

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Невинномысский технологический институт

Кафедра информационных систем, электропривода и автоматики

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАНЫХ

Методические указания к выполнению лабораторных работ
для направления подготовки 15.04.04 –
Автоматизация технологических процессов и производств

Невинномысск 2024

Настоящие методические указания предназначены для студентов направления 15.03.04 – Автоматизация технологических процессов и производств. Они разработаны в соответствии с федеральным государственным образовательным стандартом и образовательной программой направления.

В методических указаниях приведены необходимые теоретические сведения, определены требования к содержанию лабораторной работы, даны варианты заданий и приведен список рекомендуемых литературных источников.

Составитель *канд. техн. наук, доцент Д.В. Болдырев*

Отв. редактор *канд. техн. наук, доцент А.А. Евдокимов*

Лабораторная работа №1

ДИСПЕРСИОННЫЙ АНАЛИЗ

Цель работы – изучение способов исследования взаимосвязи данных, разбитых на группы, с помощью дисперсионного анализа.

Содержание работы – выполнение дисперсионного анализа, представление полученных результатов в форме отчета и доказательство их правильности.

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1 Понятие о дисперсионном анализе

Критерии различия (Стьюдента, Вилкоксона и т.п.) позволяют сравнить среднее для двух выборок. Если же требуется сравнить три или более выборки, использовать критерии различия не следует. Дело в том, что при попарном сравнении выборок при помощи критериев различия накапливается неконтролируемая ошибка. Поэтому для сравнения средних более чем в двух выборках используют специально разработанный метод, называемый дисперсионным анализом.

При помощи дисперсионного анализа обычно решают вопрос о влиянии одного (однофакторный дисперсионный анализ) или нескольких (многофакторный дисперсионный анализ) факторов на значение изучаемо признака. Например, типа корма на удои, региона проживания на продолжительность жизни, способа отбора проб на их достоверность и представительность и т.п. При этом рассматривается нулевая гипотеза: фактор не влияет на признак, средние выборок принадлежат одной генеральной совокупности. Если нулевая гипотеза отвергается при уровне значимости α , то с доверитель-

ной вероятностью $1-\alpha$ можно сделать вывод, что фактор влияет на признак.

В случае, если есть основания считать выборки нормально распределенными, а их дисперсии одинаковыми, выполняется собственно дисперсионный анализ, использующий F-критерий Фишера (параметрический). Следует отметить однако, что F-критерий по выборкам достаточно большого объема устойчив как к умеренным отклонениям от нормального закона, так и к умеренному различию в дисперсиях. Если возможность применения F-критерия все же вызывает сомнения, следует использовать непараметрические аналоги дисперсионного анализа, например критерий Краскала-Уоллиса.

1.2 Однофакторный дисперсионный анализ

При **равномерном** однофакторном дисперсионном анализе число замеров значений изучаемого признака на разных уровнях (при разных значениях) факторного признака одинаковое. Данные замеров сводятся в таблицу:

Таблица 1.1. Данные для равномерного однофакторного дисперсионного анализа

Номер измерения	Уровень фактора			
	A_1	A_2	...	A_p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
...
q	x_{q1}	x_{q2}	...	x_{qp}
Групповые средние	\bar{x}_{gp_1}	\bar{x}_{gp_2}	...	\bar{x}_{gp_p}

По этим данным рассчитываются следующие статистики:

1) Общая сумма квадратов отклонений наблюдаемых значений признака от общей средней \bar{x}

$$C_{общ} = \sum_{j=1}^p \sum_{i=1}^q (x_{ij} - \bar{x})^2 ;$$

факторная (межгрупповая) сумма квадратов отклонений групповых средних от общей средней, характеризующая рассеяние между группами:

$$C_{\text{факт}} = q \sum_{j=1}^p (x_{2p j} - \bar{x})^2 ;$$

остаточная (групповая) сумма квадратов отклонений наблюдаемых значений от своей групповой средней, характеризующая рассеяние внутри групп:

$$C_{\text{ост}} = C_{\text{общ}} - C_{\text{факт}} .$$

2) общая, факторная и остаточная дисперсии:

$$S_{\text{общ}}^2 = C_{\text{общ}} / (pq - 1) ; S_{\text{факт}}^2 = C_{\text{факт}} / (p - 1) ; S_{\text{ост}}^2 = C_{\text{ост}} / p(q - 1) ;$$

3) значение критерия Фишера:

$$F = S_{\text{факт}}^2 / S_{\text{ост}}^2 .$$

Значение критерия Фишера сравнивается с критическим для заданного уровня значимости α и числа степеней свободы $k_1 = p - 1$ и $k_2 = p(q - 1)$. Если $F > F_{\text{кр}}$, то гипотеза об отсутствии влияния фактора на признак отвергается с доверительной вероятностью $1 - \alpha$.

При **неравномерном** однофакторном дисперсионном анализе число измерений изучаемого признака на разных уровнях факторного признака различно. Обозначим q_j – количество измерений на уровне A_j . Тогда:

$$1) C_{\text{общ}} = [P_1 + P_2 + \dots + P_p] - [(R_1 + R_2 + \dots + R_p)^2 / n] ,$$

$$\text{где } P_j = \sum_{i=1}^{q_j} x_{ij}^2, \quad R_j = \sum_{i=1}^{q_j} x_{ij}, \quad n = q_1 + q_2 + \dots + q_p ;$$

$$C_{\text{факт}} = [(R_1^2 / q_1) + (R_2^2 / q_2) + \dots + (R_p^2 / q_p)] - [(R_1 + R_2 + \dots + R_p)^2 / n] ;$$

$$C_{\text{ост}} = C_{\text{общ}} - C_{\text{факт}} .$$

$$2) S_{\text{общ}}^2 = C_{\text{общ}} / (n - 1) ; S_{\text{факт}}^2 = C_{\text{факт}} / (p - 1) ; S_{\text{ост}}^2 = C_{\text{ост}} / (n - p) .$$

$$3) F = S_{\text{факт}}^2 / S_{\text{ост}}^2 ;$$

$$k_1 = p - 1 \text{ и } k_2 = n - p .$$

В таблицах Excel для проведения однофакторного равномерного или неравномерного дисперсионного анализа организуются вычисления по приведенным формулам с использованием функций СРЗНАЧ, СУММ, СУММКВ. Критическое значение критерия Фишера вычисляется при помощи функции FРАСПОБР. Также может быть использована процедура «Однофакторный дисперсионный анализ» из пакета анализа.

Пример 1.1. Проверить, существенны ли различия содержания загрязняющего вещества на трех уровнях (глубинах взятия проб)

№	Уровни замеров		
	1	2	3
1	1,17	2,28	1,80
2	1,52	2,46	2,38
3	1,90	0,88	2,62
4	1,76	2,03	2,91
5	1,54	1,22	1,60
6	0,63	2,29	2,83
7	2,30	1,80	2,13
8	1,32	1,79	2,06
9	0,94	1,61	2,23
10	1,15	2,30	3,06
11	0,75	2,60	1,86
12	2,49	1,76	1,92
13	2,14	2,14	2,16
14	1,62	2,73	2,27
15	1,40		

Решение. Откроем таблицы Excel и внесем данные из таблицы. Поскольку число измерений на разных уровнях неодинаково, требуется выполнить неравномерный дисперсионный анализ.

Вычислим в строке 17 объемы выборок: введем в ячейке B17 формулу

=СЧЁТ(B2:B16)

и «растянем» результат в ячейки C17:D17. При этом вычисляется число непустых ячеек в каждом столбце. Общее число измерений n вычислим, просуммировав результаты в ячейке E17 (функция СУММ).

В строке 18 вычислим величины P_i при помощи функции СУММКВ и в ячейке E18 их сумму. В строке 19 вычислим величины R_i при помощи функции СУММ и в ячейке E19 – их сумму. В строке 20 вычислим величины R_i^2/q_i и в ячейке E20 – их сумму.

В ячейке F2 вычислим значение $S_{\text{общ}}$, введя формулу

=E18-E19^2/E17.

В ячейке G2 – значения $S_{\text{факт}}$:

=E20-E19^2/E17.

И в ячейке H2 – значение $S_{\text{ост}}$:

=F2-G2.

Далее, в ячейке G5 вычисляем значение $S_{\text{факт}}^2$, учитывая, что $p=3$, и в ячейке H5 – значение $S_{\text{ост}}^2$, введя формулу

=H2/(E17-3).

Значение F-статистики вычислим в ячейке G8. Для вычисления критического значения выберем уровень значимости и внесем его в ячейку H8. Критическое значение F-критерия в ячейке I8 вычислим (с учетом того, что $p=3$), введя формулу

=ФРАСПОБР(H8;2;E17-3).

Итог вычислений выглядит следующим образом:

		Иллюстрации			Диаграммы					
И8		fx			=ФРАСПОБР(Н8;2;E17-3)					
	A	B	C	D	E	F	G	H	I	J
1	№	1	2	3		С общ	С факт	С ост		
2	1	1,17	2,28	1,8		14,59207	4,359944	10,23213		
3	2	1,52	2,46	2,38						
4	3	1,9	0,88	2,62			S^2 факт	S^2 ост		
5	4	1,76	2,03	2,91			2,179972	0,255803		
6	5	1,54	1,22	1,6						
7	6	0,63	2,29	2,83			F	α	F_{кр}	
8	7	2,3	1,8	2,13			8,522065	0,05	3,231727	
9	8	1,32	1,79	2,06						
10	9	0,94	1,61	2,23						
11	10	1,15	2,3	3,06						
12	11	0,75	2,6	1,86						
13	12	2,49	1,76	1,92						
14	13	2,14	2,14	2,16						
15	14	1,62	2,73	2,27						
16	15	1,4			Σ					
17	q	15	14	14	43					
18	P	38,3205	59,0941	74,8873	172,3019					
19	R	22,63	27,89	31,83	82,35					
20	R^2/q	34,14113	55,56086	72,36778	162,0698					
21										

Поскольку $F > F_{кр}$, делаем вывод, что различия на разных уровнях существенные.

