, спад или стабильность. В нашей задаче $\bar{i} = \sqrt[9]{1,107} = 1,0114$, то есть ежегодно в среднем смертность от болезней системы кровообращения растет в 1,0114 раза.

Вычитанием 1 из среднего относительного изменения образуется соответствующий *средний темп изменения*, по знаку которого также можно судить о характере изменения изучаемого явления, отраженного данным рядом динамики. В нашей задаче $\bar{T} = 1,0114 - 1 = 0,0114$, то есть ежегодно в среднем смертность от болезней системы кровообращения растет на 1,14%.

Проверка ряда динамики на наличие в нем *тренда* (тенденции развития ряда) возможна несколькими способами (метод средних, Фостера и Стюарта, Валлиса и Мура и пр.), но наиболее простым является графическая модель, где на графике по оси абсцисс откладывается время, а по оси ординат – уровни ряда. Соединив полученные точки линиями, в большинстве случаев можно

выявить тренд визуально. Тренд может представлять собой прямую линию, параболу, гиперболу и т.п. В итоге приходим к трендовой модели вида:

$$Y_t = \hat{y}_t + \varepsilon_t, \quad (29)$$

где $\hat{y}_t$ – математическая функция развития; ε_t – случайное или циклическое отклонение от функции; t – время в виде номера периода (уровня ряда). Цель такого метода – выбор теоретической зависимости $\hat{y}_t$ в качестве одной из функций:

$$\hat{y}_t = a_0 + a_1 t - \text{прямая линия}; \quad \hat{y}_t = a_0 + \frac{a_1}{t} - \text{гипербола}; \quad \hat{y}_t = a_0 + a_1 t + a_2 t^2 - \text{парабола};$$

$$\hat{y}_t = a_0 t^{a_1} - \text{степенная}; \quad \hat{y}_t = a_0 + \sum_{k=1}^m (a_k \cos kt + b_k \sin kt) - \text{ряд Фурье}.$$

Для выявления *тренда* (тенденции развития ряда) в нашей задаче построим график $Y(t)$ (рис.4):

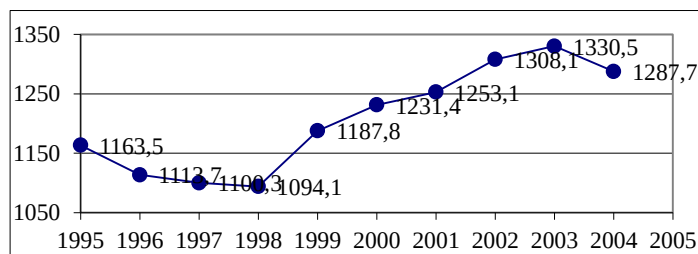


Рис.4. График динамики смертности от болезней системы кровообращения в РФ.

Из данного графика видно, что есть все основания принять уравнение тренда в виде линейной функции.

Определение параметров a_n в этих функциях может вестись несколькими способами, но самые незначительные отклонения аналитических (теоретических) уровней ($\hat{y}_t$ – читается как «игрек, выравненный по t ») от фактических (Y_t) дает *метод наименьших квадратов – МНК*. При этом методе учитываются все эмпирические уровни и должна обеспечиваться минимальная сумма квадратов отклонений эмпирических значений уровней Y_t от теоретических уровней $\hat{y}_t$:

$$\sum (\hat{y}_t - y)^2 \rightarrow \min. \quad (30)$$

В нашей задаче при выравнивании по прямой вида $\hat{y}_t = a_0 + a_1 t$ параметры a_0 и a_1 отыскиваются по МНК следующим образом. В формуле (29) вместо $\hat{y}_t$ записываем его конкретное выражение $a_0 + a_1 t$. Тогда $S = \sum (a_0 + a_1 t - y)^2 \rightarrow \min$. Дальнейшее решение сводится к задаче на экстремум, т.е. к определению того, при каком значении a_0 и a_1 функция двух переменных S может достигнуть минимума. Как известно, для этого надо найти частные производные S по a_0 и

a_1 , приравнять их к нулю и после элементарных преобразований решить систему двух уравнений с двумя неизвестными.

В соответствии с вышеизложенным найдем частные производные:

$$\begin{cases} \frac{\partial S}{\partial a_0} = 2 \sum (a_0 + a_1 t - y) = 0 \\ \frac{\partial S}{\partial a_1} = 2 \sum (a_0 + a_1 t - y) t = 0 \end{cases} \quad (31)$$

Сократив каждое уравнение на 2, раскрыв скобки и перенеся члены с y в правую сторону, а остальные – оставив в левой, получим систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases} \quad (48)$$

где n – количество уровней ряда; t – порядковый номер в условном обозначении периода или момента времени; y – уровни эмпирического ряда.

Эта система и, соответственно, расчет параметров a_0 и a_1 упрощаются, если отсчет времени ведется от середины ряда. Например, при *нечетном* числе уровней срединная точка (год, месяц) принимается за нуль. Тогда предшествующие периоды обозначаются соответственно $-1, -2, -3$ и т.д., а следующие за средним (центральным) – соответственно $1, 2, 3$ и т.д. При *четном* числе уровней два срединных момента (периода) времени обозначают -1 и $+1$, а все последующие и предыдущие, соответственно, через два интервала: $\pm 3, \pm 5, \pm 7$ и т.д.

При таком порядке отсчета времени (от середины ряда) $\sum t = 0$, поэтому, система нормальных уравнений упрощается до следующих двух уравнений, каждое из которых решается самостоятельно:

$$\begin{cases} na_0 = \sum y \Rightarrow a_0 = \frac{\sum y}{n} \\ a_1 \sum t^2 = \sum yt \Rightarrow a_1 = \frac{\sum yt}{\sum t^2} \end{cases} \quad (49)$$

Как видим, при такой нумерации периодов параметр a_0 представляет собой средний уровень ряда. Определим по формуле (49) параметры уравнения прямой, для чего исходные данные и все расчеты необходимых сумм представим в таблице 13.

Из таблицы получаем, что $a_0 = 12070,2/10 = 1207,02$ и $a_1 = 4195/330 = 12,7121$. Отсюда искомое уравнение тренда $\hat{y}_t = 1207,02 + 12,7121t$. В 6-м столбце таблицы 13 приведены трендовые уровни, рассчитанные по этому уравнению. Для иллюстрации построим график эмпирических (маркеры-кружочки) и трендовых уровней (рис.5).

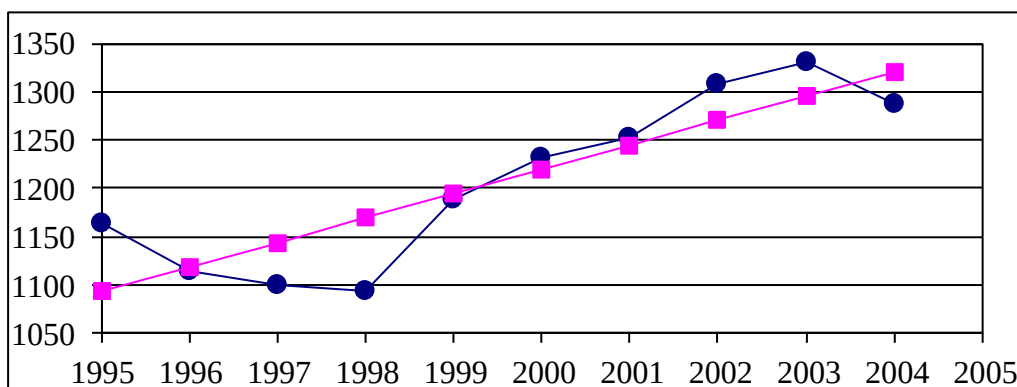


Рис.5. График эмпирических и трендовых уровней смертности от болезней системы кровообращения в РФ.

По полученной модели для каждого периода (каждой даты) определяются теоретические уровни тренда ($\hat{y}_t$) и оценивается *надежность (адекватность) выбранной модели тренда*. Оценку надежности проводят с помощью критерия Фишера, сравнивая его расчетное значение F_p с теоретическими значениями F_T (приложение 1). При этом расчетный критерий Фишера определяется по формуле:

$$F_p = \frac{(n-k)D_A}{(k-1)D_o}, \quad (32)$$

где k – число параметров (членов) выбранного уравнения тренда; D_A – аналитическая дисперсия, определяемая по формуле (34); D_o – остаточная дисперсия (35), определяемая как разность фактической дисперсии D_ϕ (33) и аналитической дисперсии:

$$D_\phi = \frac{\sum (y - \bar{y})^2}{n}; \quad (33)$$

$$D_A = \frac{\sum (\hat{y}_t - \bar{y})^2}{n}; \quad (34)$$

$$D_o = \frac{\sum (\hat{y}_t - y)^2}{n} = D_\phi - D_A. \quad (35)$$

Сравнение расчетного и теоретического значений критерия Фишера ведется обычно при уровне значимости α с учетом степеней свободы $\nu_1 = k - 1$ и $\nu_2 = n - k$. Уровень значимости α связан с вероятностью β следующей формулой $\beta = 1 - \alpha$. При условии $F_p > F_T$ считается, что выбранная математическая модель ряда динамики адекватно отражает обнаруженный в нем тренд.

Таблица 13 - Вспомогательные расчеты для решения задачи

Год	y	t	t^2	yt	$\hat{y}_t$	$(y - \hat{y}_t)^2$	$(\hat{y}_t - \bar{y})^2$	$(y - \bar{y})^2$
1995	1163,5	-9	81	-10471,5	1092,611	5025,263	13089,44	1893,9904
1996	1113,7	-7	49	-7795,9	1118,035	18,79354	7918,3033	8708,6224
1997	1100,3	-5	25	-5501,5	1143,459	1862,733	4039,9506	11389,1584
1998	1094,1	-3	9	-3282,3	1168,884	5592,592	1454,3822	12750,9264
1999	1187,8	-1	1	-1187,8	1194,308	42,35249	161,59803	369,4084
2000	1231,4	1	1	1231,4	1219,732	136,1394	161,59803	594,3844

Год	y	t	t^2	yt	$\hat{y}_t$	$(y - \hat{y}_t)^2$	$(\hat{y}_t - \bar{y})^2$	$(y - \bar{y})^2$
2001	1253,1	3	9	3759,3	1245,156	63,10136	1454,3822	2123,3664
2002	1308,1	5	25	6540,5	1270,581	1407,705	4039,9506	10217,1664
2003	1330,5	7	49	9313,5	1296,005	1189,915	7918,3033	15247,3104
2004	1287,7	9	81	11589,3	1321,429	1137,652	13089,44	6509,2624
Итого	12070,2	0	330	4195	12070,2	16476,25	53327,348	69803,596

Проверим тренд в нашей задаче на адекватность по формуле (32), для чего в 7-м столбце таблицы 13 рассчитан числитель остаточной дисперсии, а в 8-м столбце – числитель аналитической дисперсии. В формуле (32) можно использовать их числители, так как оба они делятся на число уровней n (n сократятся): $F_P = 53327,348 \cdot 8 / (16476,25 \cdot 1) = 25,893 > F_T$, значит, модель адекватна и ее можно использовать для прогнозирования ($F_T = 5,32$ находим по приложению 1 в 1-ом столбце [$\nu_1 = k - 1 = 1$] и 8-й строке [$\nu_2 = n - k = 8$]).

При составлении прогнозов уровней социально-экономических явлений обычно оперируют не точечной, а интервальной оценкой, рассчитывая так называемые *доверительные интервалы прогноза*. Границы интервалов определяются по формуле (36):

$$\hat{y}_t \pm t_\alpha \sigma_{\hat{y}}, \quad (36)$$

где $\hat{y}_t$ – точечный прогноз, рассчитанный по модели тренда; t_α – коэффициент доверия по распределению Стьюдента при уровне значимости α и числе степеней свободы $\nu = n - 1$ (приложение 2); $\sigma_{\hat{y}}$ – ошибка аппроксимации, определяемая по формуле (37):

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum (y - \hat{y}_t)^2}{n - k}}, \quad (37)$$

где y и $\hat{y}_t$ – соответственно фактические и теоретические (трендовые) значения уровней ряда динамики; n – число уровней ряда; k – число параметров (членов) в уравнении тренда.

Определим доверительный интервал в нашей задаче на 2005 год с уровнем значимости $\alpha = (1 - 0,95) = 0,05$. Для этого найдем ошибку аппроксимации по формуле (37): $\sigma_{\hat{y}} = \sqrt{16476,25 / (10 - 2)} = 45,38$. Коэффициент доверия по распределению Стьюдента $t_\alpha = 2,2622$ при $\nu = 10 - 1 = 9$.

Прогноз на 2005 с вероятностью 95% осуществим по формуле (36):

$Y_{2005} = (1207,02 + 12,7121 \cdot 11) \pm 2,2622 \cdot 45,38$ или $1244,19 < Y_{2005} < 1449,51$ (тыс.чел.).

5.2 Задачи и упражнения для самостоятельного решения

1. Имеются данные о площади посева в хозяйстве:

Годы	1993	1994	1995	1996	1997	1998
Площадь посева, га	684	704	758	737	749	743

Рассчитать и проанализировать показатели ряда динамики:

1. Абсолютный прирост (цепные и базисные);
2. Коэффициент роста (цепные и базисные);

3. Темп прироста;
4. Значение 1% прироста;
5. Средний уровень ряда.