Следует отметить, однако, что уверенности в обоснованности применения параметрического дисперсионного анализа у нас нет, поскольку нет оснований считать данные в выборках нормально распределенными, а объем выборок не позволяет проверить гипотезу о соответствии данных нормальному закону при помощи критерия χ^2 .

1.3 Однофакторный непараметрический дисперсионный анализ

Однофакторный непараметрический дисперсионный анализ производится при помощи критерия Краскала-Уоллиса (в русско-

язычной литературе его также называют критерием Краскала-Уоллеса, Крускала-Уоллеса). Этот критерий является многовыборочным обобщением критерия Уилкоксона (или Манна-Уитни).

Для применения критерия Краскала-Уоллеса следует проранжировать совмещенную выборку (из всех измерений при различных уровнях фактора); обозначим ранг i -го элемента выборки на j -м уровне фактора d_i^j . Далее находят суммы рангов R_1, \dots, R_p для каждого уровня фактора:

$$R_i = \sum_{j=1}^{q_i} d_i^j$$

При отсутствии связанных рангов статистика критерия Краскала-Уоллеса имеет вид:

$$H = \frac{12}{n(n+1)} \left(\frac{R_1^2}{q_1} + \dots + \frac{R_p^2}{q_p} \right) - 3(n+1)$$

При наличии связанных рангов используют модифицированную статистику:

$$H^* = \frac{H}{1 - \frac{1}{n^3 - n} \sum_{i=1}^k (t_i^3 - t_i)}, \text{ где } t_i \text{ — число элементов в } i\text{-й связке, } k \text{ —}$$

число связок.

Нулевая гипотеза (об отсутствии влияния фактора на признак) отклоняется, если рассчитанное значение критерия превышает критическое H_α для заданного уровня значимости. Для малых выборок ($p \leq 5$, $q_i \leq 8$) критические значения критерия Краскала-Уоллеса определяются по таблицам. При достаточно большом объеме выборки критическое значение определяется исходя из χ^2 распределения с $p-1$ степенями свободы. При организации вычислений в Excel для этого применяют функцию ХИ2ОБР($\alpha; p-1$).

Пример 4.2. Решить задачу 4.1. при помощи критерия Краскала-Уоллиса.

Решение. На новом листе Excel внесем в столбец А совмещенную выборку данных из таблицы 4.1, в столбце В для каждого данного укажем, к какому уровню факторного признака оно относится. Проведем предварительные расчеты как для критерия Вилкоксона: в столбце С вычислим ранги данных в совмещенной выборке при помощи функции РАНГ (введем формулу в ячейку С2, зафиксируем используемый массив, растянем); в столбце D вычислим длины связок t при помощи функции СЧЁТЕСЛИ. Поскольку имеются связки неединичной длины, далее в столбце Е вычислим согласованные ранги (согласованный ранг=ранг+($t-1$)/2); в столбце F – величины t^2-1 , которые затем сложим в ячейке G2 (обозначив величину, например, T). Результат вычислений имеет вид:

	A	B	C	D	E	F	G	H	I	J	K
1	данные	уровень	ранг	t	согл ранг	t^2-1	T				
2	1,17	1	6	1	6	0	24				
3	1,52	1	10	1	10	0					
4	1,9	1	21	1	21	0					
5	1,76	1	15	2	15,5	3					
6	1,54	1	11	1	11	0					
7	0,63	1	1	1	1	0					
8	2,3	1	33	2	33,5	3					
9	1,32	1	8	1	8	0					
10	0,94	1	4	1	4	0					
11	1,15	1	5	1	5	0					
12	0,75	1	2	1	2	0					
13	2,49	1	37	1	37	0					
14	2,14	1	26	2	26,5	3					
15	1,62	1	14	1	14	0					
16	1,4	1	9	1	9	0					
17	2,28	2	31	1	31	0					

Вычислим в ячейке I2 величину R_1 – сумму согласованных рангов данных, относящихся к 1 уровню факторного признака:

=СУММЕСЛИ(\$B2:\$B44;1;\$E2:\$E44)

В этой формуле массивы зафиксированы для того, чтобы можно было скопировать ее в ячейки J2 и K2. Сделав это и заменив в формулах значение критерия «1» на «2» и «3» соответственно, получаем значения R_2 и R_3 :

К2											
fx =СУММЕСЛИ(\$B2:\$B44;3;\$E2:\$E44)											
	A	B	C	D	E	F	G	H	I	J	K
1	данные	уровень	ранг	t	согл ранг	t^2-1	T		R1	R2	R3
2	1,17	1	6	1	6	0	24		203,5	334	408,5
3	1,52	1	10	1	10	0					
4	1,9	1	21	1	21	0					
5	1,76	1	15	2	15,5	3					
6	1,54	1	11	1	11	0					
7	0,63	1	1	1	1	0					
8	2,3	1	33	2	33,5	3					
9	1,32	1	8	1	8	0					
10	0,94	1	4	1	4	0					
11	1,15	1	5	1	5	0					
12	0,75	1	2	1	2	0					
13	2,49	1	37	1	37	0					
14	2,14	1	26	2	26,5	3					
15	1,62	1	14	1	14	0					
16	1,4	1	9	1	9	0					
17	2,28	2	31	1	31	0					

Значение исправленной статистики Краскала-Уоллиса можно вычислить теперь по приведенной выше формуле, но она достаточно громоздка, поэтому для облегчения вычислений проведем вспомогательные построения: укажем объемы выборок $q_1=15$, $q_2=14$, $q_3=14$; вычислим их сумму, получив объем совмещенной выборки n ; вычислим величины R_i^2/q_i (это удобно сделать, введя формулу один раз и скопировав в соседние ячейки) и их сумму.

Теперь вычислим значение исправленной статистики Краскала-Уоллиса, введя в ячейке I7 формулу:

$$=(12*L4/(L3*(L3+1))-3*(L3+1))/(1-G2/(L3^3-L3)).$$

Поскольку объемы выборок превышают 8, в таблицах критическое значение критерия найти нельзя; вычислим его при помощи функции ХИ2ОБР:

	D	E	F	G	H	I	J	K	L	M	N
1	t	согл ранг	t ² -1	T		R1	R2	R3			
2	1	6	0	24		203,5	334	408,5	Σ		
3	1	10	0		q	15	14	14	43		
4	1	21	0		R ² /q	2760,817	7968,286	11919,45	22648,55		
5	2	15,5	3								
6	1	11	0			H	H кр				
7	1	1	0			11,65182	5,991465				
8	2	33,5	3								
9	1	8	0								
10	1	4	0								
11	1	5	0								
12	1	2	0								
13	1	37	0								
14	2	26,5	3								

Поскольку рассчитанное значение критерия превышает критическое для уровня значимости 0,05, значит с доверительной вероятностью 0,95 можно сделать вывод о влиянии фактора на исследуемый признак, то есть различия в загрязнении на разных уровнях существенны. Заметим, что этот результат согласуется с выводом дисперсионного анализа, проведенного в примере 1.1.

2 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

До выполнения работы необходимо самостоятельно изучить правила выполнения дисперсионного анализа.

При выполнении работы необходимо решить следующие задачи.

1) Проверить, существенны ли различия уровня безработицы в разных регионах России по данным 2006 года (первая колонка данных) при помощи дисперсионного анализа и критерия Краскала-Уоллиса

2) Выбрать подходящий метод и проверить, существенны ли различия в уровне безработицы в России в 2006, 2005, 2004 годах

Содержание отчета о лабораторной работе:

- общая постановка задачи;
- таблицы дисперсионного анализа;
- заключения и выводы.

Варианты заданий приведены в приложении А.

КОНТРОЛЬНЫЕ ВОПРОСЫ

- 1 Что такое дисперсионный анализ?
- 2 Как вычислить p -значение для F -теста?
- 3 Однофакторный (односторонний) дисперсионный анализ. Требования для проведения данного анализа в Microsoft Excel относительно категориальных данных.
- 4 Что такое сумма квадратов ошибок?
- 5 Дайте определения терминам «внутригрупповая сумма квадратов» и «межгрупповая сумма квадратов».
- 6 Однофакторный дисперсионный анализ и анализ регрессии. Индикаторы.
- 7 Двухфакторный дисперсионный анализ. Модель влияния для двухфакторного дисперсионного анализа.
- 8 Как определить величину достоверности аппроксимации

R2?

9 Что такое сбалансированные данные?

ЛИТЕРАТУРА

1. Математика в примерах и задачах: учеб. пособие/ Л. Н. Журбенко и др. - М.: ИНФРА-М, 2012.
2. Богомолов Н. В. Математика: учебник/ Н. В. Богомолов, Самойленко П. И. – М.: ЮРАЙТ, 2013

Лабораторная работа №2

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Цель работы – изучение способов исследования взаимосвязи данных, разбитых на группы, с помощью дисперсионного анализа.

Содержание работы – выполнение дисперсионного анализа, представление полученных результатов в форме отчета и доказательство их правильности.

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Явления внешнего мира представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих явлений необходимо выявлять главные, основные причины, абстрагируясь от второстепенных. При изучении конкретных зависимостей одни признаки выступают в качестве факторов, обуславливающих изменение других признаков. Признаки этой группы называются *факторными*, а признаки, которые являются результатом влияния этих факторов, называются *результативными*. Связи между явлениями и их признаками классифицируются по степени тесноты связи, направлению и аналитическому выражению.

В статистике различают функциональную связь и стохастическую зависимость. *Функциональной* называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений, то такая зависимость называется *стохастической*.

Частным случаем стохастической связи является корреляционная связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

По степени тесноты связи различают количественные критерии оценки тесноты связи. Оценка линейного коэффициента корреляции может быть произведена по таблице 1.

Таблица 1. Количественные критерии оценки тесноты связи

Величина коэффициента корреляции	Характер связи
$ \pm 0,01 - \pm 0,15 $	Отсутствие связи
$ \pm 0,16 - \pm 0,20 $	Плохая связь
$ \pm 0,21 - \pm 0,30 $	Слабая связь
$ \pm 0,31 - \pm 0,40 $	Умеренная связь
$ \pm 0,41 - \pm 0,60 $	Средняя связь
$ \pm 0,61 - \pm 0,80 $	Высокая связь
$ \pm 0,81 - \pm 0,90 $	Очень высокая связь
$ \pm 0,91 - \pm 1,00 $	Полная связь

Либо укрупнено по приведенной ниже таблице:

Величина коэффициента корреляции	Характер связи
до $ \pm 0,3 $	Практически отсутствуют
$ \pm 0,3 - \pm 0,5 $	Слабая
$ \pm 0,5 - \pm 0,7 $	Умеренная
$ \pm 0,7 - \pm 1,0 $	Сильная

По направлению выделяют связь прямую и обратную. При прямой связи с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного. В случае обратной связи значения результативного признака изменяются под воздействием факторного, но в противоположном направлении по сравнению с изменением факторного признака.

По аналитическому выражению выделяют связи прямолинейные и криволинейные. Если статистическая связь между явлениями может быть приближенно выражена уравнением прямой линии, то ее называют линейной связью; если же она выражается уравнением какой-либо кривой линии (параболы, гиперболы, показательной и др.), то такую связь называют криволинейной.

Графически взаимосвязь двух признаков отображается с помощью поля корреляции. В системе координат по оси абсцисс

откладываются значения факторного признака, а на оси ординат – результативного. Каждое пересечение линий, проводимых через эти оси, обозначается точкой. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи.

Корреляционно-регрессионный анализ как общее понятие включает в себя измерение тесноты, направления связи и установление аналитического выражения (формы) связи.

Наиболее разработанной в теории статистики является методология парной корреляции, рассматривающая влияние вариации факторного признака x на результативный y и представляющая собой однофакторный корреляционный и регрессионный анализ.

Выбор типа функции может опираться на теоретические знания об изучаемом явлении, опыт предыдущих аналогичных исследований, или осуществляться эмпирически – перебором и оценкой функций разных типов.

При изучении связи экономических показателей используют различного вида уравнения прямолинейной и криволинейной связи:

$$y_x = a_0 + a_1 \cdot x \quad - \text{линейная функция} \quad (1)$$

$$y_x = a_0 + a_1 \cdot \lg x \quad - \text{полулогарифмическая функция} \quad (2)$$

$$y_x = a_0 \cdot a_1^x \quad - \text{показательная функция} \quad (3)$$

$$y_x = a_0 \cdot x^{a_1} \quad - \text{степенная функция} \quad (4)$$

$$y_x = a_0 + a_1 \cdot x + a_2 \cdot x^2 \quad - \text{параболическая функция} \quad (5)$$

$$y_x = a_0 + a_1 \cdot \frac{1}{x} \quad - \text{гипербола} \quad (6)$$

$$y_x = e^{a+b \cdot x} \quad - \text{экспоненциальная} \quad (7)$$

и другие.

Оценка параметров уравнений регрессии осуществляется методом наименьших квадратов (МНК), в основе которого лежит предположение о независимости наблюдений исследуемой совокупности. Сущность МНК заключается в нахождении параметров модели, при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии.

$$\sum (y_i - \hat{y})^2 = \sum (y_i - a_0 - a_1 \cdot x_i)^2 \rightarrow \min \quad .$$

для уравнения прямой. (8)

Рассмотрим расчет параметров однофакторной линейной модели. Внимание к линейным связям объясняется ограниченной вариацией переменных и тем, что в большинстве случаев нелинейные формы связи для выполнения расчетов преобразуют (путем логарифмирования или замены переменных) в линейную форму.

Уравнение однофакторной линейной корреляционной связи имеет вид (1): $y_x = a_0 + a_1 \cdot x$.

Система нормальных уравнений для нахождения параметров линейной регрессии методом наименьших квадратов имеет вид:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum x = \sum y \\ a_0 \cdot \sum x + a_1 \cdot \sum x^2 = \sum yx \end{cases} , \quad (9)$$

где n – объем исследуемой совокупности.

В уравнениях регрессии параметр a_0 показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов; параметр a_1 (в уравнении параболы и a_2) – коэффициент регрессии показывает, насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

Можно воспользоваться готовыми формулами нахождения параметров:

$$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \quad (10)$$

$$a_0 = \bar{y} - a_1 \cdot \bar{x} \quad (11)$$

Определив значения a_0 , a_1 и подставив их в уравнение связи $y_x = a_0 + a_1 \cdot x$, находим значения y_x , зависящие только от заданного значения x .

Для практического использования моделей регрессии большое значение имеет их адекватность, т.е. соответствие фактическим статистическим данным. Значимость коэффициентов простой линейной регрессии (при $n < 30$) осуществляется с помощью t – критерия Стьюдента. При этом вычисляют расчетные (фактические) значения t – критерия

для параметра a_0 :
$$t_{a_0} = |a_0| \cdot \frac{\sqrt{n-2}}{\sigma_{\text{ост}}}, \quad (12)$$

для параметра a_1 :
$$t_{a_1} = |a_1| \cdot \frac{\sqrt{n-2}}{\sigma_{\text{ост}}} \cdot \sigma_x, \quad (13)$$

где n – объем выборки;

$$\sigma_{\text{ост}} = \sqrt{\frac{\sum (y - y_x)^2}{n}} \quad (14)$$

- среднее квадратическое отклонение результативного признака y от выровненных значений y_x ;

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad (15)$$

или

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \quad (16)$$

- среднее квадратическое отклонение факторного признака x от общей средней \bar{x} .

Вычисленные значения t_{a_0} и t_{a_1} сравнивают с критическими (табличными) t , которые определяют по таблице Стьюдента с учетом принятого уровня значимости α и числом степеней свободы вариации $\nu = n - 2$. В социально-экономических исследованиях уровень значимости α обычно принимают равным 0,05. Параметр признается значимым при условии, если $t_{\text{расч}} > t_{\text{табл}}$.

Наряду с проверкой значимости отдельных параметров осуществляется проверка значимости уравнения регрессии в целом на основе расчета величины средней ошибки аппроксимации $\bar{\varepsilon}$.

Средняя ошибка аппроксимации рассчитывается по формуле:

$$\bar{\varepsilon} = \frac{1}{n} \cdot \sum \frac{|y_i - y_x|}{y_i} \cdot 100, \quad (17)$$

где y_i – эмпирические значения результативного признака;

y_x – теоретические значения результативного признака.

Значение средней ошибки аппроксимации не должно превышать 10-15%.

Проверка адекватности регрессионной модели может быть дополнена корреляционным анализом. Тесноту корреляционной связи между переменными x и y определяется с помощью теоретического корреляционного отношения

$$\eta = \sqrt{\frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \bar{y})^2}} \quad (18)$$

или

$$\eta = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}. \quad (19)$$

Подкоренное выражение – коэффициент детерминации – показывает долю вариации результативного признака под влиянием вариации признака-фактора.

Теоретическое корреляционное отношение может находиться в пределах от 0 до 1. Чем ближе корреляционное отношение к 1, тем теснее связь между признаками.

Кроме того, при линейной форме уравнения применяется другой показатель тесноты связи – линейный коэффициент корреляции. Для практических вычислений при малом числе наблюдений ($n < 30$) его удобнее исчислять по следующей формуле:

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}} . \quad (20)$$

Линейный коэффициент корреляции принимает значения от -1 до $+1$. Отрицательные значения указывают на обратную связь, положительные – на прямую. При $r=0$ связь отсутствует. Чем ближе коэффициент корреляции по абсолютной величине к 1, тем теснее связь между признаками. При $r=\pm 1$ связь функциональная.

Квадрат линейного коэффициента корреляции r^2 называется линейным коэффициентом детерминации, его числовое значение всегда заключено в пределах от 0 до 1.

Факт совпадений и несовпадений значений теоретического корреляционного отношения η и линейного коэффициента корреляции r используется для оценки формы связи. Если значения η и r совпадают, можно сделать вывод о наличии прямолинейной связи. Несовпадение этих величин свидетельствует, что связь между изучаемыми признаками не линейная, а криволинейная.

Показатели тесноты связи, исчисленные по данным сравнительно небольшой статистической совокупности, могут искажаться

действием случайных причин. Это вызывает необходимость проверки их существенности, дающей возможность распространять выводы по результатам выборки на генеральную совокупность.

Для оценки значимости коэффициента корреляции r используют t – критерий Стьюдента. При линейной однофакторной связи t – критерий можно рассчитать по формуле:

$$t_{\text{расч}} = r \cdot \sqrt{\frac{n-2}{1-r^2}}, \quad (21)$$

где $n-2$ – число степеней свободы при заданном уровне значимости α и объеме выборки n .

Полученное значение $t_{\text{расч}}$ сравнивают с табличным значением t – критерия (для $\alpha=0,05$ и $0,01$). Если $t_{\text{расч}} > t_{\text{табл}}$, то коэффициент корреляции признается значимым.

После проверки адекватности, установления точности и надежности построенной модели, ее необходимо проанализировать. Из уравнения регрессии следует, что при изменении факторного признака x на единицу результативный признак изменится на параметр a_1 .