2. Производство картофеля в регионе характеризуется следующими данными, млн. тонн:

Годы	Производство картофеля, млн. тонн
1990	84
1995	78
1996	83
1997	85
1998	82
1999	86
2000	89

Для анализа производства картофеля в регионе за 1995-2000 гг. вычислите:

1. Абсолютные приросты, темпы роста и темпы прироста по годам и к 1995 году. Полученные показатели представьте в таблице.
2. Среднегодовое производство картофеля.
3. Среднегодовой темп роста и прироста производства картофеля за 1995-2000 гг. и за 1990-1995 гг.
4. Постройте график производства картофеля в регионе за 1990-2000гг. Сделайте выводы.

3. Имеются два ряда динамики:

Год	Сумма капитальных вложений в постоянных ценах 1991 г., тыс. тнг.	Год	Сумма капитальных вложений в текущих ценах, тыс. тнг.	Коэффициент роста цен к 1991 г., раз.
1986	436	1992	3979	22,23
1987	464	1993	29624	264,5
1988	479	1994	75549	1349,1
1989	507	1995	131648	3872,0
1990	501	1996	152288	6621,2
1991	513	1997	144671	7614,3

Провести смыкание рядов динамики для последующего анализа.

4. По данному динамическому ряду провести выравнивание, используя способ наименьших квадратов:

Год	Фактический объем производства, млн. т.
1990	30,8
1991	34,3
1992	38,3
1993	33,7
1994	33,8
1995	39,9
1996	38,7
1997	37,0
1998	31,4

5. По статистическим данным по РК за 2010 – 2015 гг. вычислить: абсолютные, относительные, средние изменения и их темпы базисным и цепным способами. Проверить ряд на наличие в нем линейного тренда, на основе которого рассчитать интервальный прогноз на 2016 год с вероятностью 95%.

Год	Вариант									
	1	2	3	4	5	6	7	8	9	10
	Валовой сбор сахарной свеклы, млн. т.	Валовой сбор картофеля, млн. т.	Число заключенных браков, тыс.	Число построенных жилых домов, млн.м ²	Поголовье крупного рогатого скота, млн. голов (на конец года)	Производство мяса, млн. т.	Производство яиц, млрд. шт.	Численность населения, тыс. чел. (на начало года)	Среднегодовая численность занятых в экономике, тыс. чел.	Доля расходов на оплату ЖКХ в бюджете домохозяйств, %
2010	14,1	34	897,3	30,3	16,5	4,4	34,1	146890	64327	4,6
2011	14,6	35	1001,6	31,7	15,8	4,5	35,2	146304	64710	5,2
2012	15,7	32,9	1019,8	33,8	15,0	4,7	36,3	145649	65359	6,2
2013	19,4	36,7	1091,8	36,4	13,5	4,9	36,5	144964	65666	7,2
2014	21,8	35,9	979,7	41,0	12,1	5,0	35,8	144168	66407	7,7
2015	21,4	37,3	1066,4	43,6	11,1	4,9	36,8	143474	66939	8,3

Контрольные вопросы

1. Какие задачи решаются с помощью анализа рядов динамики?
2. Назовите виды рядов динамики.
3. В каких случаях используются аналитические показатели динамического ряда? Перечислите данные показатели.
4. Дайте общую характеристику средних показателей динамического ряда.
5. Какой метод расчета среднего темпа роста уровней ряда динамики вы знаете?
6. С какой целью рассчитывается средний темп прироста?
7. Что понимается под колебаниями уровней временного ряда?
8. Раскройте понятие «тренд» и объясните, с какой целью используется уравнение тренда.
9. В чем суть метода скользящей средней?
10. Назовите способы измерения сезонных колебаний. Как рассчитываются индексы сезонности?
11. Что такое автокорреляция?

Тема 6. Индексы

6.1 Методические указания и решения типовых задач

Индексы – это относительные величины, характеризующие соотношение во времени или пространстве социально-экономических явлений. При всём их разнообразии индексы можно разделить на два класса:

1. Индивидуальные, элементарные (i)
2. Сводные, сложные (I).

Под индивидуальными понимают относительные величины, характеризующие изменение во времени показателей, относящиеся к одному объекту, или сравнивающие размеры показателей для одновременно существующих однородных объектов:

$$i = a_i / a_0 \quad \text{или} \quad i = a / b, \quad (55)$$

где a_i , a_0 – показатель в отчетном и базисном периоде :

a , b – показатели, характеризующие однородные объекты.

Сводными (сложными) индексами называют относительные показатели, характеризующие изменения сложного явления в целом. Сложные индексы зависят от влияния двух или более факторов. В сложных индексах могут изменяться все показатели, влияющие на данное явление, но можно определить влияние только одного из факторов. В этом случае фактор, влияние которого изучается, будет меняться, т.е. индексироваться, другие же факторы, влияние которых устраняется, будут фиксироваться, т.е. оставаться неизменными.

Общий индекс затрат на производство основных видов продукции, выпускаемых данным цехом (предприятием) определяется по формуле:

$$I_{zq} = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_0} \quad (56)$$

где z_1 , z_0 – себестоимость единицы продукции в отчётном и базисном периоде, тыс. тен./т.

q_1 , q_0 – объём производства данного вида продукции в отчётном и базисном периоде, тыс. т.

Для того, чтобы определить влияние каждого из этих показателей в отдельности строят соответствующие индексы:

Индекс себестоимости:

$$I_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1} \quad (57)$$

где q_1 – объём производства продукции, тыс. т.,

z_1, z_0 – себестоимость единицы продукции соответственно в отчетном и базисном периоде, тыс. тнг./т.

Индекс физического объема:

$$I_q = \frac{\sum z_0 \cdot q_1}{\sum z_0 \cdot q_0} \quad (58)$$

где q_1, q_0 – объем производства продукции соответственно в отчетном и базисном периоде, тыс. т.,

z_0 – себестоимость единицы продукции, тыс. тнг./т.

Эти индексы образуют систему взаимосвязанных индексов, т.е.

$$I_{zq} = I_z \cdot I_q \quad (59)$$

где I_z – индекс себестоимости,

I_q – индекс физического объема.

Индекс товарооборота определяется по следующей формуле:

$$I_{pq} = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_0} \quad (60)$$

где q_1, q_0 – объем производства продукции соответственно в отчетном и базисном периоде, тыс. т.,

p_1, p_0 – цена за единицу продукции соответственно в отчетном и базисном периоде, тыс. тнг./т.

Для того, чтобы определить влияние каждого из этих показателей в отдельности строят соответствующие индексы:

Индекс цен:

$$I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \quad (61)$$

где q_1 – объем производства продукции, тыс. т.,

p_1, p_0 – цена за единицу продукции соответственно в отчетном и базисном периоде, тыс. тнг./т.

Индекс физического объема:

$$I_q = \frac{\sum p_0 \times q_1}{\sum p_0 \times q_0} \quad (62)$$

где q_1, q_0 – объем производства продукции соответственно в отчетном и базисном периоде, тыс. т.,

p_1 – цена за единицу продукции, тыс. тнг./т.

Эти индексы образуют систему взаимосвязанных индексов, т.е.

$$I_{pq} = I_p \cdot I_q \quad (63)$$

где I_p – индекс цен,

I_q - индекс физического объема.

Индекс трудоёмкости определяется по следующей формуле:

$$I_{tq} = \frac{\sum t_1 \cdot q_1}{\sum t_0 \cdot q_0} \quad (64)$$

где q_1, q_0 – объем производства продукции соответственно в отчетном и базисном периоде, тыс. т.,

t_1, t_0 – затраты труда на единицу продукции соответственно в отчетном и базисном периоде, час/т.

Для того, чтобы определить влияние каждого из этих показателей в отдельности строят соответствующие индексы:

Индекс затрат труда. Он характеризует влияние изменения затрат труда на единицу продукции на общее изменение трудоемкости. В нем будут изменяться затраты труда на единицу продукции, а объемный показатель фиксируется на уровне отчетного периода. Данный индекс показывает во сколько раз изменится трудоемкость за счет изменения затрат труда на единицу продукции:

$$I_t = \frac{\sum t_1 \cdot q_1}{\sum t_0 \cdot q_1} \quad (65)$$

где q_1 – объем производства продукции, тыс. т.,

t_1, t_0 – затраты труда на единицу продукции соответственно в отчетном и базисном периоде, час/т.

Индекс физического объема.

$$I_q = \frac{\sum t_0 \times q_1}{\sum t_0 \times q_0} \quad (66)$$

где q_1, q_0 – объем производства продукции соответственно в отчетном и базисном периоде, тыс. т.,

t_0 – затраты труда на единицу продукции в базисном периоде, час/т.

Эти индексы образуют систему взаимосвязанных индексов, т.е.

$$I_{tq} = I_t \times I_q \quad (67)$$

где I_t - индекс затрат труда,

I_q - индекс физического объема.

Пример 1. Показатели товарооборота и затрат на производство различных видов продукции

Основные виды продукции	Объем производства продукции, тыс. т.		Себестоимость ед. продукции, тыс. руб./т.		Цена за единицу, тыс. руб./т.		Затраты труда на единицу час/т.	
	Баз.	Отч.	Баз.	Отч.	Баз.	Отч.	Баз.	Отч.
А	109	108	4,0	4,1	5,4	5,4	4,0	3,5
Б	5,8	4,8	5,4	5,6	6,9	6,8	7,8	7,6
В	4,1	5,3	5,8	5,8	4,7	4,7	7,8	8,0
Г	25,6	20,5	16,4	16,5	17,8	17,6	8,5	8,0

На основании исходных данных, представленных в виде таблиц 11 рассчитаем индивидуальные индексы цен, себестоимости и физического объёма по различным видам продукции и цехам, а также общие индексы товарооборота, цен и себестоимости в целом для данного цеха или предприятия, индексы переменного состава структурных сдвигов и фиксированного состава по одному из основных видов продукции, выпускаемому разными цехами (производствами).

Решение. Индивидуальные индексы определяются по следующим формулам:

индекс цен – $i_p = p_1 / p_0$,

индекс себестоимости – $i_z = z_1 / z_0$,

индекс физического объёма – $i_q = q_1 / q_0$,

где i_p , i_z , i_q – соответственно индивидуальные индексы цены, себестоимости и объёма;

p_1 , z_1 , q_1 – соответственно цена, себестоимость и объём продукции в отчётном периоде;

p_0 , z_0 , q_0 – соответствующие показатели в базисном периоде.

$$i_{p(A)} = \frac{5,3}{5,4} \times 100\% = 98\%$$

$$i_{p(B)} = \frac{4,7}{4,7} \times 100\% = 100\%$$

$$i_{p(B)} = \frac{6,8}{6,9} \times 100\% = 98,6\%$$

$$i_{p(\Gamma)} = \frac{17,6}{17,8} \times 100\% = 98,9\%$$

$$i_{z(A)} = \frac{4,1}{4,0} \times 100\% = 102,5\%$$

$$i_{z(\hat{A})} = \frac{5,8}{5,8} \times 100\% = 100\%$$

$$i_{z(B)} = \frac{5,6}{5,4} \times 100\% = 103,7\%$$

$$i_{z(\Gamma)} = \frac{16,5}{16,4} \times 100\% = 100,6\%$$

$$i_{z(\text{цех1})} = \frac{12,1}{11,6} \times 100\% = 104,3\%$$

$$i_{z(\text{цех3})} = \frac{11,6}{11,4} \times 100\% = 101,8\%$$

$$i_{z(\text{цех2})} = \frac{12,0}{12,6} \times 100\% = 95,2\%$$

$$i_{q(A)} = \frac{108}{109} \times 100\% = 99\%$$

$$i_{q(B)} = \frac{5,3}{4,1} \times 100\% = 129,3\%$$

$$i_{q(B)} = \frac{4,8}{5,8} \times 100\% = 82,8\%$$

$$i_{q(\Gamma)} = \frac{20,5}{25,6} \times 100\% = 80,1\%$$

$$i_{q(\text{цех1})} = \frac{50}{40} \times 100\% = 125\%$$

$$i_{q(\text{цех3})} = \frac{104}{100} \times 100\% = 104\%$$

$$i_{q(\text{цех2})} = \frac{240}{200} \times 100\% = 120\%$$

Общий индекс затрат на производство основных видов продукции, выпускаемых данным цехом (предприятием):

$$I_{zq} = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_0}$$

$$I_{zq} = \frac{4,1 \times 108 + 5,6 \times 4,8 + 5,8 \times 5,3 + 16,5 \times 20,5}{4 \times 109 + 5,4 \times 5,8 + 5,8 \times 4,1 + 16,4 \times 25,6} \times 100\% = 92\%$$

Общий индекс себестоимости основных видов продукции:

$$I_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1}$$

$$I_z = \frac{4,1 \times 108 + 5,6 \times 4,8 + 5,8 \times 5,3 + 16,5 \times 20,5}{4 \times 108 + 5,4 \times 4,8 + 5,8 \times 5,3 + 16,4 \times 20,5} \times 100\% = 101,7\%$$

Общий индекс физического объёма:

$$I_q = \frac{\sum z_0 \cdot q_1}{\sum z_0 \cdot q_0}$$

$$I_q = \frac{4 \times 108 + 5,4 \times 4,8 + 5,8 \times 5,3 + 16,4 \times 20,5}{4 \times 109 + 5,4 \times 5,8 + 5,8 \times 4,1 + 16,4 \times 25,6} \times 100\% = 90,6\%$$

Проверка: Проверка основана на взаимосвязи данных индексов.

$$I_{zq} = I_z \times I_q$$

$$I_{zq} = \frac{101,7 \times 90,6}{100} = 92\%$$

Аналогичные расчёты проводим и по группе общих индексов товарооборота, цен и физического объёма.