Для удобства интерпретации параметра a_1 используют коэффициент эластичности. Он показывает средние изменения результативного признака при изменении факторного признака на 1%, %:

$$\mathcal{E} = a_1 \cdot \frac{\bar{x}}{\bar{y}}. \quad (22)$$

Имеет смысл вычислить остатки $\varepsilon_i = y - y_x$, характеризующие отклонение i -х наблюдений от значений, которые следует ожидать в среднем.

Многофакторный корреляционно-регрессионный анализ сводится к решению следующих задач:

- обосновать взаимосвязи факторов, влияющих на исследуемый показатель;

- определить степень влияния каждого фактора на результативный признак путем построения модели – уравнения множественной регрессии, которая позволяет устанавливать, в каком направлении и на какую величину изменится результативный показатель при изменении каждого фактора, входящего в модель;
- количественно оценить тесноту связи между результативным признаком и факторными.

Математически задача формулируется следующим образом. Требуется найти аналитическое выражение, наилучшим образом отражающее установленную теоретическим анализом связь независимых признаков с результативным, т.е. функцию

$$y_x = f(x_1, x_2, x_3, \dots, x_n) + \varepsilon_i \quad (23)$$

Выбрать форму связи довольно сложно. Все реально существующие зависимости между социально-экономическими явлениями можно описать, используя пять типов моделей:

1. линейная:

$$y_x = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (24)$$

2. степенная:

$$y_x = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \dots \cdot x_n^{a_n} \quad (25)$$

3. показательная:

$$y_x = e^{a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n} \quad (26)$$

4. параболическая:

$$y_x = a_0 + a_1 x_1^2 + a_2 x_2^2 + \dots + a_n x_n^2 \quad (27)$$

5. гиперболическая:

$$y_x = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_n}{x_n} \quad (28)$$

Следует отметить, что основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации.

В условиях использования ЭВМ выбор аппроксимирующей математической функции осуществляется перебором решений, наи-

более часто применяемых в анализе корреляции уравнений регрессии.

После выбора типа аппроксимирующей функции приступают к многофакторному корреляционному и регрессионному анализу, задачей которого является построение уравнения множественной регрессии и нахождение его неизвестных параметров a_0, a_1, \dots, a_n . Параметры уравнения множественной регрессии, как и в случае парной регрессии, находят по способу наименьших квадратов. Затем с помощью корреляционного анализа осуществляют проверку адекватности полученной модели. Адекватную модель экономически интерпретируют.

Построение и статистический анализ двухфакторной линейной модели (трехмерной регрессии).

Для расчета параметров простейшего уравнения множественной линейной двухфакторной регрессии

$$\hat{y}_{x_1 x_2} = a_0 + a_1 x_1 + a_2 x_2, \quad (29)$$

где $\hat{y}_{x_1 x_2}$ - расчетные значения зависимой переменной (результативного признака),

x_1, x_2 – независимые переменные (факторные признаки),

a_0, a_1, a_2 – параметры уравнения.

Построим следующую *систему нормальных уравнений*:

$$\begin{aligned} a_0 n + a_1 \sum x_1 + a_2 \sum x_2 &= \sum y; \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 &= \sum y x_1; \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 &= \sum y x_2. \end{aligned} \quad (30)$$

Параметры этой системы могут быть найдены, например, методом К. Гаусса.

После построения регрессионной модели необходимо исчислить различного рода характеристики тесноты связи между за-

висимой и независимой переменными: парные, частные и множественные коэффициенты корреляции, множественный коэффициент детерминации, а затем проверить адекватность данной модели.

Парные коэффициенты корреляции. Для измерения тесноты связи между двумя из рассматриваемых переменных (без учета их взаимодействия с другими переменными) применяются парные коэффициенты корреляции. Методика расчета таких коэффициентов и их интерпретация аналогичны методике расчета линейного коэффициента корреляции в случае однофакторной связи. Если известны средние квадратические отклонения анализируемых величин, то *парные коэффициенты корреляции* можно рассчитать проще по следующим формулам:

$$r_{yx_1} = \frac{\overline{x_1 y} - \bar{x}_1 \bar{y}}{\sigma_{x_1} \sigma_y}; \quad (31)$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \bar{x}_2 \bar{y}}{\sigma_{x_2} \sigma_y}; \quad (32)$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}}, \quad (33)$$

где $\sigma_y = \sqrt{y^2 - \bar{y}^2}; \quad (34)$

$$\sigma_{x_1} = \sqrt{x_1^2 - \bar{x}_1^2}; \quad (35)$$

$$\sigma_{x_2} = \sqrt{x_2^2 - \bar{x}_2^2}. \quad (36)$$

или используя формулу линейного коэффициента корреляции для малых выборок.

Частные коэффициенты корреляции.

Однако в реальных условиях все переменные, как правило, взаимосвязаны. Теснота этой связи определяется частными коэффициентами корреляции, которые характеризуют степень и влияние од-

ного из аргументов на функцию при условии, что остальные независимые переменные закреплены на постоянном уровне. В зависимости от количества переменных, влияние которых исключается, частные коэффициенты корреляции могут быть различного порядка: при исключении влияния одной переменной получаем частный коэффициент корреляции первого порядка; при исключении влияния двух переменных – второго порядка и т.д. Парный коэффициент корреляции между функцией и аргументом обычно не равен соответствующему частному коэффициенту.

Частный коэффициент корреляции первого порядка между признаками x_1 и y при исключении влияния признака x_2 вычисляют по формуле:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} , \quad (37)$$

то же – зависимость y от x_2 при исключении влияния x_1 :

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}} . \quad (38)$$

Можно рассчитать взаимосвязь факторных признаков при устранении влияния результативного признака:

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{yx_2}^2)}} , \quad (39)$$

где r - парные коэффициенты корреляции между соответствующими признаками.

Совокупный коэффициент множественной корреляции.

Показателем тесноты связи, устанавливаемой между результативными и двумя или более факторными признаками, является со-

вокупный коэффициент множественной корреляции $R_{yx_1, x_2, \dots, x_n}$. В случае линейной двухфакторной связи совокупный коэффициент множественной корреляции может быть рассчитан по формуле:

$$R_{yx_1, x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1, x_2}}{1 - r_{x_1, x_2}^2}}, \quad (40)$$

где r – линейные коэффициенты корреляции (парные); подстрочные индексы показывают, между какими признаками они исчисляются.

Совокупный коэффициент множественной корреляции измеряет одновременное влияние факторных признаков на результативный. Его значения находятся в пределах от -1 до $+1$. Чем меньше наблюдаемые значения изучаемого показателя отклоняются от линии множественной регрессии, тем корреляционная связь является более интенсивной, а следовательно, значение R ближе к единице.

Совокупный коэффициент множественной детерминации.

Величина R^2 называется *совокупным коэффициентом множественной детерминации*. Она показывает, какая доля вариации изучаемого показателя объясняется влиянием факторов, включенных в уравнение множественной регрессии. значение совокупного коэффициента множественной детерминации находится в пределах от 0 до 1. Поэтому, чем ближе R^2 к единице, тем вариация изучаемого показателя в большей мере характеризуется влиянием отобранных факторов.

Проверку значимости уравнения регрессии производят на основе вычисления F – критерия Фишера-Снедекора:

$$F_{расч} = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k} \quad (41)$$

или F – критерия Фишера:

$$F = \frac{\sigma_y^2}{\sigma_{\text{ост}}^2} \cdot \frac{n-m}{m-1}, \quad (42)$$

$$\text{где } \sigma_{\text{ост}}^2 = \frac{\sum (y - y_{x_1 x_2})^2}{n}; \quad (43)$$

k – число факторных признаков;

m – число параметров в уравнении регрессии.

Полученное значение – критерия $F_{\text{расч}}$ сравнивают с критическим (табличным) для принятого уровня значимости 0,05 или 0,01 и чисел степеней свободы $\nu_1 = m-1$ и $\nu_2 = n-m$. Если оно окажется больше соответствующего табличного значения, то данное уравнение регрессии статистически значимо, т.е. доля вариации, обусловленная регрессией, намного превышает случайную ошибку.

Для оценки значимости коэффициентов регрессии при линейной зависимости y от x_1 и x_2 – (двух факторов) используют t – критерий Стьюдента при $n-k-1$ степенях свободы:

$$t_{\text{расч}}^i = \frac{|a_i|}{\sigma_{a_i}}; \quad (44)$$

$$\sigma_{a_i}^2 = \frac{\sigma_y^2}{k}; \quad (45)$$

$$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}, \quad (46)$$

где k – число факторных признаков;

a_j – значение коэффициента регрессии при факторе x_j ;

σ_{a_i} – среднее квадратическое отклонение коэффициента регрессии;

σ_y^2 – дисперсия результативного признака.

Существенность совокупного коэффициента корреляции определяют по формуле:

$$t_{R_{y_{x_1}x_2}} = \frac{R_{y_{x_1}x_2} \sqrt{n-k-1}}{1 - R_{y_{x_1}x_2}^2}, \quad (47)$$

где k – число факторных признаков;

n – число единиц совокупности.

Значения оцениваемых $a_1, a_2, R_{y_{x_1}x_2}$ берутся по модулю.

При анализе адекватности уравнения регрессии исследуемому процессу возможны следующие варианты:

1. Построенная модель на основе ее проверки по F – критерию Фишера в целом адекватна, и все ее коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов.

2. Модель по F – критерию Фишера адекватна, но часть коэффициентов регрессии незначима. Следовательно, модель пригодна для принятия некоторых решений, но не для прогнозов.

3. Модель по F – критерию Фишера адекватна, но все коэффициенты незначимы. В этом случае модель признается полностью незначимой (неадекватной). На ее основе не принимаются решения и не осуществляются прогнозы.