$$I_{pq} = \frac{\sum p_1 \times q_1}{\sum p_0 \times q_0}$$

$$I_{pq} = \frac{5,3 \times 108 + 6,8 \times 4,8 + 4,7 \times 5,3 + 17,6 \times 20,5}{5,4 \times 109 + 6,9 \times 5,8 + 4,7 \times 4,1 + 17,8 \times 25,6} \times 100\% = 89,8\%$$

$$I_p = \frac{\sum p_1 \times q_1}{\sum p_0 \times q_1}$$

$$I_p = \frac{5,3 \times 108 + 6,8 \times 4,8 + 4,7 \times 5,3 + 17,6 \times 20,5}{5,4 \times 108 + 6,9 \times 4,8 + 4,7 \times 5,3 + 17,8 \times 20,5} \times 100\% = 98,5\%$$

$$I_q = \frac{\sum p_0 \times q_1}{\sum p_0 \times q_0}$$

$$I_q = \frac{5,4 \times 109 + 6,9 \times 5,8 + 4,7 \times 4,1 + 17,8 \times 25,6}{5,4 \times 108 + 6,9 \times 4,8 + 4,7 \times 5,3 + 17,8 \times 20,5} \times 100\% = 91,2\%$$

Проверка: Проверка основана на взаимосвязи данных индексов.

$$I_{pq} = I_p \times I_q$$

$$I_{pq} = \frac{98,5 \times 91,2}{100} = 89,8\%$$

И по группе общих индексов затрат труда на производство продукции, трудоёмкости и физического объёма:

$$I_{tq} = \frac{\sum t_1 \times q_1}{\sum t_0 \times q_0}$$

$$I_{tq} = \frac{3,5 \times 108 + 7,6 \times 4,8 + 8 \times 5,3 + 8 \times 20,5}{4 \times 109 + 7,8 \times 5,8 + 7,8 \times 4,1 + 8,5 \times 25,6} \times 100\% = 85\%$$

$$I_t = \frac{\sum t_1 \times q_1}{\sum t_0 \times q_1}$$

$$I_t = \frac{3,5 \times 108 + 7,6 \times 4,8 + 8 \times 5,3 + 8 \times 20,5}{4 \times 108 + 7,8 \times 4,8 + 7,8 \times 5,3 + 8,5 \times 20,5} \times 100\% = 90,6\%$$

$$I_q = \frac{\sum t_0 \times q_1}{\sum t_0 \times q_0}$$

$$I_q = \frac{4 \times 108 + 7,8 \times 4,8 + 7,8 \times 5,3 + 8,5 \times 20,5}{4 \times 109 + 7,8 \times 5,8 + 7,8 \times 4,1 + 8,5 \times 25,6} \times 100\% = 93,7\%$$

Проверка: Проверка основана на взаимосвязи данных индексов.

$$I_{tq} = I_t \times I_q$$

$$I_{tq} = \frac{90,6 \times 93,7}{100} = 85\%$$

Показатели затрат на производство основного вида продукции в различных цехах

Участок, цех, предприятие	Объем производства продукции, тыс. т.		Себестоимость единицы продукции, тыс. тнг./т.	
	Баз.	Отч.	Баз.	Отч.
Цех 1	40	50	11,6	12,1
Цех 2	200	240	12,6	12,0
Цех 3	100	104	11,4	11,6
.....				

Данные таблицы служат основой для расчёта индексов переменного, фиксированного состава и структурных сдвигов.

$$d_i = \frac{q_i}{\sum q_i}$$

$$d_1^{1y} = \frac{50}{50 + 240 + 104} = 0,127$$

$$d_1^{2y} = \frac{240}{50 + 240 + 104} = 0,609$$

$$d_1^{3y} = \frac{104}{50 + 240 + 104} = 0,264$$

$$d_0^{1y} = \frac{40}{40 + 200 + 100} = 0,118$$

$$d_0^{2y} = \frac{200}{40 + 200 + 100} = 0,588$$

$$d_0^{3y} = \frac{100}{40 + 200 + 100} = 0,294$$

$$I_{nc} = \frac{\sum z_1 \times d_1}{\sum z_0 \times d_0}$$

$$I_{nc} = \frac{12,1 \times 0,127 + 12 \times 0,609 + 11,6 \times 0,264}{11,6 \times 0,118 + 12,6 \times 0,588 + 11,4 \times 0,294} \times 100\% = 98,2\%$$

$$I_{\phi c} = \frac{\sum z_1 \times d_1}{\sum z_0 \times d_1}$$

$$I_{\phi c} = \frac{12,1 \times 0,127 + 12 \times 0,609 + 11,6 \times 0,264}{11,6 \times 0,127 + 12,6 \times 0,609 + 11,4 \times 0,264} \times 100\% = 98\%$$

$$I_{cmp} = \frac{\sum z_0 \times d_1}{\sum z_0 \times d_0}$$

$$I_{cmp} = \frac{11,6 \times 0,127 + 12,6 \times 0,609 + 11,4 \times 0,264}{11,6 \times 0,118 + 12,6 \times 0,588 + 11,4 \times 0,294} \times 100\% = 100,2\%$$

Проверка: Проверка основана на взаимосвязи данных индексов.

$$I_{nc} = I_{\phi c} \times I_{cmp}$$

$$I_{nc} = \frac{98 \times 100,2}{100} = 98,2\%$$

Абсолютная величина влияния изменения доли, удельного веса каждого предприятия (цеха) в общем объеме производства и изменения себестоимости на отдельных предприятиях на изменение средней себестоимости:

$$A_{nc} = \sum z_1 \times d_1 - \sum z_0 \times d_0$$

$$A_{\tilde{n}} = (12,1 \times 0,127 + 12 \times 0,609 + 11,6 \times 0,264) - (11,6 \times 0,118 + 12,6 \times 0,588 + 11,4 \times 0,294) = -0,222 \text{ руб.}$$

Абсолютная величина влияния изменения себестоимости на отдельных предприятиях на общее изменение средней себестоимости:

$$A_{\phi c} = \sum (z_1 - z_0) \times d_1$$

$$A_{\tilde{o}} = (12,1 - 11,6) \times 0,127 + (12 - 12,6) \times 0,609 + (11,6 - 11,4) \times 0,264 = -0,249 \text{ руб.}$$

Абсолютная величина влияния изменения доли, удельного веса каждого предприятия в общем объеме производства на изменение средней себестоимости:

$$A_{cmp} = \sum (d_1 - d_0) \times z_0$$

$$A_{\tilde{\delta}} = (0,127 - 0,118) \times 11,6 + (0,609 - 0,588) \times 12,6 + (0,264 - 0,294) \times 11,4 = 0,027 \text{ руб.}$$

Проверка: Проверка основана на взаимосвязи данных индексов.

$$A_{nc} = A_{\phi c} + A_{cmp}$$

$$\sum z_1 d_1 - \sum z_0 d_0 = \sum (z_1 - z_0) \times d_1 + \sum (d_1 - d_0) \times z_0$$

$$A_{\tilde{n}} = -0,249 + 0,027 = -0,222 \text{ руб.}$$

Пример 2. Динамика себестоимости и объема производства продукции характеризуется следующими данными:

Вид продукции	Выработано продукции за период, тыс.ед.		Себестоимость единицы продукции за период, руб.	
	базисный	отчетный	базисный	отчетный
Завод №1:				
ВН-25	7	7,4	150	180
НС-26	6,5	5,4	100	120
Завод №2:				
ВН-25	6,8	7,0	140	150

На основании имеющихся данных вычислите:

1. Для завода №1 (по двум видам продукции вместе):

- общий индекс затрат на производство продукции;
- общий индекс себестоимости продукции;
- общий индекс физического объема произведенной продукции.

Определите сумму изменения затрат в отчетном периоде по сравнению с базисным и разложите по факторам (за счет изменения себестоимости и за счет изменения физического объема продукции).

2. Для двух заводов вместе (по продукции ВН-25) определите:

- индекс себестоимости переменного состава;
- индекс себестоимости постоянного состава;

в) индекс изменения структуры. Сделайте выводы.

Решение. Рассмотрим вначале завод №1. Сформируем для него из исходных данных следующую таблицу:

Вид продукции	Выработано продукции за период, тыс.ед.		Себестоимость единицы продукции за период, тнг.	
	базисный q_i^0	отчетный q_i^1	базисный p_i^0	отчетный p_i^1
ВН-25, $i = 1$	7	7,4	150	180
НС-26, $i = 2$	6,5	5,4	100	120

Используя в качестве соизмерителя неизменные цены, получим следующую формулу для определения общего индекса физического объема произведенной продукции:

$$I_q = \frac{\sum_i q_i^1 p_i^0}{\sum_i q_i^0 p_i^0} = \frac{7,4 \cdot 150 + 5,4 \cdot 100}{7 \cdot 150 + 6,5 \cdot 100} = \frac{1650}{1700} = 0,9706, \text{ или } 97,06\%.$$

Общий индекс физического объема произведенной продукции определяется по формуле:

$$I_{pq} = \frac{\sum_i q_i^1 p_i^1}{\sum_i q_i^0 p_i^0} = \frac{7,4 \cdot 180 + 5,4 \cdot 120}{7 \cdot 150 + 6,5 \cdot 100} = \frac{1980}{1700} = 1,1647, \text{ или } 116,47\%.$$

Отсюда, используя взаимосвязь индексов, вычислим общий индекс себестоимости продукции:

$$I_p = \frac{I_{pq}}{I_q} = \frac{1,1647}{0,9706} = 1,20, \text{ или } 120,0\%.$$

Сумма изменения затрат в отчетном периоде по сравнению с базисным составила: $\Delta pq = \sum_i q_i^1 p_i^1 - \sum_i q_i^0 p_i^0 = 1980 - 1700 = 280$ тыс. тнг.

Разложим теперь эту сумму изменения затрат по факторам. Сумма изменения затрат в отчетном периоде по сравнению с базисным за счет изменения себестоимости составила:

$$\Delta pq(p) = \sum_i q_i^1 p_i^1 - \sum_i q_i^1 p_i^0 = 1980 - 1650 = 330 \text{ тыс. тнг.}$$

Сумма изменения затрат в отчетном периоде по сравнению с базисным за счет изменения физического объема продукции составила:

$$\Delta pq(q) = \sum_i q_i^1 p_i^0 - \sum_i q_i^0 p_i^0 = 1650 - 1700 = -50 \text{ тыс. тнг.}$$

Рассмотрим теперь оба завода вместе (по продукции ВН-25). Сформируем для них из исходных данных следующую таблицу:

Номер завода	Выработано продукции за период, тыс.ед.		Себестоимость единицы продукции за период, тнг.	
	базисный q_i^0	отчетный q_i^1	базисный p_i^0	отчетный p_i^1
1	7	7,4	150	180
2	6,8	7	140	150

Индекс себестоимости переменного состава представляет собой отношение двух взвешенных средних величин с переменными весами, характеризующее изменение индексируемого (осредняемого) показателя:

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\sum_i p_i^1 q_i^1}{\sum_i q_i^1} : \frac{\sum_i p_i^0 q_i^0}{\sum_i q_i^0} = \frac{180 \cdot 7,4 + 150 \cdot 7}{7,4 + 7} : \frac{150 \cdot 7 + 140 \cdot 6,8}{7 + 6,8} = \frac{165,417}{145,072} = 1,1402,$$

или 114,02%.

Индекс себестоимости постоянного состава представляет собой отношение двух взвешенных средних величин с одними и теми же весами:

$$I_p = \frac{\sum_i p_i^1 q_i^1}{\sum_i q_i^1} : \frac{\sum_i p_i^0 q_i^1}{\sum_i q_i^1} = \frac{\sum_i p_i^1 q_i^1}{\sum_i p_i^0 q_i^1} = \frac{180 \cdot 7,4 + 150 \cdot 7}{150 \cdot 7,4 + 140 \cdot 7} = \frac{2382}{2090} = 1,1397, \text{ или } 113,97\%.$$

Индекс изменения структуры равен:

$$I_{\bar{p} \rightarrow \bar{p}} = \frac{I_{\bar{p}}}{I_p} = \frac{1,1402}{1,1397} = 1,0005, \text{ или } 100,05\%.$$

Вывод.

1. По результатам отчетного периода рост затрат 1-го завода произошел исключительно за счет увеличения себестоимости продукции. Более того, за год наблюдалось незначительное сокращение затрат за счет уменьшения физического объема продукции

2. Изменение структуры выпуска продукции ВН-25в общем объеме практически не повлияло на увеличение себестоимости продукции по двум заводам.

Произошедший рост средней себестоимости вызван ростом себестоимости одновременно на двух заводах.

Пример 3. Имеются следующие данные о товарообороте коммерческого магазина:

Товарная группа	Продано товаров в фактических ценах за период, тыс.тнг.		Изменение цен в отчетном периоде по сравнению с базисным периодом, %
	базисный	отчетный	
Хлеб и хлебобулочные изделия	48	54	+15
Кондитерские изделия	68	69,2	без изменения
Цельномолочная продукция	38	42,3	+3

Вычислите:

1. Общий индекс товарооборота в фактических ценах.
2. Общий индекс цен.
3. Общий индекс физического объема товарооборота, используя взаимосвязь индексов. Сделайте выводы.