На основе коэффициентов регрессии нельзя сказать какой из факторных признаков оказывает наибольшее влияние на результативный признак, так как коэффициенты регрессии между собой не сопоставимы, поскольку они измерены разными единицами. На их основе нельзя также установить, в развитии каких факторных признаков заложены наиболее крупные резервы изменения результативного показателя, потому что в коэффициентах регрессии не учтена вариация факторных признаков.

Чтобы иметь возможность судить о сравнительной силе влияния отдельных факторов и о тех резервах, которые в них заложены, должны быть вычислены *частные коэффициенты эластичности* ε_i , а также *бета-коэффициенты* β_i .

Различия в единицах измерения факторов устраняют с помощью *частных коэффициентов эластичности*, которые рассчитывают по формуле:

$$\varepsilon_i = a_i \frac{\bar{x}_i}{\bar{y}_i}, \quad (48)$$

где a_i – коэффициент регрессии при i -м факторе;

\bar{x}_i – среднее значение i -го фактора;

\bar{y}_i – среднее значение изучаемого показателя.

Частные коэффициенты эластичности показывают, на сколько процентов в среднем изменяется анализируемый показатель с изменением на 1% каждого фактора при фиксированном положении других факторов.

Для определения факторов, в развитии которых заложены наиболее крупные резервы улучшения изучаемого показателя, необходимо учесть различия в степени варьирования вошедших в уравнение факторов. Это можно сделать с помощью β – *коэффициентов*, которые вычисляют по формуле:

$$\beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y}, \quad (49)$$

где σ_{x_i} – среднее квадратическое отклонение i -го фактора;

σ_y – среднее квадратическое отклонение показателя; β – *коэффициент* показывает, на какую часть среднего квадратического отклонения изменяется результативный признак с изменением соответствующего факторного признака на величину его среднего

квадратического отклонения. Исходя из соотношения $\sum_{i=1}^n \beta_i r_i = R^2$ и, принимая во внимание, что коэффициент множественной детерминации R^2 есть доля изучаемых факторов в наличном приращении результативного показателя в анализируемой совокупности, можно

сделать вывод, что произведение $\beta_i r_i (1 \leq i \leq n)$ является показателем силы влияния соответствующего фактора на данный показатель.

Поделив произведение $\beta_i r_i$ на коэффициент множественной детерминации R^2 , получим коэффициент, который показывает, какова доля вклада анализируемого фактора в суммарное влияние всех отобранных факторов. Обозначив этот коэффициент Δ_i , получим

$$\Delta_i = \frac{\beta_i r_i}{R^2} . \quad (50)$$

2 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

До выполнения работы необходимо самостоятельно изучить правила выполнения корреляционного анализа.

При выполнении работы необходимо по исходным данным построить регрессионную модель

$$f(x) = \sum_{j=0}^m b_j \cdot x^j,$$

где значение m определяется заданием.

Оценить основные показатели качества модели: среднюю и максимальную погрешность приближения, а также показатель адекватности

$$R^2 = 1 - \frac{\sum_{i=1}^n \{f(x_i) - y_i\}^2}{\sum_{i=1}^n y_i^2},$$

который должен быть как можно ближе к единице.

Содержание отчета о лабораторной работе:

- общая постановка задачи;
- оценка параметров линейной модели.

Варианты заданий приведены в приложении А.

КОНТРОЛЬНЫЕ ВОПРОСЫ

- 1 Функциональная и корреляционная связи между величинами; приведите примеры таких связей.
- 2 Форма, направление и степень корреляционных связей.
- 3 Линейная корреляционная связь и количественный показатель этой связи.
- 4 Оценка достоверности выборочного коэффициента корреляции.
- 5 Понятие о функциях регрессии.

ЛИТЕРАТУРА

3. Математика в примерах и задачах: учеб. пособие/ Л. Н. Журбенко и др. - М.: ИНФРА-М, 2012.
4. Богомолов Н. В. Математика: учебник/ Н. В. Богомолов, Самойленко П. И. – М.: ЮРАЙТ, 2013

Лабораторная работа №3 ПОЛУЧЕНИЕ ЛИНЕЙНОЙ МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Цель работы – усвоение основных правил построения регрессионных моделей систем и оценки их качества с использованием методов математической статистики.

Содержание работы – определение параметров уравнения множественной линейной регрессии, представление полученных результатов в форме отчета и доказательство их правильности.

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1 Понятие о регрессии

Если все параметры объекта определяются точно, его модель можно считать детерминированной и представить в виде функциональной зависимости. На практике вследствие неточности измерения или влияния физической природы объекта эти параметры находятся с некоторой ошибкой ε , что делает величины на его входе и выходе ξ и η случайными (см. рисунок 1.1).

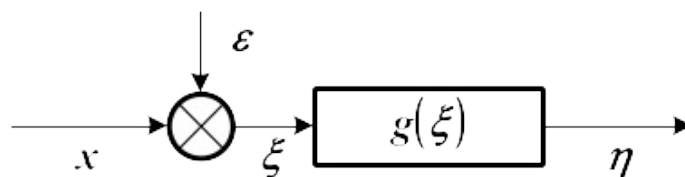


Рисунок 1.1 – Структурная схема модели объекта со случайной ошибкой входного сигнала

Информация о структуре объекта также может отсутствовать. В этих условиях зависимость $\eta = g(\xi)$ будет вероятностной (стохастической). Параметр ξ считается независимой переменной

(**фактором** или **предиктором**), а параметр η – зависимой переменной (или **откликом**).

Среди всех функций $g(\xi)$ необходимо найти ту, которая наилучшим образом воспроизводит значения η . Общепринятым способом решения этой задачи является *метод наименьших квадратов*.

Зависимость $g(\xi)$ называется наилучшим приближением η по методу наименьших квадратов, если математическое ожидание $M\{\eta - g(\xi)\}^2$ принимает наименьшее возможное значение. Зависимость $g(\xi)$ называется **среднеквадратической регрессией** величины η на величину ξ .

В случае нормального распределения ξ и η эта регрессия является линейной функцией

$$\frac{g(\xi) - M\eta}{\sigma_\eta} = \rho_{\xi\eta} \cdot \frac{\xi - M\xi}{\sigma_\xi}, \quad (1.1)$$

где $M\xi$, $M\eta$, σ_ξ , σ_η – математические ожидания и среднеквадратические отклонения ξ и η , $\rho_{\xi\eta}$ – коэффициент их корреляции.

Уравнение (1.1) определяет *прямую регрессии*

$$\hat{\eta} = g(\xi) = M\eta + \rho_{\xi\eta} \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot (\xi - M\xi) = \beta_0 + \beta_1 \cdot \xi, \quad (1.2)$$

где β_0 – смещение линии регрессии по оси ординат (т. н. *пересечение*), β_1 – коэффициент регрессии η на ξ . Их значения находятся по формулам

$$\beta_1 = \rho_{\xi\eta} \cdot \frac{\sigma_\eta}{\sigma_\xi}, \quad \beta_0 = M\eta - \beta_1 \cdot M\xi. \quad (1.3)$$

Разность $\eta - \hat{\eta}$ называется *остатком* η относительно ξ . Математическое ожидание квадрата остатка

$$M(\eta - \hat{\eta})^2 = \sigma_{\eta}^2 \cdot (1 - \rho_{\xi\eta}^2) \quad (1.4)$$

называется *остаточной дисперсией* η относительно ξ . Она показывает, насколько можно уменьшить дисперсию η , вычитая из переменной значение регрессии.

Для оценки точности регрессии можно рассмотреть следующее тождество (n – число наблюдений зависимой переменной).

$$(\eta_i - M\eta) = (\hat{\eta}_i - M\eta) + (\eta_i - \hat{\eta}_i), i = 1, \dots, n. \quad (1.5)$$

Геометрическая интерпретация (1.5) показана на рисунке 1.2.

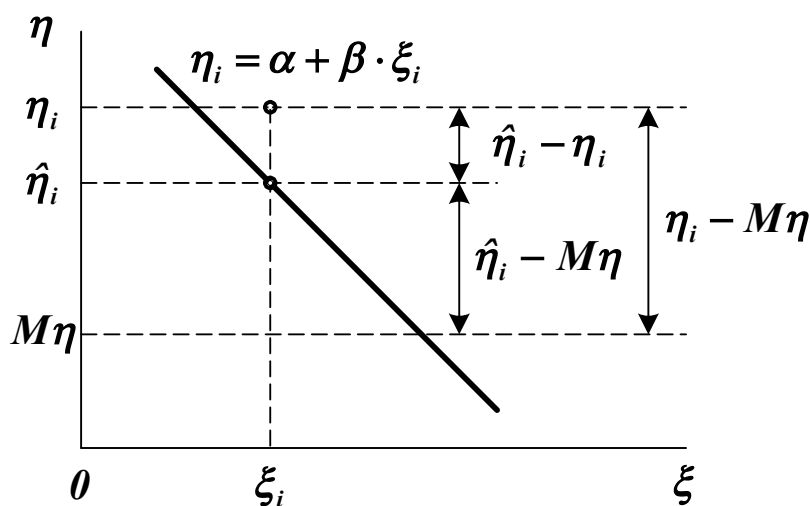


Рисунок 1.2 – Геометрическая интерпретация тождества (1.5)

Возводя обе части (1.5) в квадрат и используя формулу (1.2), можно получить основное уравнение дисперсионного анализа

$$\sum_{i=1}^n (\eta_i - M\eta)^2 = \sum_{i=1}^n (\hat{\eta}_i - M\eta)^2 + \sum_{i=1}^n (\eta_i - \hat{\eta}_i)^2, \quad (1.6)$$

$$SS_M = SS_R + SS_D,$$

где SS_M – общая сумма квадратов (*Square Sum*) отклонений от среднего (*Mean*), SS_R – сумма квадратов отклонений, обусловленная регрессией (*Regression*), SS_D – сумма квадратов остатка

(Deviation). Регрессия считается удовлетворительной, если близок к единице коэффициент детерминации

$$R^2 = \frac{SS_R}{SS_M} = 1 - \frac{SS_D}{SS_M}. \quad (1.7)$$

Можно составить таблицу дисперсионного анализа, включающую значения элементов уравнения (1.6).