Решение. Используя исходные данные, и приняв цены в базисном периоде за 1, получим следующую таблицу:

Товарная группа	Продано товаров в фактических ценах за период, тыс. тнг.		Цены, усл. д. ед.	
	базисный, $q_i^0 p_i^0$	отчетный, $q_i^1 p_i^1$	базисный, $\tilde{p}_i^0$	отчетный, $\tilde{p}_i^1$
Хлеб и хлебобулочные изделия	48	54	1	1,15
Кондитерские изделия	68	69,2	1	1
Цельномолочная продукция	38	42,3	1	1,03

Общий индекс товарооборота в фактических ценах равен:

$$I_{pq} = \frac{\sum_i q_i^1 p_i^1}{\sum_i q_i^0 p_i^0} = \frac{54 + 69,2 + 42,3}{48 + 68 + 38} = \frac{165,5}{154} = 1,0747, \text{ или } 107,47\%.$$

Общий индекс цен равен:

$$I_p = \frac{\sum_i q_i^1 p_i^1}{\sum_i q_i^1 \tilde{p}_i^0} = \frac{\sum_i q_i^1 p_i^1}{\sum_i \frac{q_i^1 p_i^1}{\tilde{p}_i^0}} = \frac{165,5}{\frac{54}{1,15} + \frac{69,2}{1} + \frac{42,3}{1,03}} = \frac{165,5}{157,224} = 1,0526, \\ \text{или } 105,26\%.$$

Общий индекс физического объема товарооборота, используя взаимосвязь индексов, определим как:

$$I_p = \frac{I_{pq}}{I_q} = \frac{1,0747}{1,0526} = 1,0209, \text{ или } 102,09\%.$$

Вывод.

1. За отчетный год цены выросли на 5,26%.
2. За отчетный год физический объем товарооборота вырос на 2,09%.
3. За отчетный год товарооборот в фактических ценах вырос на 7,47%.

Пример 4. Имеются следующие данные о продажах торговой точкой двух видов товара:

Товар	Цена за кг, тенге		Объем продаж, тыс. кг	
	Январь	Февраль	Январь	Февраль
Апельсины	20	18	100	160
Бананы	22	25	150	120

Определить: 1) индивидуальные индексы цен, физического объема и выручки; 2) общие индексы цен, физического объема и выручки; 3) абсолютное изменение выручки за счет изменений цен, структурного сдвига и объемов продаж (для каждого фактора в отдельности) по всей продукции и по каждому товару в отдельности. По итогам расчетов сделать аргументированные выводы.

Решение. В основе решения задачи лежит формула (38):

$$Q = pq, \quad (38)$$

где p – цена товара, q – физический объем (количество), Q – выручка (товарооборот).

Применив формулу (38) к нашей задаче, рассчитаем выручку по каждому товару в январе (Q_{0j}) и феврале (Q_{1j}) в таблице 14.

Таблица 14 - Расчет выручки и ее изменения по каждому товару

Товар j	Январь Q_{0j}	Февраль Q_{1j}	Изменение выручки $\Delta Q_j = Q_{1j} - Q_{0j}$
Апельсины	$20 \cdot 100 = 2000$	$18 \cdot 160 = 2880$	880
Бананы	$22 \cdot 150 = 3300$	$25 \cdot 120 = 3000$	-300
Итого	5300	5880	580

Из таблицы видно, что абсолютное изменение общей выручки составило:
 $\Delta \sum Q = \sum Q_1 - \sum Q_0 = 5880 - 5300 = 580$ тыс. тенге, то есть она выросла на 580 тыс. тенге

Общий индекс изменения выручки равняется:

$I_Q = \sum Q_1 / \sum Q_0 = 5880 / 5300 = 1,1094$, то есть выручка от продажи фруктов увеличилась в 1,1094 раза или на 10,94% в феврале по сравнению с январем.

Определим индивидуальные индексы цен (i_p), физического объема (i_q), выручки (i_Q) и доли товара (i_d) по формуле (5), используя в качестве X_i цены (p), физический объем (q), выручки (Q) и доли товара ($d = q / \sum q$) каждого вида фруктов соответственно. Результаты расчетов представим в таблице 15.

Таблица 15- Расчет индивидуальных индексов

Индивидуальный индекс	апельсины	бананы
количества i_q	$160/100 = 1,6$	$120/150 = 0,8$
отпускных цен i_p	$18/20 = 0,9$	$25/22 = 1,136$
выручки i_Q	$2880/2000 = 1,44$	$3000/3300 = 0,909$
доли товара i_d	$(160/280)/(100/250) = 1,429$	$(120/280)/(150/250) = 0,714$

Правильность выполненных расчетов проверяется следующим образом:

1) общее изменение выручки должно равняться сумме ее частных (по каждому товару в отдельности) изменений: $\Delta \sum Q = 880 + (-300) = 580$ (тыс. тенге);

2) произведение факторных индивидуальных индексов по периодам должно равняться соответствующему индивидуальному индексу выручки: $i_{QA} = 1,6 \cdot 0,9 = 1,44$; $i_{QB} = 0,8 \cdot 1,136 = 0,909$.

Из таблицы видно, что в феврале по сравнению с январем:

– количество проданных апельсинов увеличилось в 1,6 раза или на 60%, а бананов – уменьшилось в 0,8 раза или на 20%;

– цена апельсинов понизилась в 0,9 раза или на 10%, а бананов – повысилась в 1,136 раза или на 13,6%;

– выручка по апельсинам выросла в 1,44 раза или на 44%, а по бананам – снизилась в 0,909 раза или на 9,1%;

– доля проданных апельсинов увеличилась в 1,429 раза или на 42,9%, а бананов – уменьшилась в 0,714 раза или на 28,6%.

Агрегатный общий индекс физического объема Ласпейреса определяется по формуле (39):

$$I_q^L = \frac{\sum q_1 p_0}{\sum q_0 p_0}. \quad (39)$$

В нашей задаче $I_q^L = \frac{160*20+120*22}{100*20+150*22} = 5840/5300 = 1,10189$, то есть количество проданных фруктов в базисных (январских) ценах выросло в 1,10189 раза или на 10,189% в феврале по сравнению с январем.

Агрегатный общий индекс цен Пааше рассчитывается по формуле (40):

$$I_p^L = \frac{\sum q_1 p_1}{\sum q_1 p_0}. \quad (40)$$

В нашей задаче $I_p^L = \frac{160*18+120*25}{5840} = 5880/5840 = 1,00685$, то есть цена проданных фруктов при объемах продаж отчетного (февральского) периода выросла в 1,00685 раза или на 0,685% в феврале по сравнению с январем.

Контроль осуществляется по формуле: $I_Q = I_q^L I_p^L = 1,10189*1,00685 = 1,1094$.

Агрегатный общий индекс цен Ласпейреса вычисляется по формуле (41):

$$I_p^L = \frac{\sum p_1 q_0}{\sum p_0 q_0}. \quad (41)$$

В нашей задаче $I_p^L = \frac{18*100+25*150}{5300} = 5550/5300 = 1,04717$, то есть цена проданных фруктов при объемах продаж базисного (январского) периода выросла в 1,04717 раза или на 4,717% в феврале по сравнению с январем.

Агрегатный общий количественный индекс Пааше рассчитывается по формуле (42):

$$I_q^L = \frac{\sum p_1 q_1}{\sum p_1 q_0}. \quad (42)$$

В нашей задаче $I_q^L = 5880/5550 = 1,05946$, то есть количество проданных фруктов в отчетных (февральских) ценах выросло в 1,05946 раза или на 5,946% в феврале по сравнению с январем.

Контроль осуществляется по формуле: $I_Q = I_p^L I_q^L = 1,04717*1,05946 = 1,1094$.

Средняя геометрическая величина определяется из индексов Ласпейреса и Пааше (по методике Фишера) по формуле (43) для количества товаров и по формуле (44) – для цен:

$$I_q^\Phi = \sqrt{I_q^L I_q^L}; \quad (43)$$

$$I_p^\Phi = \sqrt{I_p^L I_p^L}. \quad (44)$$

В нашей задаче $I_q^\Phi = \sqrt{1,10189*1,05946} = 1,0805$, то есть в среднем количество проданных фруктов выросло в 1,0805 раза или на 8,05%; $I_p^\Phi = \sqrt{1,04717*1,00685} = 1,0268$, то есть в среднем цена проданных фруктов выросла в 1,0268 раза или на 2,68%.

Далее выполняется факторный анализ общей выручки. В его основе лежит следующая трехфакторная мультипликативная модель выручки:

$$I_Q = I'_q I_d I_p'' , \quad (45)$$

где $I'_q = \frac{\sum q_1}{\sum q_0}$, I_d – индекс структурных сдвигов, показывающий как изменилась выручка под влиянием фактора изменения долей проданных фруктов в отчетном периоде по сравнению с базисным периодом. Он определяется по формуле (46):

$$I_d = \frac{\sum \frac{q_1}{\sum q_1} p_0}{\sum \frac{q_0}{\sum q_0} p_0} = \frac{\sum d_1 p_0}{\sum d_0 p_0} . \quad (46)$$

В нашей задаче $I_d = \frac{160/280*20+120/280*22}{100/250*20+150/250*22} = 0,9838$, то есть структурный сдвиг должен был уменьшить отчетную выручку в базисных ценах в 0,9838 раза или на 1,62%.

Тогда изменение выручки за счет изменения общего количества фруктов определяется по формуле (47):

$$\Delta \sum Q_q = (I'_q - 1) \sum Q_0 . \quad (47)$$

В нашей задаче $\Delta \sum Q_q = (1,12-1)*5300 = 636$ (тыс. тенге), то есть изменение количества проданных фруктов увеличило выручку на 636 тыс. тенге

Изменение общей выручки за счет структурных сдвигов находится по формуле (48):

$$\Delta \sum Q_d = I'_q (I_d - 1) \sum Q_0 . \quad (48)$$

В нашей задаче $\Delta \sum Q_d = 1,12*(0,9838-1)*5300 = -96$ (тыс. тенге), то есть структурный сдвиг в количестве проданных фруктов уменьшил выручку на 96 тыс. тенге

Изменение общей выручки за счет изменения отпускных цен рассчитывается по формуле (49):

$$\Delta \sum Q_p = I'_q I_d (I_p'' - 1) \sum Q_0 . \quad (49)$$

В нашей задаче $\Delta \sum Q_p = 1,12*0,9838*(1,00685-1)*5300 = 40$ (тыс. тенге), то есть изменение цен на фрукты увеличило выручку на 40 тыс. тенге

Контроль правильности расчетов производится по формуле (50), согласно которой общее изменение выручки равно сумме ее изменений за счет каждого фактора в отдельности.

$$\Delta \sum Q = \sum Q_1 - \sum Q_0 = \Delta \sum Q_q + \Delta \sum Q_d + \Delta \sum Q_p . \quad (50)$$

В нашей задаче $\Delta \sum Q = 636 + (-96) + 40 = 580$ тыс. тенге

Результаты факторного анализа общей выручки заносятся в последнюю строку факторной таблицы 16.

Таблица 16 - Результаты факторного анализа выручки

Товар j	Изменение выручки, тыс. тенге	В том числе за счет		
		количества продукта	структурных сдвигов	отпускных цен
А	880	240	960	-320
Б	-300	396	-1056	360
Итого	580	636	-96	40

Наконец, ведется факторный анализ изменения частной (по каждому j -му товару в отдельности) выручки на основе следующей трехфакторной мультипликативной модели:

$$Q_{1j} = I'_q i_{dj} i_{pj} Q_{0j}. \quad (51)$$

Тогда изменение частной выручки за счет каждого из 3-х факторов (количество, структурный сдвиг и цена) по j -му виду товара определяется соответственно по формулам (52) – (54).

$$\Delta Q_{qj} = (I'_q - 1)Q_{0j}; \quad (52)$$

$$\Delta Q_{dj} = I'_q (i_{dj} - 1)Q_{0j}; \quad (53)$$

$$\Delta Q_{pj} = I'_q i_{dj} (i_{pj} - 1)Q_{0j}. \quad (54)$$

Так, по апельсинам изменение выручки за счет первого фактора (изменения общего количества проданных фруктов) по формуле (52) равно:

$$\Delta Q_{qA} = (1,12 - 1) * 2000 = 240 \text{ (тыс. тенге)}.$$

$$\text{Аналогично по бананам: } \Delta Q_{qB} = (1,12 - 1) * 3300 = 396 \text{ (тыс. тенге)}$$

Контроль правильности расчетов:

$$\sum \Delta Q_{qj} = \Delta \sum Q_q, \text{ то есть } 240 + 396 = 636 \text{ (тыс. тенге)}.$$

Так, по апельсинам изменение выручки за счет второго фактора (структурных сдвигов в количестве проданных фруктов) по формуле (53) равно:

$$\Delta Q_{dA} = 1,12 * (1,429 - 1) * 2000 = 960 \text{ (тыс. тенге)}.$$

$$\text{Аналогично по бананам: } \Delta Q_{dB} = 1,12 * (0,714 - 1) * 3300 = -1056 \text{ (тыс. тенге)}.$$

Контроль правильности расчетов:

$$\sum \Delta Q_{dj} = \Delta \sum Q_d, \text{ то есть } 960 + (-1056) = -96 \text{ (тыс. тенге)}.$$

И, наконец, по апельсинам изменение выручки за счет 3-го фактора (изменения отпускной цены) по формуле (54) равно:

$$\Delta Q_{pA} = 1,12 * 1,429 * (0,9 - 1) * 2000 = -320 \text{ (тыс. тенге)}.$$

Аналогично по бананам: $\Delta Q_{pB} = 1,12 * 0,714 * (1,136 - 1) * 3300 = 360 \text{ (тыс. тенге)}$.

Контроль правильности расчетов:

$$\sum \Delta Q_{pj} = \Delta \sum Q_p, \text{ то есть } (-320) + 360 = 40 \text{ (тыс. тенге)}$$

Результаты факторного анализа частной выручки также заносятся в таблицу 16, в которой все числа оказались взаимно согласованными.

6.2 Задачи и упражнения для самостоятельного решения

1. Имеются следующие данные о выполнении работ специализированной строительной организацией за 2011-2014 гг. Объем работ (м²).

Вид работ	2011г.	2012г.	2013г.	2014г.	Сметная стоимость за 1 м ² , тнг
Штукатурные	50500	65000	78000	90000	60,0
Малярные	130000	140000	135000	150000	24,0
Устройство полов	5000	5500	6000	7000	288,0

Определите цепные и базисные индексы физического объема строительных работ: 1) индивидуальные индексы объема каждого вида работ; 2) агрегатные индексы общего объема работ. При расчете базисных индексов за базу сравнения примите объем работ за 2011 год.

2. Имеются следующие данные о выполнении СМР строительной фирмой:

Вид работ	Сметная стоимость работ в базисном периоде, тыс. тнг.	Увеличение (+) или уменьшение (-) объема работ в отчетном периоде по сравнению с базисным, %
Земляные	1200	Без изменений
Строительные	80000	+3
Монтаж металлоконструкций	48000	-2

Определите общий индекс физического объема СМР.

3. Определите индексы динамики себестоимости производства кирпича во втором квартале по сравнению с первым: 1) индивидуальные индексы себестоимости кирпича по каждому заводу; 2) общие индексы себестоимости кирпича переменного состава, фиксированного состава и влияния структурных сдвигов. Проверьте взаимосвязь исчисленных индексов. Производство и себестоимость кирпича по двум заводам характеризуется следующими данными:

Завод	Производство кирпича, тыс. шт.		Себестоимость 1 тыс.шт., тнг.	
	1 квартал	2 квартал	1 квартал	2 квартал
1	600	800	650	640
2	700	550	700	680

4. Имеются следующие данные о динамике себестоимости готовой строительной продукции по двум строительным объектам:

Строительный объект	Себестоимость 1 м ² тнг.		Площадь введенных в эксплуатацию объектов, тыс.м ²	
	Базисный период	Отчетный период	Базисный период	Отчетный период
1	6500	6450	6500	6450
2	6200	6100	6200	6100

Определите как изменилась в отчетном периоде по сравнению с базисным средняя себестоимость единицы готовой продукции в целом по двум объектам.

5. Средняя месячная заработная плата рабочих по двум строительным организациям характеризуется следующими данными:

Строительная организация	Средняя месячная зарплата, р.		Число рабочих	
	Базисный период	Отчетный период		Базисный период
1	3900	3920	100	120
2	3980	3890	90	80

Определите общие индексы средней заработной платы переменного и фиксированного состава, а также влияния структурных сдвигов. Разложите на слагаемые абсолютный прирост фонда заработной платы рабочих в отчетном периоде по сравнению с базисным за счет изменения числа рабочих и их среднего заработка.

6. Средняя себестоимость 1 м² площади введенных в эксплуатацию строительных объектов по двум СУ составила в базисном периоде 4950р., а в отчетном-4948,5 р. Индекс динамики себестоимости фиксированного состава по этим объектам равен 0,9.

Определите, как изменилась средняя себестоимость строительных объектов в отчетном периоде по сравнению с базисным за счет структурных сдвигов.

7. Имеются следующие данные о выполнении отделочных работ строительной организацией за отчетный год:

Вид работ	Объем работ по кварталам, м, ²				Сметная стоимость единицы работ, тнг.
	1	2	3	4	
Окраска дверей масляной краской	500	550	600	700	35
Окраска окон масляной краской	850	950	1000	1100	45

Определите базисные и цепные индексы физического объема работ: индивидуальные и общие. При вычислении базисных индексов за базу сравнения примите объем работ за 1 квартал.

8. Имеются следующие данные о динамике себестоимости готовой строительной продукции по трем строительным организациям:

Строительная организация	Себестоимость 1 м ² жилой площади, р.		Введенная в эксплуатацию жилая площадь, тыс. м ²	
	Базисный период	Отчетный период		Базисный период
1	5300	5400	40	35
2	5250	5200	30	50
3	5600	5600	55	52

Определите индексы себестоимости готовой строительной продукции в отчетном периоде по сравнению с базисным: 1) индивидуальные индексы себестоимости 1м² жилой площади по каждой организации; 2) общие индексы себестоимости переменного состава, фиксированного состава и влияние структурных сдвигов. Проверьте взаимосвязь исчисленных индексов.

9. На основании приведенных данных о строительстве жилого комплекса определите индекс физического объема выполнения капитальных работ:

Виды работ	Объем работ в сметных ценах		Изменение физического объема работ в отчетном периоде по сравнению с базисным, %
	Базисный период	Отчетный период	
Проектно-изыскательные	130	140	+43
Земляные	1640	1720	+14
Кирпичная кладка	3280	2780	-24
Отделочные	640	960	+38

10. Имеются следующие данные по двум строительно-монтажным фирмам за два квартала:

Показатели	Строительная фирма 1		Строительная фирма 2	
	2 квартал	3 квартал	2 квартал	3 квартал
Сметная стоимость СМР, тыс. тенге	1800	2380	2880	3040
Среднесписочная численность рабочих	130	135	140	138
Средняя фактическая продолжительность рабочего квартала, дней	58	52	60	56

Определите индивидуальные индексы среднедневной производительности труда по каждой строительной фирме, а так же общие индексы по двум строительным фирмам переменного, постоянного состава и влияния структурных сдвигов. Сделайте выводы.

11. Имеются данные о цене и количестве проданных товаров в магазине:

Товар	Цена, тнг. за 1 кг.		Количество проданных товаров		Стоимость проданной продукции, тыс. тнг.	
	Январь	Февраль	Январь	Февраль	Январь	Февраль
Колбаса вареная, кг.	95,9	117,9	1904	2017	182,6	237,8
Масло животное, кг.	71,9	74,8	370	414	26,6	30,9
Масло растительное, л.	28,8	30,9	539	566	15,5	17,5

Определить: Индивидуальные индексы цен, физического объема продукции, стоимости; Агрегатный индекс товарооборота; Агрегатный индекс физического объема продукции; Агрегатный индекс цен; Средние индексы цен и физического объема продукции; Индекс цен Ласпейерса.

12. Имеются следующие данные о продажах минимаркетом 3-х видов товаров (А, В и С): Определить: Индивидуальные индексы цен, физического объема и товарооборота; Общие индексы цен, физического объема и товарооборота; Абсолютные приросты товарооборота за счет изменений цен, структурного сдвига и объемов продаж (для каждого фактора в отдельности) по всей продукции и по каждому товару в отдельности.

Товар	Цена за единицу продукта, тенге		Объем продаж, тыс. штук	
	1 квартал	2 квартал	1 квартал	2 квартал
1 вариант				
A	102	105	205	195
B	56	51	380	423
C	26	30	510	490
2 вариант				
A	112	109	202	260
B	51	48	365	420
C	22	26	477	316
3 вариант				
A	99	103	198	182
B	55	59	370	361
C	20	18	502	456
4 вариант				
A	99	109	188	182
B	55	56	380	385
C	20	21	508	444
5 вариант				
A	120	110	170	220
B	60	58	350	390
C	19	20	550	490
Товар	Цена за единицу продукта, тенге		Объем продаж, тыс. штук	
	1 квартал	2 квартал	1 квартал	2 квартал
6 вариант				
A	130	125	138	198
B	50	56	339	264
C	20	21	613	511
7 вариант				
A	107	110	220	189
B	46	44	490	550
C	18	20	720	680
8 вариант				
A	95	98	264	197
B	48	50	360	294
C	26	25	448	640
9 вариант				
A	89	92	360	294
B	58	56	410	482
C	24	25	558	593
10 вариант				
A	120	125	150	108
B	44	46	513	461
C	16	19	891	550

По итогам расчетов сделать аргументированные выводы.

Контрольные вопросы

1. Как понимается роль индексного метода в статистических исследованиях?
2. Объясните разницу между индивидуальными и общими индексами.
3. Что такое агрегатный индекс?
4. Какова роль средних индексов?
5. Какие факторы положены в основу различия агрегатных индексов Ласпейреса и Пааше?
6. Какова роль индексов фиксированного состава?
7. Объясните принцип взаимосвязи индексов.
8. Чем отличается факторный индекс от результативного?
9. В чем принципиальное различие методов цепных и базисных индексов?
10. Какой принцип положен в основу последовательно-цепного метода?

Тема 7. Статистическое изучение взаимосвязей

7.1 Методические указания и решения типовых задач

Основные определения и решаемые задачи. Изучение механизма рыночных связей, взаимодействия спроса и предложения, влияние объема и состава предложения товаров на объем и структуру товарооборота, формирование товарных запасов, издержек обращения, прибыли и других качественных показателей имеет первостепенное значение для прогнозирования конъюнктуры рынка, рациональной организации торговых процессов и решения многих вопросов успешного ведения бизнеса.

Статистические методы позволяют проанализировать взаимосвязи, складывающиеся в общественных явлениях. Все взаимосвязи по своему характеру можно разбить на три группы: балансовые, компонентные, факторные.

Балансовая связь — характеризует зависимость между источниками формирования ресурсов (средств) и их использованием.

$$O_H + P = B + O_K$$

где O_H - остаток товаров на начало отчетного периода;

P - поступление товаров за период;

B - выбытие товаров в изучаемом периоде;

O_K - остаток товаров на конец отчетного периода.

Левая часть формулы характеризует предложение товаров

$(O_H + P)$, а правая часть - использование товарных ресурсов $(B + O_K)$.

Компонентные связи - показателей экономической деятельности характеризуются тем, что изменение статистического показателя определяется изменением компонентов, входящих в этот показатель, как множители:

$$a = b * c$$

В статистике экономической деятельности компонентные связи используются в индексном методе. Например, индекс товарооборота в фактических ценах I_{qp} представляет произведение двух компонентов — индекса товарооборота в сопоставимых ценах I_q и индекса цен I_p , т.е.

$$I_{qp} = I_p * I_q.$$

Важное значение компонентной связи состоит в том, что она позволяет определять величину одного из неизвестных компонентов:

$$I_q = \frac{I_{pq}}{I_p} \quad \text{или} \quad I_p = \frac{I_{pq}}{I_q}$$

Факторные связи характеризуются тем, что они проявляются в согласованной вариации изучаемых показателей. При этом одни показатели выступают как факторные, а другие - как результативные.

Метод анализа корреляций и регрессий — корреляционно-регрессионный анализ (КРА) заключается в построении и анализе экономико-математической модели в виде уравнения регрессии (корреляционной связи),

выражающего зависимость явления от определяющих его факторов: $y_x = f(x_1, x_2, \dots, x_n)$.

Если при сравнении вариации различных признаков обнаруживается закономерное изменение одного признака (результативного, следственного) под влиянием другого (факторного, причинного), то можно говорить о связи между ними. При этом связи могут быть функциональными или корреляционными.

Функциональные связи являются связями полными, жесткими между переменными x и y , то есть каждому возможному значению x сопоставлено в однозначное соответствие определенное значение y .

Корреляционные связи – связи соотносительные, неполные. Одному значению признака фактора соответствует несколько значений признака следствия. Связь проявляется лишь в изменении средних величин результативного признака. Средние величины результативного признака изменяются под влиянием изменения многих факторных признаков, некоторые из них могут быть и неизвестны.

Особенность корреляционных связей состоит в том, что они обнаруживаются не в единичных случаях, а в массе и требуют для своего исследования массовых наблюдений, другими словами проявление корреляционных связей связано с действием закона больших чисел.

Корреляционная связь является неполной, так как существует множество других факторов, влияющих на конкретное значение признака-следствия.

Следует отметить еще одну особенность корреляционных связей: они необратимы. Например, если производительность труда зависит от энерговооруженности, то это не значит, что энерговооруженность зависит от производительности.

Предварительный теоретический анализ должен доказать, что между признаком, который мы избираем как фактор, и признаком-следствием имеется причинная связь, а также по возможности установить форму этой связи.

В теории корреляции решаются две задачи:

- определить теоретическую форму связи (регрессионный анализ);
- измерить тесноту связи (корреляционный анализ).

По форме корреляционные связи бывают прямые и обратные, прямолинейные и криволинейные (линейные и нелинейные), однофакторные и многофакторные.

Корреляционно-регрессионный анализ заключается в построении и анализе статистической модели в виде уравнения регрессии, приближенно выражающей зависимость результативного признака от одного или нескольких признаков-факторов и в оценке степени тесноты связи.

Традиционные методы корреляционно-регрессионного анализа позволяют не только оценить тесноту связи, но и выразить эту связь аналитически. Применению корреляционно-регрессионного анализа должен предшествовать качественный, теоретический анализ исследуемого социально-экономического явления или процесса.

Связь между факторами аналитически выражается уравнениями:

- $\bar{y}_x = a_0 + a_1 x$ прямой;
- $\bar{y}_x = a_0 + \frac{a_1}{x}$ гиперболы;
- $\bar{y}_x = a_0 + a_1 x + a_2 x^2$ параболы;
- $\bar{y}_x = a_0 x^{a_1}$ степенной функции.

Параметр a_0 показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов. Параметр a_1 – коэффициент регрессии показывает, на сколько изменяется в среднем значение результативного признака при увеличении факторного на единицу. На основе этого параметра вычисляются коэффициенты эластичности по формуле:

$$\varepsilon = a_1 \frac{\bar{x}}{\bar{y}}.$$

Коэффициенты эластичности показывают изменение результативного признака в процентах в зависимости от изменения факторного признака на 1%.

Для определения параметров уравнений используется метод наименьших квадратов (МНК), на основании которого строится соответствующая система уравнений.

$$\begin{cases} a_1 \sum_{i=0}^n x_i^2 + a_0 \sum_{i=0}^n x_i = \sum_{i=0}^n x_i y_i \\ a_1 \sum_{i=0}^n x_i - n a_0 = \sum_{i=0}^n y_i \end{cases}$$

Степень тесноты связи измеряется с помощью линейного коэффициента корреляции, например, в случае наличия линейной зависимости между признаками, определяется по формуле:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \cdot \sigma_x \sigma_y}$$

где: n - число наблюдений.