Примечание. Любая сумма квадратов связана с числом ее степеней свободы, которое показывает, как много независимых элементов выборки из n независимых чисел требуется для образования данной суммы квадратов.

Таблица 1.1 – Таблица дисперсионного анализа

<i>Источник вариации</i>	<i>Число степеней свободы</i>	<i>SS</i>
Регрессия	1	$\sum (\hat{\eta}_i - M\eta)^2$
Остаток	$n - 2$	$\sum (\eta_i - \hat{\eta}_i)^2$
Общий	$n - 1$	$\sum (\eta_i - M\eta)^2$

Регрессия считается статистически значимой, если отношение

$$F = \frac{\sum (\hat{\eta}_i - M\eta)^2}{\sum (\eta_i - \hat{\eta}_i)^2} \cdot (n - 2) \quad (1.8)$$

превышает критерий Фишера $F_\alpha(1, n - 2)$ для заданного уровня доверительной вероятности α .

Оценки коэффициентов β_0 и β_1 считаются статистически значимыми, если значения выражений

$$t_0 = \frac{|\beta_0|}{s} \cdot \sqrt{\frac{n \cdot \sum (\xi_i - M\xi)^2}{\sum \xi_i^2}}, t_1 = \frac{|\beta_1|}{s} \cdot \sqrt{\sum (\xi_i - M\xi)^2} \quad (1.9)$$

превышают критерий Стьюдента $t_{1-\alpha/2}(n-2)$ для заданного

уровня доверительной вероятности α .

1.2 Множественная линейная регрессия

Пусть ξ_1, \dots, ξ_m – случайные величины, имеющие математические ожидания $M\xi_i$ и дисперсии $D\xi_i$ ($i = 1, \dots, m$). Если ξ_i – отклик, а все остальные – факторы, можно определить зависимость

$$\hat{\eta} = g(\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_m), \quad (1.10)$$

где $\hat{\eta} = \xi_i$. Форма $g(\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_m)$ должна подбираться так, чтобы выполнялось условие

$$M\{\eta - g(\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_m)\}^2 \rightarrow \min. \quad (1.11)$$

Если зависимость (1.10) представляется линейной функцией

$$\hat{\eta} = M\xi_i + \sum_{j=1, j \neq i}^m \beta_{ij} \cdot (\xi_j - M\xi_j), \quad (1.12)$$

то ее называют **множественной линейной регрессией** величины ξ_i на остальные величины. Коэффициенты регрессии β_{ij} находятся по формуле

$$\beta_{ij} = -\frac{D_{ij}^{-1}}{D_{ii}^{-1}}, \quad (1.13)$$

где D – матрица центральных корреляционных моментов. Каждый ее элемент равен

$$D_{ij} = M\{(\xi_i - M\xi_i) \cdot (\xi_j - M\xi_j)\} = \begin{cases} D\xi_i, & i = j, \\ \text{cov}\{\xi_i, \xi_j\}, & i \neq j. \end{cases} \quad (1.14)$$

Если матрица D положительно определена, то коэффициенты регрессии всегда определяются однозначно.

Таблица дисперсионного анализа для множественной линейной регрессии приведена ниже.

Таблица 1.2 – Таблица дисперсионного анализа для множественной линейной регрессии

<i>Источник вариации</i>	<i>Число степеней свободы</i>	<i>SS</i>
Регрессия	$m - 1$	$\sum (\hat{\eta}_i - M\eta)^2$
Остаток	$n - m$	$\sum (\eta_i - \hat{\eta}_i)^2$
Общий	$n - 1$	$\sum (\eta_i - M\eta)^2$

Величина R^2 в данном случае является мерой статистической взаимосвязи между ξ_i и всеми остальными величинами и называется *сводным коэффициентом корреляции*.

Регрессия считается статистически значимой, если отношение

$$F = \frac{\sum (\hat{\eta}_i - M\eta)^2}{\sum (\eta_i - \hat{\eta}_i)^2} \cdot \frac{n - m}{m - 1} \quad (1.15)$$

превышает критерий Фишера $F_\alpha(m - 1, n - m)$ для заданного уровня доверительной вероятности α .

Оценки коэффициентов β_{ij} считаются статистически значимыми, если значения выражений

$$t_j = \sqrt{\frac{\rho_{ij}^2}{1 - \rho_{ij}^2} \cdot (n - m)} \quad (1.16)$$

превышают критерий Стьюдента $t_{1-\alpha/2}(n - m)$ для заданного уровня доверительной вероятности α .

Параметры ρ_{ij} , входящие в (1.16), называемые *частными коэффициентами корреляции*, определяют меру статистической взаимосвязи ξ_i и ξ_j после устранения влияния всех остальных факторов. Они вычисляются по формуле

$$\rho_{ij} = - \frac{D_{ij}^{-1}}{\sqrt{D_{ii}^{-1} \cdot D_{jj}^{-1}}}. \quad (1.17)$$

1.3 Пошаговая регрессия

Для одного и того же объекта можно построить несколько регрессионных моделей, отличающихся набором независимых переменных. Для выбора оптимальной модели обычно используют следующие критерии:

- минимальная сумма квадратов остатка;
- минимальное число коэффициентов, совместимое с разумной ошибкой;
- достаточные физические основания для включения фактора в модель.

Одним из методов получения оптимальной модели является пошаговая регрессия. Она заключается в последовательном включении в уравнение независимых переменных и исследовании его на значимость.

На первом шаге выбирается переменная ξ_j , наиболее сильно коррелированная с величиной η , составляется уравнение $\hat{\eta} = g(\xi_j)$, проверяется общая значимость регрессии по критерию (1.15) и значимость коэффициента β_{ij} по критерию (1.16). Если результат проверки отрицательный, наилучшей считают модель $\hat{\eta} = M\eta$.

На втором шаге вычисляется остаток $\eta - M\eta - \beta_{ij} \cdot (\xi_j - M\xi_j)$, и определяется наиболее сильно коррелированная с ним переменная ξ_k . Составляется уравнение $\hat{\eta} = g(\xi_j, \xi_k)$, проверяется общая значимость регрессии и значимость коэффициентов β_{ij} и β_{ik} . Незначимые переменные из уравнения регрессии исключаются.

На третьем шаге вычисляется остаток регрессии $\eta - M\eta - \beta_{ij} \cdot (\xi_j - M\xi_j) - \beta_{ik} \cdot (\xi_k - M\xi_k)$, определяется наиболее сильно коррелированная с ним переменная ξ_l , и процедура форми-

рования уравнения регрессии повторяется.

Процесс прекращается, если дальнейшее расширение или сокращение модели делают регрессию незначимой.

2 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

До выполнения работы необходимо самостоятельно изучить основы регрессионного анализа и правила построения множественных регрессионных моделей.

При выполнении работы необходимо по исходным данным получить полностью значимую при уровне доверительной вероятности $\alpha = 0,95$ полиномиальную модель вида

$$y = \sum_{j=0}^m b_j \cdot x^j. \quad (1.18)$$

Для приведения зависимости (1.18) к виду (1.12) необходимо сформировать матрицу переменных ξ_i следующего вида

$$\xi = \begin{bmatrix} y_1 & x_1 & x_1^2 & \dots & x_1^m \\ \dots & \dots & \dots & \dots & \dots \\ y_n & x_n & x_n^2 & \dots & x_n^m \end{bmatrix}, \quad (1.19)$$

где ξ_1 – отклик, $\xi_j = x^{j-1}$ ($j = 2, \dots, m+1$) – факторы. С помощью пошаговой процедуры в регрессию

$$\hat{\eta} = g(\xi_2, \dots, \xi_m) \quad (1.20)$$

должны быть включены только значимые факторы.

Содержание отчета о лабораторной работе:

- общая постановка задачи;
- расчет основных статистических характеристик независимых переменных;
- решение задачи пошаговой регрессии.

Результаты расчетов должны приводиться как в числовой, так и в графической форме (например, в виде графика остатков). Для каждой модели должны определяться среднеквадратическая ошибка и коэффициент детерминации.

Варианты заданий приведены в приложении **Б**.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что считается среднеквадратической регрессией одной переменной относительно другой переменной?
2. Как строится уравнение линейной регрессии? Как определяются коэффициенты ее уравнения?
3. Как строится уравнение множественной линейной регрессии? Как определяются коэффициенты ее уравнения?
4. Как строится таблица дисперсионного анализа для уравнения регрессии?
5. Как определяется значимость регрессии в целом? Как определяется значимость переменных, входящих в уравнение регрессии?
6. Как выполняется процедура пошаговой регрессии?

ЛИТЕРАТУРА

1. Вентцель, Е.С., Овчаров, Л.А. Теория случайных процессов и ее инженерные приложения [Текст] : Учеб. пособие для втузов. / Вентцель Е.С., Овчаров Л.А. – 2-е изд., стер. – М. : Высш. шк., 2000. – 383 с. : ил. – (Высшая математика для втузов). – ISBN 5-06-003831-9.
2. Случайные процессы [Текст] : Учеб. для вузов. / Волков И.К., Зуев С.М., Цветкова Г.М. ; Под ред. В.С. Зарубина, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 448 с. – (Математика в техническом университете. Вып. XVIII). – ISBN 5-7038-1267-4 (Вып. XVIII) ; ISBN 5-7038-1270-4.
3. Вентцель, Е.С., Овчаров, Л.А. Теория вероятностей и ее инженерные приложения [Текст] : Учеб. пособие для втузов. / Вентцель Е.С., Овчаров Л.А. – 2-е изд., стер. – М. : Высш. шк., 2000. – 480 с. : ил. – (Высшая математика для втузов). – ISBN 5-06-003830-0.
4. Теория вероятностей [Текст] : Учеб. для вузов. / А.В. Печин-

кин, О.И. Тескин, Г.М. Цветкова и др. ; Под ред. В.С. Зарубина, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 456 с. : ил. – (Математика в техническом университете. Вып. XVI). – ISBN 5-7038-1307-7 (Вып. XVI) ; ISBN 5-7038-1270-4.

5. Дрейпер, Н., Смит, Г. Прикладной регрессионный анализ [Текст] : монография : в 2 т. / Дрейпер Н., Смит Г. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 1986-1987. – (Математико-статистические методы за рубежом).

Т. 1 : – М. : Финансы и статистика, 1986. – 366 с. : ил.

Т. 2 : – М. : Финансы и статистика, 1987. – 351 с. : ил.

6. Химмельблау, Д. Анализ процессов статистическими методами [Текст] / Д. Химмельблау. – М. : Мир, 1973. – [957 с.] : ил.

Лабораторная работа №4

ПОЛУЧЕНИЕ НЕЛИНЕЙНОЙ МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Цель работы – усвоение основных правил построения детерминированных статических моделей систем и оценки их качества.

Содержание работы – определение параметров линейной и нелинейной статических моделей, представление полученных результатов в форме отчета и доказательство их правильности.

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1 Понятие об аппроксимации

Если все параметры объекта в его установившемся режиме определяются точно, его статическую модель можно считать *детерминированной* и представить в виде функциональной зависимости. Для ее получения необходима выборка значений входных и выходных параметров объекта $\{x_i, y_i, i=1, \dots, n\}$ (n – число наблюдений), полученная в ходе его исследования (см. рисунок 1.1).

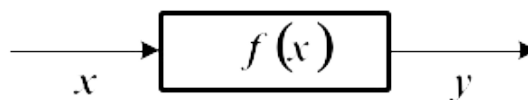


Рисунок 1.1 – Структурная схема детерминированной модели объекта

Точная форма зависимости y от x в большинстве случаев неизвестна. Поэтому ее **аппроксимируют** (приближенно заменяют) некоторой функцией $y=f(x)$. Традиционным способом приближения является **метод наименьших квадратов** (МНК), стремящийся минимизировать сумму квадратов отклонений расчетных выходных

значений модели от наблюдаемых

$$S = \sum_{i=1}^n [f(x_i) - y_i]^2. \quad (1.1)$$

Выражение (1.1) определяет **функцию невязки**.

Процедура аппроксимации включает в себя два этапа.

1. Выбирается форма модели $f(x, \beta_1, \dots, \beta_m)$, исходя из физической природы задачи или из интуиции исследователя (β_1, \dots, β_m – коэффициенты модели). Если характер зависимости неизвестен, предпочтение отдается более простым формулам.

2. Определяются параметры (коэффициенты) выбранной модели из условия минимума функции невязки

$$S = \sum_{i=1}^n [f(x_i, \beta_1, \dots, \beta_m) - y_i]^2 \rightarrow \min. \quad (1.2)$$

Минимум (1.2) находится путем приравнивания нулю частных производных по β_1, \dots, β_m .

$$\begin{cases} \frac{\partial S}{\partial \beta_1} = 2 \cdot \sum_{i=1}^n \left\{ [f(x_i, \beta_1, \dots, \beta_m) - y_i] \cdot \frac{\partial f(x_i, \beta_1, \dots, \beta_m)}{\partial \beta_1} \right\} = 0, \\ \dots \\ \frac{\partial S}{\partial \beta_m} = 2 \cdot \sum_{i=1}^n \left\{ [f(x_i, \beta_1, \dots, \beta_m) - y_i] \cdot \frac{\partial f(x_i, \beta_1, \dots, \beta_m)}{\partial \beta_m} \right\} = 0. \end{cases} \quad (1.3)$$

Полученные уравнения составляют систему, решая которую можно определить неизвестные коэффициенты. Если модель линейна относительно своих параметров, то эта система также будет линейной и решаться аналитически. Если форма модели нелинейная, то решение такой системы может быть получено путем итерационной процедуры.

1.2 Построение линейных статических моделей

Линейная статическая модель представляется в виде

$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_m \cdot x_m + \varepsilon = \beta_0 + \sum_{j=1}^m \beta_j \cdot x_j + \varepsilon = \hat{y} + \varepsilon, \quad (1.4)$$

где x_1, \dots, x_m – независимые переменные, входящие в уравнение; y – наблюдаемая зависимая переменная; \hat{y} – оценка значения y посредством модели; ε – ненаблюдаемая ошибка, определяемая разностью между измеренным значением y и величиной, полученной по уравнению модели.

Для получения оценок параметров β_0, \dots, β_m по схеме МНК необходимо определить минимум следующей функции

$$S = \sum_{i=1}^n \left\{ w_i \cdot \left[\beta_0 + \sum_{j=1}^m \beta_j \cdot x_{ij} - y_i \right]^2 \right\} \rightarrow \min, \quad (1.5)$$

где w_i имеет смысл весового коэффициента i -го наблюдения величины y . Он может выбираться по одному из следующих принципов.

- Если наблюдения каждого значения y_i неоднократные, то $w_i = n_i$, где n_i – число повторных измерений зависимой переменной y_i при фиксированном значении x_i .
- Если наблюдения каждого значения y_i неравноточные, то $w_i = 1 / \sigma_i^2$, где σ_i – среднеквадратичная погрешность i -го измерения зависимой переменной y_i .
- Если наблюдения каждого значения y_i однократные и равноточные, то $w_i = 1$.

Учитывая выражение для \hat{y} , можно составить схему МНК

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial \beta_0} = 2 \cdot \sum_{i=1}^n \left\{ w_i \cdot \left[\beta_0 + \sum_{j=1}^m \beta_j \cdot x_{ij} - y_i \right] \right\} = 0, \\ \frac{\partial S}{\partial \beta_k} = 2 \cdot \sum_{i=1}^n \left\{ w_i \cdot \left[\beta_0 + \sum_{j=1}^m \beta_j \cdot x_{ij} - y_i \right] \cdot x_{ik} \right\} = 0, k = 1, \dots, m. \end{array} \right. \quad (1.6)$$

Отсюда можно получить эквивалентную систему линейных уравнений, решением которой будут параметры уравнения (1.4)

$$\left\{ \begin{array}{l} \beta_0 \cdot \sum_{i=1}^n w_i + \sum_{j=1}^m \left\{ \beta_j \cdot \sum_{i=1}^n w_i \cdot x_{ij} \right\} = \sum_{i=1}^n w_i \cdot y_i \\ \beta_0 \cdot \sum_{i=1}^n w_i \cdot x_{ik} + \sum_{j=1}^m \left\{ \beta_j \cdot \sum_{i=1}^n w_i \cdot x_{ij} \cdot x_{ik} \right\} = \sum_{i=1}^n w_i \cdot y_i \cdot x_{ik}, \\ k=1, \dots, m \end{array} \right. \quad (1.7)$$

Используя обозначения

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}, W = \begin{bmatrix} w_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & w_n \end{bmatrix},$$

$$B = \begin{bmatrix} \beta_0 \\ \dots \\ \beta_m \end{bmatrix}, E = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{bmatrix},$$

можно записать схему МНК в матричной форме

$$\begin{aligned} Y &= X \cdot B + E, \\ S &= (X \cdot B - Y)^T \cdot W \cdot (X \cdot B - Y), \\ \frac{\partial S}{\partial B} &= 2 \cdot X^T \cdot W \cdot (X \cdot B - Y) = 0. \end{aligned} \quad (1.8)$$

Система (1.7) представится в виде

$$X^T \cdot W \cdot X \cdot B = X^T \cdot W \cdot Y. \quad (1.9)$$

Если $\det(X^T \cdot W \cdot X) \neq 0$, то она имеет единственное решение

$$B = (X^T \cdot W \cdot X)^{-1} \cdot X^T \cdot W \cdot Y. \quad (1.10)$$

Эта схема определения коэффициентов применима для линейной комбинации произвольных функций

$$y = \beta_0 \cdot f_0(x) + \dots + \beta_m \cdot f_m(x) = \sum_{j=0}^m \beta_j \cdot f_j(x), \quad (1.11)$$

и для полиномиальных уравнений

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \dots + \beta_m \cdot x^m = \sum_{j=0}^m \beta_j \cdot x^j. \quad (1.12)$$

Матрицы X' для комбинации функций и X'' для полного полинома (содержащего все степени x от 0 до m) будут иметь вид

$$X' = \begin{bmatrix} f_0(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_0(x_n) & \dots & f_m(x_n) \end{bmatrix}, \quad X'' = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix}.$$

Основным критерием качества модели является оценка средне-квадратичной (стандартной) погрешности расчета параметра y

$$S_y^2 = \frac{1}{n - m'} \cdot \sum_{i=1}^n [\hat{y}_i - y_i]^2, \quad (1.13)$$

где $m' = m + 1$ – число коэффициентов уравнения модели.

С учетом введенных матричных обозначений

$$S_y^2 = \frac{(X \cdot B - Y)^T \cdot (X \cdot B - Y)}{n - m'}. \quad (1.14)$$

Если эта величина превышает допустимый предел, модель отвергается, или ее форма модифицируется.