Для практических вычислений при малом числе наблюдений, $n \leq (20 \div 30)$, линейный коэффициент корреляции удобнее вычислять по формуле:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}}$$

Линейный коэффициент принимает любые значения от $-1 \leq r \leq 1$. Чем ближе $|r|$ к единице, тем теснее связь между признаками.

При $r = \pm 1$ корреляционная связь представляет линейную функциональную зависимость, при этом все наблюдаемые значения располагаются на прямой.

При $r = 0$ корреляционная связь отсутствует, при этом линия регрессии параллельна оси ОХ.

Значимость найденного значения r_{xy} проверяется с помощью t -критерия Стьюдента по таблице приложения 1 с количеством степеней свободы df , которое исчисляется по формуле:

$$df = n - m - 1$$

где n – объем выборки;

m – количество факторов.

В данном случае $m=1$ и соответственно $df = n - 2$

Фактическое значение $t_{факт}$ вычисляется по формуле:

$$t_{факт} = r_{le} \sqrt{\frac{df}{1 - r_{le}^2}}$$

Найденное значение r_{xy} значимо с заданным уровнем значимости α , если выполняется условие

$$|t_{факт}| > t_{кр.д}(\alpha; df)$$

В случае значимости r_{xy} качественная оценка силы связи в зависимости от $|r_{xy}|$ будет следующей

По степени тесноты связи различают количественные критерии оценки тесноты связи (табл. 17).

Таблица 17 - Качественная характеристика силы связи между факторами

Значение $ r_{xy} $	Характеристика связи
Меньше 0,3	Отсутствует
0,3 - 0,7	Средняя
0,7 - 0,9	Высокая
0,9 - 0,99	Весьма высокая
1	Связь не статистическая, а функциональная

Множественная (многофакторная) регрессия. Изучение связи между тремя и более связанными между собой признаками носит название множественной (многофакторной) регрессии, описываемой функцией вида:

$$y_x = f(x_1, x_2, \dots, x_n).$$

Построение моделей множественной регрессии включает этапы:

- выбор формы связи (уравнение регрессии);
- отбор факторных признаков;
- обеспечение достаточного объема совокупности для получения несмещенных оценок.

Практика построения многофакторных моделей взаимосвязи показывает, что все реально существующие зависимости между социально-экономическими явлениями можно описать, используя пять типов моделей:

- линейную: $\bar{y}_{1,2,\dots,k} = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k$;
- степенную: $\bar{y}_{1,2,\dots,k} = a_0x_1^{a_1}x_2^{a_2}\dots x_k^{a_k}$;
- показательную: $\bar{y}_{1,2,\dots,k} = e^{a_0+a_1x_1+a_2x_2+\dots+a_kx_k}$;
- параболическую: $\bar{y}_{1,2,\dots,k} = a_0 + a_1x_1^2 + a_2x_2^2 + \dots + a_kx_k^2$;

- гиперболическую: $\bar{y}_{1,2,\dots,k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_k}{x_k}$.

Основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации. Нелинейные формы зависимости приводятся к линейным путем линеаризации.

Пример 1. По данным о стоимости основных фондов и объеме валовой продукции нужно определить уравнение связи и тесноту связи, сделать вывод табл. 101.

В первых двух столбцах табл. 18 приведены условные данные о стоимости основных фондов и объеме валовой продукции в остальных столбцах и последней строке таблицы содержатся расчетные показатели, необходимых при решении задачи.

Решение. Построения корреляционного поля позволяет, определить связь между признаками в данном случае связь между признаками линейная. Принимая для этой связи уравнение прямой линии, определим его параметры на основе метода наименьших квадратов, решив следующую систему нормальных уравнений.

$$\begin{cases} a_1 \sum_{i=0}^n x_i^2 + a_0 \sum_{i=0}^n x_i = \sum_{i=0}^n x_i y_i \\ a_1 \sum_{i=0}^n x_i - n a_0 = \sum_{i=0}^n y_i \end{cases}$$

Таблица 18 - Расчетная таблица

Стоимость основных производственных фондов, млн. тенге x	Объем валовой продукции, млн. тенге y	$x \cdot y$	x^2	y^2	$\bar{y}_x$
1	20	20	1	400	19,4
2	25	50	4	625	25,0
3	31	93	9	961	30,6
4	31	124	16	961	36,2
5	40	200	25	1600	41,8
6	56	336	36	3136	47,4
7	52	364	49	2704	53,0
8	60	480	64	3600	58,6
9	60	540	81	3600	64,2
10	70	700	100	4900	69,8
$\Sigma=55$	$\Sigma=445$	$\Sigma=2907$	$\Sigma=385$	$\Sigma=22487$	$\Sigma=446$

$$\begin{cases} 10a_0 + 55a_1 = 445 \\ 55a_0 + 385a_1 = 2907 \end{cases};$$

$$a_0 = 13,8; a_1 = 5,6;$$

Составим уравнение регрессии:

$$\bar{y}_x = 13,8 + 5,6x$$

Таким образом, параметр уравнения регрессии $a_1 = 5,6$ показывает, что с увеличением стоимости основных производственных фондов на 1 млн. тенге объем валовой продукции увеличивается в среднем на 5,6 млн. тенге, или с увеличением стоимости основных производственных фондов на 1% объем валовой продукции увеличивается на 0,69%.

Определим коэффициент эластичности:

$$\varepsilon = 5,6 * \frac{5,5}{44,5} = \frac{30,8}{44,5} = 0,6921.$$

Рассчитаем величину линейного коэффициента корреляции:

$$r = \frac{\frac{2907}{10} - \frac{55}{10} * \frac{445}{10}}{\sqrt{\left[\frac{22487}{10} - \left(\frac{445}{10} \right)^2 \right] * \left[\frac{385}{10} - \left(\frac{55}{10} \right)^2 \right]}} = 0,98$$

В соответствии с формулами $df = n - m - 1$, $t_{факт} = r_{xy} \sqrt{\frac{df}{1 - r_{xy}^2}}$ находим

следующие значения: $df = 10 - 2 = 8$, $t_{факт} = 0,98 * \sqrt{\frac{8}{1 - 0,98^2}} = 13,93$, $t_{факт} = 13,93$.

При уровне значимости $\alpha = 0,05$ и $df = 8$ из таблицы в Приложение 1 следует, что $t_{кр.д} = 2,306$. Поскольку $|t_{факт}| > t_{кр.д}$, то согласно формуле $|t_{факт}| > t_{кр.д}(\alpha; df)$ найденное значение $r = 0,98$ существенно и указывает на тесную связь между стоимостью основных производственных фондов (x) и объемом валовой продукции (y).

Пример 2. Для характеристики зависимости между товарооборотом и товарными запасами рассчитайте линейный коэффициент корреляции на основании данных:

№ торгового предприятия	Товарооборот, тыс. тенге	Товарные запасы, тыс. тенге
1	25	10
2	34	12
3	45	15
4	50	18
5	52	22
6	55	25
7	61	30
8	72	31

Рассчитать коэффициент корреляции по двум формулам:

Решение. В первых двух столбцах ниже приведенной таблице приведены условные данные о товарообороте и товарных запасах в остальных столбцах и последней строке таблицы содержатся расчетные показатели, необходимых при решении задачи.

Расчетная таблица

x	y	$x \cdot y$	x^2	y^2	$x - x_{\bar{n}\bar{d}}$	$y - y_{\bar{n}\bar{d}}$	$(x - x_{\bar{n}\bar{d}}) \cdot (y - y_{\bar{n}\bar{d}})$	$(x - x_{\bar{n}\bar{d}})^2$	$(y - y_{\bar{n}\bar{d}})^2$
25	10	250	625	100	-24,25	-10,38	251,59	588,06	107,64
34	12	408	1156	144	-15,25	-8,38	127,72	232,56	70,14
45	15	675	2025	225	-4,25	-5,38	22,84	18,06	28,89
50	18	900	2500	324	0,75	-2,38	-1,78	0,56	5,64
52	22	1144	2704	484	2,75	1,63	4,47	7,56	2,64
55	25	1375	3025	625	5,75	4,63	26,59	33,06	21,39
61	30	1830	3721	900	11,75	9,63	113,09	138,06	92,64
72	31	2232	5184	961	22,75	10,63	241,72	517,56	112,89
394	163	8814	20940	3763			786,25	1535,50	441,88

$$x_{\bar{n}\bar{d}} = 49,3$$

$$y_{\bar{n}\bar{d}} = 20,4$$

Рассчитаем коэффициент корреляции по первой формуле:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \cdot \sigma_x \sigma_y}$$

Для этого найдем среднее квадратическое отклонение (σ)

$$\sigma_x = \sqrt{\frac{\sum (x - x_{\bar{n}\bar{d}})^2}{n}} = \sqrt{\frac{1535,5}{8}} = 13,85$$

$$\sigma_y = \sqrt{\frac{\sum (y - y_{\bar{n}\bar{d}})^2}{n}} = \sqrt{\frac{441,88}{8}} = 7,43$$

$$r = \frac{786,25}{8 \cdot 13,85 \cdot 7,43} = 0,95$$

Рассчитаем коэффициент корреляции по второй формуле:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}}$$

$$r = \frac{8814 - \frac{394 \cdot 163}{8}}{\sqrt{\left(20940 - \frac{(394)^2}{8} \right) \cdot \left(3763 - \frac{(163)^2}{8} \right)}} = 0,95$$

Вывод:

Коэффициент корреляции составил 0,95, следовательно, зависимость между товарооборотом и товарными запасами характеризуется как прямая, тесная.

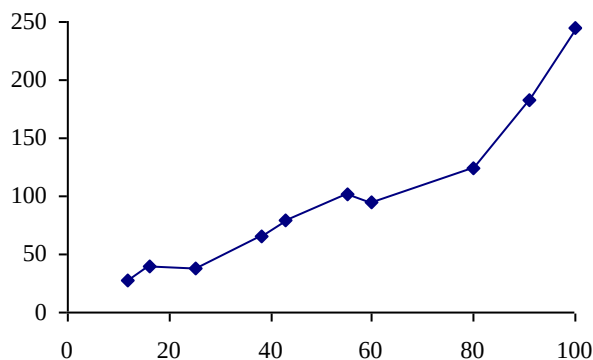
Пример 3. По условным данным таблицы 1 о стоимости основных фондов x и валовом выпуске продукции y (в порядке возрастания стоимости основных фондов) выявить наличие и характер корреляционной связи между признаками x и y .

Таблица 1 - Стоимость основных фондов и валовой выпуск по 10 однотипным предприятиям

Предприятия i	Основные производственные фонды, млн. тенге x_i	Валовой выпуск продукции, млн. тенге y_i	$x_i - \bar{x}$	$y_i - \bar{y}$
1	12	28	—	—
2	16	40	—	—
3	25	38	—	—
4	38	65	—	—
5	43	80	—	—
6	55	101	+	+
7	60	95	+	—
8	80	125	+	+
9	91	183	+	+
10	100	245	+	+
Итого	520	1000		

Решение. Для выявления наличия и характера корреляционной связи между двумя признаками в статистике используется **ряд методов**.

1. Графический метод, когда корреляционную зависимость для наглядности можно изобразить графически. Для этого, имея n взаимосвязанных пар значений x и y и пользуясь прямоугольной системой координат, каждую такую пару изображают в виде точки на плоскости с координатами x и y . Соединяя последовательно нанесенные точки, получают ломаную линию, именуемую *эмпирической линией регрессии* (см. рисунок справа). Анализируя эту линию, визуально можно определить характер зависимости между признаками x и y . В нашей задаче эта линия похожа на восходящую прямую, что позволяет выдвинуть гипотезу о наличии прямой зависимости между величиной основных фондов и валовым выпуском продукции.



2. Рассмотрение параллельных данных (значений x и y в каждой из n единиц). Единицы наблюдения располагают по возрастанию значений факторного признака x и затем сравнивают с ним (визуально) поведение результативного признака y . В нашей задаче в большинстве случаев по мере увеличения значений x увеличиваются и значения y (за несколькими исключениями – 2 и 3, 6 и 7 предприятия), поэтому, можно говорить о прямой связи между x и y (этот вывод подтверждает и эмпирическая линия регрессии). Теперь необходимо ее измерить, для чего рассчитывают несколько коэффициентов.

3. Коэффициент корреляции знаков (Фехнера) – простейший показатель тесноты связи, основанный на сравнении поведения отклонений

индивидуальных значений каждого признака (x и y) от своей средней величины. При этом во внимание принимаются не величины отклонений ($x_i - \bar{x}$) и ($y_i - \bar{y}$), а их знаки («+» или «-»). Определив знаки отклонений от средней величины в каждом ряду, рассматривают все пары знаков и подсчитывают число их совпадений (C) и несовпадений (H). Тогда коэффициент Фехнера рассчитывается как отношение разности чисел пар совпадений и несовпадений знаков к их сумме, т.е. к общему числу наблюдаемых единиц:

$$K_{\phi} = \frac{\sum C - \sum H}{\sum C + \sum H}. \quad (55)$$

Очевидно, что если знаки всех отклонений по каждому признаку совпадут, то $K_{\phi}=1$, что характеризует наличие прямой связи. Если все знаки не совпадут, то $K_{\phi}=-1$ (обратная связь). Если же $\sum C = \sum H$, то $K_{\phi}=0$. Итак, как и любой показатель тесноты связи, коэффициент Фехнера может принимать значения от 0 до ± 1 . Однако, если $K_{\phi}=1$, то это ни в коей мере нельзя воспринимать как свидетельство функциональной зависимости между x и y .

$$\text{В нашей задаче } \bar{x} = \frac{\sum x}{n} = \frac{520}{10} = 52; \quad \bar{y} = \frac{\sum y}{n} = \frac{1000}{10} = 100.$$

В двух последних столбцах таблицы 1 приведены знаки отклонений каждого x и y от своей средней величины. Число совпадений знаков – 9, а несовпадений – 1. Отсюда $K_{\phi} = \frac{9-1}{9+1} = 0,8$. Обычно такое значение показателя тесноты связи характеризует сильную зависимость, однако, следует иметь в виду, что поскольку K_{ϕ} зависит только от знаков и не учитывает величину самих отклонений x и y от их средних величин, то он практически характеризует не столько тесноту связи, сколько ее наличие и направление.

4. Линейный коэффициент корреляции применяется в случае линейной зависимости между двумя количественными признаками x и y . В отличие от K_{ϕ} в линейном коэффициенте корреляции учитываются не только знаки отклонений от средних величин, но и значения самих отклонений, выраженные для сопоставимости в единицах среднего квадратического отклонения t :

$$t_x = \frac{x - \bar{x}}{\sigma_x} \quad \text{и} \quad t_y = \frac{y - \bar{y}}{\sigma_y}.$$

Линейный коэффициент корреляции r представляет собой среднюю величину из произведений нормированных отклонений для x и y :

$$r = \frac{\sum \left(\frac{x - \bar{x}}{\sigma_x} \right) \left(\frac{y - \bar{y}}{\sigma_y} \right)}{n}, \quad (56)$$

$$\text{или} \quad r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}. \quad (57)$$

Числитель формулы (57), деленный на n , т.е. $\frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \overline{(x - \bar{x})(y - \bar{y})}$, представляет собой среднее произведение отклонений значений двух признаков от их средних значений, именуемое *ковариацией*. Поэтому можно сказать, что линейный коэффициент корреляции представляет собой частное от деления ковариации между x и y на произведение их средних квадратических отклонений. Путем несложных математических преобразований можно

получить и другие модификации формулы линейного коэффициента корреляции, например:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}. \quad (58)$$

Линейный коэффициент корреляции может принимать значения от -1 до $+1$, причем знак определяется в ходе решения. Например, если $\overline{xy} > \bar{x}\bar{y}$, то r по формуле (58) будет положительным, что характеризует прямую зависимость между x и y , в противном случае ($r < 0$) – обратную связь. Если $\overline{xy} = \bar{x}\bar{y}$, то $r = 0$, что означает отсутствие линейной зависимости между x и y , а при $r = 1$ – функциональная зависимость между x и y . Следовательно, всякое промежуточное значение r от 0 до 1 характеризует степень приближения корреляционной связи между x и y к функциональной. Таким образом, коэффициент корреляции при линейной зависимости служит как мерой тесноты связи, так и показателем, характеризующим степень приближения корреляционной зависимости между x и y к линейной. Поэтому близость значения r к 0 в одних случаях может означать отсутствие связи между x и y , а в других свидетельствовать о том, что зависимость не линейная.

В нашей задаче для расчета r построим вспомогательную таблицу 21

Таблица 21- Вспомогательные расчеты линейного коэффициента корреляции

i	x_i	y_i	$(x - \bar{x})^2$	$(y - \bar{y})^2$	t_x	t_y	$t_x t_y$	$(x - \bar{x})(y - \bar{y})$	$\overline{xy}$
1	12	28	1600	5184	-1,36526	-1,10032	1,502223	288	33,6
2	16	40	1296	3600	-1,22873	-0,91693	1,126667	216	64
3	25	38	729	3844	-0,92155	-0,9475	0,873167	167,4	95
4	38	65	196	1225	-0,47784	-0,53488	0,255587	49	247
5	43	80	81	400	-0,30718	-0,30564	0,093889	18	344
6	55	101	9	1	0,102394	0,015282	0,001565	0,3	555,5
7	60	95	64	25	0,273052	-0,07641	-0,02086	-4	570
8	80	125	784	625	0,955681	0,382056	0,365124	70	1000
9	91	183	1521	6889	1,331128	1,268425	1,688436	323,7	1665,3
10	100	245	2304	21025	1,638311	2,215924	3,630373	696	2450
Итого	520	1000	8584	42818			9,516166	1824,4	7024,4

В нашей задаче: $\sigma_x = \sqrt{8584/10} = 29,299$; $\sigma_y = \sqrt{42818/10} = 65,436$. Тогда по формуле (56) $r = 9,516166/10 = 0,9516$. Аналогичный результат получаем по формуле (57): $r = 1824,4/(29,299*65,436) = 0,9516$ или по формуле (58): $r = (7024,4 - 52*100) / (29,299*65,436) = 0,9516$, то есть связь между величиной основных фондов и валовым выпуском продукции очень близка к функциональной.

Проверка коэффициента корреляции на значимость (существенность). Интерпретируя значение коэффициента корреляции, следует иметь в виду, что он рассчитан для ограниченного числа наблюдений и подвержен случайным колебаниям, как и сами значения x и y , на основе которых он рассчитан.

Другими словами, как любой выборочный показатель, он содержит случайную ошибку и не всегда однозначно отражает действительно реальную связь между изучаемыми показателями. Для того, чтобы оценить существенность (значимость) самого r и, соответственно, реальность измеряемой связи между x и y , необходимо рассчитать среднюю квадратическую ошибку коэффициента корреляции σ_r . Оценка существенности (значимости) r основана на сопоставлении значения r с его средней квадратической ошибкой: $\frac{|r|}{\sigma_r}$.

Существуют некоторые особенности расчета σ_r в зависимости от числа наблюдений (объема выборки) – n .

1. Если число наблюдений достаточно велико ($n > 30$), то σ_r рассчитывается по формуле (59):

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}. \quad (59)$$

Обычно, если $\frac{|r|}{\sigma_r} > 3$, то r считается значимым (существенным), а связь – реальной. Задавшись определенной вероятностью, можно определить *доверительные пределы (границы)* $r = (r \pm t\sigma_r)$, где t – коэффициент доверия, рассчитываемый по интегралу Лапласа (см. таблицу **Error! Reference source not found.**).

2. Если число наблюдений небольшое ($n < 30$), то σ_r рассчитывается по формуле (60):

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}, \quad (60)$$

а значимость r проверяется на основе t -критерия Стьюдента, для чего определяется расчетное значение критерия по формуле (61) и сопоставляется с $t_{ТАБЛ}$.

$$t_{РАСЧ} = \frac{|r|}{\sigma_r} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \quad (61)$$

Табличное значение $t_{ТАБЛ}$ находится по таблице распределения t -критерия Стьюдента (см. приложение 2) при уровне значимости $\alpha = 1 - \beta$ и числе степеней свободы $\nu = n - 2$. Если $t_{РАСЧ} > t_{ТАБЛ}$, то r считается значимым, а связь между x и y – реальной. В противном случае ($t_{РАСЧ} < t_{ТАБЛ}$) считается, что связь между x и y отсутствует, и значение r , отличное от нуля, получено случайно.

В нашей задаче число наблюдений небольшое, значит, оценивать существенность (значимость) линейного коэффициента корреляции будем по формулам (60) и (61): $\sigma_r = \frac{\sqrt{1-0,9516^2}}{\sqrt{10-2}} = 0,3073/2,8284 = 0,1086$; $t_{РАСЧ} = \frac{|r|}{\sigma_r} = 0,9516/0,1086 = 8,7591$. При вероятности 95% $t_{табл} = 2,306$, а при вероятности 99% $t_{табл} = 3,355$, значит, $t_{РАСЧ} > t_{ТАБЛ}$, что дает возможность считать линейный коэффициент корреляции $r = 0,9516$ значимым.

5. Подбор уравнения регрессии представляет собой математическое описание изменения взаимно коррелируемых величин по эмпирическим (фактическим) данным. Уравнение регрессии должно определить, каким будет среднее значение результативного признака y при том или ином значении факторного признака x , если остальные факторы, влияющие на y и не связанные с x , не учитывать, т.е. абстрагироваться от них. Другими словами, уравнение регрессии можно рассматривать как вероятностную гипотетическую функциональную связь величины результативного признака y со значениями факторного признака x .

Уравнение регрессии можно также назвать *теоретической линией регрессии*. Рассчитанные по уравнению регрессии значения результативного признака называются *теоретическими*. Они обычно обозначаются $\bar{y}_x$ (читается: «игрек, выравненный по x ») и рассматриваются как функция от x , т.е. $\bar{y}_x = f(x)$. (Иногда для простоты записи вместо $\bar{y}_x$ пишут $\hat{y}$.)

Найти в каждом конкретном случае тип функции, с помощью которой можно наиболее адекватно отразить ту или иную зависимость между признаками x и y , — одна из основных задач регрессионного анализа. Выбор теоретической линии регрессии часто обусловлен формой эмпирической линии регрессии; теоретическая линия как бы сглаживает изломы эмпирической линии регрессии. Кроме того, необходимо учитывать природу изучаемых показателей и специфику их взаимосвязей.

Для аналитической связи между x и y могут использоваться следующие простые виды уравнений:

$$\hat{y} = a_0 + a_1 x - \text{прямая линия};$$

$$\hat{y} = a_0 + a_1 x + a_2 x^2 - \text{парабола};$$

$$\hat{y} = a_0 + \frac{a_1}{x} - \text{гипербола};$$

$$\hat{y} = a_0 a_1^x - \text{показательная функция};$$

$$\hat{y} = a_0 + a_1 \lg x - \text{логарифмическая функция и др.}$$

Обычно зависимость, выражаемую уравнением прямой, называют *линейной* (или *прямолинейной*), а все остальные — *криволинейными зависимостями*.

Выбрав тип функции, по эмпирическим данным определяют параметры уравнения. При этом отыскиваемые параметры должны быть такими, при которых рассчитанные по уравнению теоретические значения результативного признака $\bar{y}_x$ были бы максимально близки к эмпирическим данным.

Существует несколько методов нахождения параметров уравнения регрессии. Наиболее часто используется *метод наименьших квадратов* (МНК). Его суть заключается в следующем требовании: искомые теоретические значения результативного признака $\bar{y}_x$ должны быть такими, при которых бы обеспечивалась минимальная сумма квадратов их отклонений от эмпирических значений, т.е.

$$S = \sum (y - \bar{y}_x)^2 \rightarrow \min .$$

Поставив данное условие, легко определить, при каких значениях a_0 , a_1 и т.д. для каждой аналитической кривой эта сумма квадратов отклонений будет минимальной. Данный метод уже использовался нами в методических указаниях к теме 4 «Ряды динамики», поэтому, воспользуемся формулой (48) для нахождения параметров теоретической линии регрессии в нашей задаче, заменив параметр t на x .

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum yx \end{cases} \quad (62)$$

Исходные данные и все расчеты необходимых сумм представим в таблице 22.

Таблица 22 - Вспомогательные расчеты для решения задачи

i	x	y	$x*x$	$y*x$	y'	$(y_i - \bar{y})^2$	$(\bar{y}_x - \bar{y})^2$
1	12	28	144	336	15	5184	7225
2	16	40	256	640	23,5	3600	5852,25
3	25	38	625	950	42,625	3844	3291,891
4	38	65	1444	2470	70,25	1225	885,0625
5	43	80	1849	3440	80,875	400	365,7656
6	55	101	3025	5555	106,375	1	40,64063
7	60	95	3600	5700	117	25	289
8	80	125	6400	10000	159,5	625	3540,25
9	91	183	8281	16653	182,875	6889	6868,266
10	100	245	10000	24500	202	21025	10404
Итого	520	1000	35624	70244	1000	42818	38762,125

$$\begin{cases} 10a_0 + a_1 520 = 1000 \\ a_0 520 + a_1 35624 = 70244 \end{cases}; \quad \begin{cases} a_0 = 100 - 52a_1 \\ a_0 520 + a_1 35624 = 70244 \end{cases}; \quad (100 - 52a_1)520 + 35624a_1 = 70244;$$

$$52000 - 27040a_1 + 35624a_1 = 70244; \quad 8584a_1 = 18244; \quad a_1 = 2,125; \quad a_0 = 100 - 52 \cdot 2,125 = -10,5.$$

Отсюда искомая линия регрессии: $\hat{y}_t = -10,5 + 2,125x$. Для иллюстрации построим график эмпирической (маркеры-кружочки) и теоретической (маркеры-квадратики) линий регрессии.

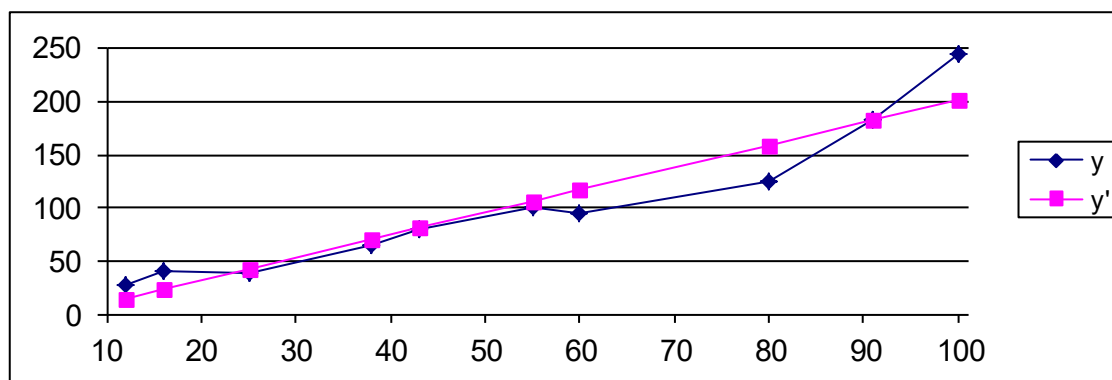


Рис.6. График эмпирической и теоретической линий регрессии.

6. Теоретическое корреляционное отношение представляет собой универсальный показатель тесноты связи. Измерить тесноту связи между коррелируемыми величинами – это значит определить, насколько вариация результативного признака обусловлена вариацией факторного признака. Ранее были рассмотрены показатели, с помощью которых можно выявить наличие корреляционной связи между двумя признаками x и y и измерить тесноту этой связи: коэффициент Фехнера и линейный коэффициент корреляции.

Наряду с ними существует универсальный показатель – *корреляционное отношение* (или *коэффициент корреляции по Пирсону*), применимое ко всем случаям корреляционной зависимости независимо от формы этой связи. Следует различать эмпирическое и теоретическое корреляционные отношения. *Эмпирическое корреляционное отношение* рассчитывается на основе правила сложения дисперсий как корень квадратный из отношения межгрупповой дисперсии к общей дисперсии, т.е.

$$\eta_{эмп} = \sqrt{\frac{\delta^2}{\sigma^2}}. \quad (63)$$

Теоретическое корреляционное отношение $\eta_{теор}$ определяется на основе выравненных (теоретических) значений результативного признака $\bar{y}_x$, рассчитанных по уравнению регрессии. $\eta_{теор}$ представляет собой относительную величину, получаемую в результате сравнения среднего квадратического отклонения в ряду теоретических значений результативного признака со средним квадратическим отклонением в ряду эмпирических значений. Если обозначить дисперсию эмпирического ряда игроков через σ_y^2 , а теоретического ряда – δ^2 , то каждая из них выразится формулами:

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}, \quad (64)$$

$$\delta^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n}. \quad (65)$$

Сравнивая вторую дисперсию с первой, получим *теоретический коэффициент детерминации*:

$$\eta_{теор}^2 = \frac{\delta^2}{\sigma_y^2} = \frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}, \quad (66)$$

который показывает, какую долю в общей дисперсии результативного признака занимает дисперсия, выражающая влияние вариации фактора x на вариацию y . Извлекая корень квадратный из коэффициента детерминации, получаем *теоретическое корреляционное отношение*:

$$\eta_{теор} = \sqrt{\frac{\delta^2}{\sigma_y^2}} = \sqrt{\frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}}. \quad (67)$$

Оно может находиться в пределах от 0 до 1. Чем ближе его значение к 1, тем теснее связь между вариацией y и x . При $\eta < 0,3$ говорят о малой зависимости между коррелируемыми величинами, при $0,3 < \eta < 0,6$ – о средней, при $0,6 < \eta < 0,8$ – о зависимости выше средней, при $\eta > 0,8$ – о большой, сильной зависимости. Корреляционное отношение применимо как для парной, так и для множественной корреляции независимо от формы связи. При линейной зависимости $\eta_{теор} \equiv r$.

В нашей задаче расчет необходимых сумм для использования в формуле (66) приведен в последних двух столбцах таблицы 22. Тогда теоретический коэффициент детерминации по формуле (66) равен: $\eta^2_{теор} = 38762,125 / 42818 = 0,9053$, то есть дисперсия, выражающая влияние вариации фактора x на вариацию y , составляет 90,53%.

Теоретическое корреляционное отношение по формуле (67) равно: $\eta_{теор} = \sqrt{0,9053} = 0,9515$, что совпадает со значением линейного коэффициента корреляции и, следовательно, можно говорить о большой, сильной зависимости между коррелируемыми величинами.

7.2 Задачи и упражнения для самостоятельного решения

1. Исследуйте графически с помощью поля корреляции связь фондовооруженности и производительности труда.

Показатель	Номер предприятия									
	1	2	3	4	5	6	7	8	9	10
Производительность работающих, тнг.	7560	8100	6110	7900	6610	7830	8400	7820	8560	7010
Фондовооруженность, тнг.	3930	4200	2800	4100	3500	4010	4600	4030	4730	3110

2. По десяти однородным семьям имеются следующие данные о доходах и потреблении молока за месяц (на одного члена семьи):

Номер семьи	1	2	3	4	5	6	7	8	9	10
Доход (х), тыс.тнг	14,7	16,4	18,9	34,5	14,56	22,65	25	36,78	21,34	24,56
Потребление молока (у), л	10	12	7	8	5	20	17	23	18	13

Рассчитать коэффициент корреляции и определить степень зависимости между доходом семьи и потреблением молока.

3. По десяти однородным семьям имеются следующие данные о доходах и расходах на промышленные товары за месяц (на одного члена семьи):

Номер семьи	1	2	3	4	5	6	7	8	9	10
Доход (х), тыс.тнг	14,7	16,4	18,9	34,5	14,56	22,65	25	36,78	21,34	24,56
Расходы на пром. товары (у), тыс.тнг	4,6	5,8	8,9	20	2,5	7,6	5,6	6,7	1,2	1,2

Рассчитать коэффициент корреляции и определить степень зависимости между доходом семьи и расходами на промышленные товары.

4. По десяти однородным семьям имеются следующие данные о доходах и расходах на продовольственные товары за месяц (на одного члена семьи):

Номер семьи	1	2	3	4	5	6	7	8	9	10
Доход (х), тыс.тнг	14,7	16,4	18,9	34,5	14,56	22,65	25	36,78	21,34	24,56
Расходы на прод. товары (у), тыс.тнг	7,6	8,8	11,5	21	8,5	9,6	8,6	7,7	4,2	5,6

Рассчитать коэффициент корреляции и определить степень зависимости между доходом семьи и расходами на продовольственные товары.

5. На основе исходных данных контрольных заданий по теме 2 определить наличие и характер корреляционной связи между признаками х и у 6-ю методами.

При- знак	Вариант									
	1	2	3	4	5	6	7	8	9	10
х	Рост	Доход	Возраст	IQ	Доход	Возраст	рост/вес	Стаж	Доход	IQ
у	Вес	Вес	Доход	Доход	Тетрадь	рост/вес	Кол-во друзей	Доход	Кол-во друзей	Время решения

Контрольные вопросы

1. Как вы понимаете сущность корреляционной связи? В чем ее отличие от функциональной связи?

2. Каковы признаки парной корреляции?

3. Что значит найти уравнение регрессии?

4. Какой вид имеет система нормальных уравнений?

5. С помощью, каких коэффициентов можно определить степень тесноты парной линейной зависимости?

6. В каких целях используются ранговые коэффициенты связи Спирмэна и Кендэла? Дайте их формулы.

7. Что такое коэффициент ассоциации и коэффициент контингенции? В чем особенности их расчета?

8. Какой коэффициент используется в случае существования небольшого объема исследуемой информации?

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

Учебники и учебные пособия:

1. Бельгибаева К.К. Сборник задач по социально-экономической статистике: Учебное пособие с методическими указаниями.- Алматы: Экономика, 2002.-159 с.
2. Елисеева И.И., Юзбашев М.М. Общая теория статистики /Учебник-М: Финансы и статистика, 2002.-366 с.
3. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред. Чл.-корр. РАН И.И.Елисеевой. – М.: Финансы и статистика, 1996.
4. Ефимова М.Р., Петров Е.В., Румянцев В.Н. Общая теория статистики –М. Инфра-М, 2000.- 414 с.
5. Ефимова М.Р., Рябушкин М.Р. Общая теория статистики. – СПб., 2002
6. Закон РК от 7 мая 1997 года № 98-І «О государственной статистике» (с изменениями и дополнениями, внесенными Законами РК от 30.01.01 г. №154-ІІ; от 15.01.02 г. №280 –ІІ; от 09.08.2002 г. №346- ІІ)
7. Молдакулова Г.М. Международная экономическая статистика: Учебное пособие.- Алматы: Экономика, 2004.- 172 с.
8. Октябрьский П.Я. Статистика: Учебник.- М.: ТК Велби, Изд-во Проспект, 2003.- 328 с
9. Словарь современной экономической теории Макмиллана. – М.: Инфра-М, 1997.
10. Социальная статистика. /Под ред. чл.-корр. АН И.И. Елисеевой. –М.: Финансы и статистика, 2002
11. Теория статистики: Учеб. для вузов / под ред. Р.А. Шмойловой - М: Финансы и статистика, 2001. – 415 с
12. Практикум по теории статистики: учеб.пособие/Шмойлова Р.А., Минашкин В.Г., Садовникова Н.А.; под ред. Шмойловой Р.А -3-е изд. – М.: Финансы и статистика, 2009. – 416с.
13. Шокаманов Ю.К., Бельгибаева К.К. Статистика: Учебник для студентов высших учебных заведений/Под ред. Ю.К. Шокаманова – Алматы: ТОО «Радгел», Экономика, 2008.-522с.
14. Харламов А.И. и др. Общая теория статистики. - М: Финансы и статистики, 1998.- 318 с.

Адреса Web-сайтов Агентства РК по статистике:

<http://www.oblstat.kost.kz/>

www.kostanai.stat.kz

www.stat.kz

Методические пособия:

1. Хасенова А.А. УМК по дисциплине «Статистика» для студентов экономических специальностей.– Костанай, КГУ им.А.Байтурсынова, 2007.
2. Хасенова А.А., Абдрахманова А.Б., Серкебаева Л.Т. Электронный учебник по статистике для студентов экономических специальностей.– Костанай, КГУ им.А. Байтурсынова, 2007 .

Журналы:

1. Казахстан и его регионы (Агентство РК по статистике), Алматы, 2016г.
2. Казахстан и страны СНГ (Агентство РК по статистике), Алматы, 2016г.
3. Костанайская область в цифрах: статистический справочник. Костанай, 2007г.
4. Основные показатели деятельности малых предприятий Костанайской области. Региональный статистический ежегодник Костанайской области, Костанай, 2015г., 2016г., 2017г.
5. Статистические ежегодники Казахстана, 2014г., 2015г., 2016г. – Алматы: Агентство Республики Казахстан, 2014г., 2015г., 2016г.
6. Статистические справочники: Национальные счета РК 1995-1998гг. – Алматы: Агентство РК по статистике, 2017г.
7. Статистический ежегодник Костанайской области (Управление статистики Костанайской области), Костанай, 2015г., 2016г., 2017г.
8. Статистическое обозрение Казахстана (Агентство РК по статистике), Алматы, 2016г..

Приложения

Приложение 1. Значения F-критерия Фишера при уровне значимости 0,05

	1	2	3	4	5	6	8	12	24	∞
1	161,5	200	215,7	224,6	230,2	234	238,9	243,9	249	254,3
2	18,5	19	19,16	19,25	19,3	19,33	19,37	19,41	19,45	19,5
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,9	2,71
10	4,96	4,1	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,2	3,09	2,95	2,79	2,61	2,4
12	4,75	3,88	3,49	3,26	3,11	3	2,85	2,69	2,5	2,3
13	4,67	3,8	3,41	3,18	3,02	2,92	2,77	2,6	2,42	2,21
14	4,6	3,74	3,34	3,11	2,96	2,85	2,7	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,9	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,2	2,96	2,81	2,7	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,9	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,1	2,87	2,71	2,6	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,3	3,44	3,05	2,82	2,66	2,55	2,4	2,23	2,03	1,78
23	4,28	3,42	3,03	2,8	2,64	2,53	2,38	2,2	2	1,76
24	4,26	3,4	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,6	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,3	2,13	1,93	1,67
28	4,2	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,7	2,54	2,43	2,28	2,1	1,9	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2	1,79	1,52
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,4	2,29	2,13	1,95	1,72	1,44
60	4	3,15	2,76	2,52	2,37	2,25	2,1	1,92	1,7	1,39
70	3,98	3,13	2,74	2,5	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,1	2,71	2,47	2,32	2,2	2,04	1,86	1,64	1,28
100	3,94	3,09	2,7	2,46	2,3	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,6	1,21
150	3,9	3,06	2,66	2,43	2,27	2,16	2	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,8	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,1
400	3,86	3,02	2,63	2,4	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3	2,61	2,38	2,22	2,1	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,6	2,37	2,21	2,09	1,94	1,75	1,52	

Приложение 2. Значения t-критерия Стьюдента
при уровне значимости α : 0,10, 0,05, 0,01

Число степеней свободы	α			Число степеней свободы	α		
	0,1	0,05	0,01		0,1	0,05	0,01
1	6,314	12,706	63,66	18	1,734	2,101	2,878
2	2,92	4,3027	9,925	19	1,729	2,093	2,861
3	2,353	3,1825	5,841	20	1,725	2,086	2,845
4	2,132	2,7764	4,604	21	1,721	2,08	2,831
5	2,015	2,5706	4,032	22	1,717	2,074	2,819
6	1,943	2,4469	3,707	23	1,714	2,069	2,807
7	1,895	2,3646	3,5	24	1,711	2,064	2,797
8	1,86	2,306	3,355	25	1,708	2,06	2,787
9	1,833	2,2622	3,25	26	1,706	2,056	2,779
10	1,813	2,2281	3,169	27	1,703	2,052	2,771
11	1,796	2,201	3,106	28	1,701	2,048	2,763
12	1,782	2,1788	3,055	29	1,699	2,045	2,756
13	1,771	2,1604	3,012	30	1,697	2,042	2,75
14	1,761	2,1448	2,977	40	1,684	2,021	2,705
15	1,753	2,1315	2,947	60	1,671	2	2,66
16	1,746	2,1199	2,921	120	1,658	1,98	2,617
17	1,74	2,1098	2,898	∞	1,645	1,96	2,576