Важным показателем качества модели является чувствительность оценок коэффициентов β_0, \dots, β_m к относительным изменениям величин y_i , вызванными влиянием погрешности их определения. Она находится по формуле

$$\gamma = \frac{\partial \beta_k / \beta_k}{\partial y_i / y_i} = \frac{\partial \ln \beta_k}{\partial \ln y_i} = \frac{\partial \beta_k}{\partial y_i} \cdot \frac{y_i}{\beta_k}. \quad (1.15)$$

С учетом (1.10) выражение для расчета чувствительности коэффициента β_k к изменению величины y_i примет вид

$$\gamma_k = \left[(X^T \cdot W \cdot X)^{-1} \cdot X^T \cdot W \right]_{ki} \cdot \frac{y_i}{\beta_k}. \quad (1.16)$$

1.3 Повышение качества линейной модели

Параметры модели всегда определяются с погрешностью, величина которой тем больше, чем ближе $\det(X^T \cdot W \cdot X)$ к нулю. Среднеквадратичная ошибка оценки коэффициента β_k равна

$$S_{\beta k}^2 = S_y^2 \cdot (X^T \cdot W \cdot X)_{kk}^{-1}. \quad (1.17)$$

Коэффициент β_k считается *незначимыми и подлежащим исключению из уравнения модели*, если значения выражения

$$t = |\beta_k| / S_{\beta k} \quad (1.18)$$

меньше критерия Стьюдента $t_{1-\alpha/2}(n-m')$ для уровня доверительной вероятности α и числа степеней свободы $n-m'$.

В ряде случаев качество линейной модели можно повысить, представив ее в виде

$$y = \beta_0 + \sum_{j=1}^m \beta_j \cdot (x_j - \bar{x}_j) + \varepsilon, \quad (1.19)$$

где \bar{x}_j – средние значения соответствующих переменных.

Тогда система (1.7) представится в форме

$$\begin{cases} \beta_0 \cdot \sum_{i=1}^n w_i + \sum_{j=1}^m \left[\beta_j \cdot \sum_{i=1}^n w_i \cdot (x_{ij} - \bar{x}_j) \right] = \sum_{i=1}^n w_i \cdot y_i, \\ \beta_0 \cdot \sum_{i=1}^n w_i \cdot (x_{ik} - \bar{x}_k) + \sum_{j=1}^m \left[\beta_j \cdot \sum_{i=1}^n w_i \cdot (x_{ij} - \bar{x}_j) \cdot (x_{ik} - \bar{x}_k) \right] = \\ \quad \sum_{i=1}^n w_i \cdot y_i \cdot (x_{ik} - \bar{x}_k), k = 1, \dots, m. \end{cases} \quad (1.20)$$

Так как $\sum_{i=1}^n w_i \cdot (x_{ij} - \bar{x}_j) = 0, j = 1, \dots, m$, эта система упрощается

$$\left\{ \begin{array}{l} \beta_0 \cdot \sum_{i=1}^n w_i = \sum_{i=1}^n w_i \cdot y_i, \\ \sum_{j=1}^m \left[\beta_j \cdot \sum_{i=1}^n w_i \cdot (x_{ij} - \bar{x}_j) \cdot (x_{ik} - \bar{x}_k) \right] = \sum_{i=1}^n w_i \cdot y_i \cdot (x_{ik} - \bar{x}_k), \\ k=1, \dots, m. \end{array} \right. \quad (1.21)$$

Отсюда можно получить выражение для определения β_0 . При равных весах наблюдений $\beta_0 = \bar{y}$.

При записи системы (1.21) в матричной форме изменяются компоненты X и B

$$X = \begin{bmatrix} x_{11} - \bar{x}_1 & \dots & x_{1m} - \bar{x}_m \\ \dots & \dots & \dots \\ x_{n1} - \bar{x}_1 & \dots & x_{nm} - \bar{x}_m \end{bmatrix}, B = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_m \end{bmatrix}.$$

Остальные матричные обозначения сохраняются.

Погрешности определения всех коэффициентов, кроме β_0 , находятся по общим формулам. Погрешность β_0 составляет

$$S_{\beta_0} = \frac{S_y}{\sqrt{\sum_{i=1}^n w_i}}. \quad (1.22)$$

Для равноточных измерений $S_{\beta_0} = S_y / \sqrt{n}$.

1.4 Построение нелинейных статических моделей

Если модель нелинейная относительно своих параметров, то в общем случае система (1.3) должна решаться численно. Популярным средством решения нелинейных задач является метод Ньютона. Он заключается в следующем.

Введем обозначения

$$\begin{cases} \phi_1(\beta_1, \dots, \beta_m) = \sum_{i=1}^n \left\{ [f(x_i, \beta_1, \dots, \beta_m) - y_i] \cdot \frac{\partial f(x_i, \beta_1, \dots, \beta_m)}{\partial \beta_1} \right\}, \\ \dots \\ \phi_m(\beta_1, \dots, \beta_m) = \sum_{i=1}^n \left\{ [f(x_i, \beta_1, \dots, \beta_m) - y_i] \cdot \frac{\partial f(x_i, \beta_1, \dots, \beta_m)}{\partial \beta_m} \right\}. \end{cases} \quad (1.23)$$

Зададимся начальными приближениями значений коэффициентов $\beta_1^{(0)}, \dots, \beta_m^{(0)}$ и разложим функции ϕ_1, \dots, ϕ_m в ряд Тейлора в окрестности начальной точки. Из разложения исключим все составляющие со степенями производных выше первой. Тогда

$$\begin{cases} \phi_1 \approx \phi_1^{(0)} + \left[\frac{\partial \phi_1}{\partial \beta_1} \right]_0 \cdot (\beta_1 - \beta_1^{(0)}) + \dots + \left[\frac{\partial \phi_1}{\partial \beta_m} \right]_0 \cdot (\beta_m - \beta_m^{(0)}), \\ \dots \\ \phi_m \approx \phi_m^{(0)} + \left[\frac{\partial \phi_m}{\partial \beta_1} \right]_0 \cdot (\beta_1 - \beta_1^{(0)}) + \dots + \left[\frac{\partial \phi_m}{\partial \beta_m} \right]_0 \cdot (\beta_m - \beta_m^{(0)}). \end{cases} \quad (1.24)$$

Так как в точке минимума функции невязки \mathcal{S} должно выполняться условие $\phi_1 = \dots = \phi_m = 0$, эту систему можно приближенно представить следующим образом

$$\begin{cases} \left[\frac{\partial \phi_1}{\partial \beta_1} \right]_0 \cdot \Delta\beta_1 + \dots + \left[\frac{\partial \phi_1}{\partial \beta_m} \right]_0 \cdot \Delta\beta_m = -\phi_1^{(0)}, \\ \dots \\ \left[\frac{\partial \phi_m}{\partial \beta_1} \right]_0 \cdot \Delta\beta_1 + \dots + \left[\frac{\partial \phi_m}{\partial \beta_m} \right]_0 \cdot \Delta\beta_m = -\phi_m^{(0)}. \end{cases} \quad (1.25)$$

где величины $\Delta\beta_j = \beta_j - \beta_j^{(0)}, j = 1, \dots, m$ – поправки к соответствующим коэффициентам.

Система (1.25) линейна относительно поправок. Решая ее, можно найти значения $\Delta\beta_j$ и определить скорректированные значе-

ния коэффициентов $\beta_j = \beta_j^{(0)} + \Delta\beta_j$.

Если все поправки по абсолютной величине меньше заданной точности ε , итерационный процесс прекращается. В противном случае выполняется переназначение переменных $\beta_j^{(0)} = \beta_j$, пересоставляется система (1.25) и определяются новые поправки.

Найдем решение нелинейной задачи МНК в матричной форме. Введем обозначения векторов и матриц

$$X = \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}, B = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_m \end{bmatrix}, F(X, B) = \begin{bmatrix} f(x_1, \beta_1, \dots, \beta_m) \\ \dots \\ f(x_n, \beta_1, \dots, \beta_m) \end{bmatrix},$$

$$\Phi = \begin{bmatrix} f(x_1, \beta_1, \dots, \beta_m) - y_1 \\ \dots \\ f(x_n, \beta_1, \dots, \beta_m) - y_n \end{bmatrix} = F(X, B) - Y, W = \begin{bmatrix} w_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & w_n \end{bmatrix}.$$

Тогда функция невязки и ее производные в матричной форме будут иметь следующий вид

$$S = [F(X, B) - Y]^T \cdot W \cdot [F(X, B) - Y] = \Phi^T \cdot W \cdot \Phi \rightarrow \min,$$

$$\frac{\partial S}{\partial B} = 2 \cdot J^T \cdot W \cdot \Phi. \quad (1.26)$$

Коэффициенты матрицы Якоби J находятся по правилу

$$J = \begin{bmatrix} \frac{\partial \Phi_1}{\partial \beta_1} & \dots & \frac{\partial \Phi_1}{\partial \beta_m} \\ \dots & \dots & \dots \\ \frac{\partial \Phi_m}{\partial \beta_1} & \dots & \frac{\partial \Phi_m}{\partial \beta_m} \end{bmatrix}.$$

Если задан вектор начальных приближений коэффициентов B_0 , то можно получить приближенное выражение для оценки производной $\partial S / \partial B$ в точке, где функция невязки имеет минимум

$$\frac{\partial S}{\partial B} \approx 2 \cdot J_0^T \cdot W \cdot \Phi. \quad (1.27)$$

Разложим функцию Φ в ряд Тейлора в окрестности точки B_0 . Исключим из этого разложения компоненты со степенями производных выше первой

$$\Phi = \Phi_0 + J_0 \cdot (B - B_0). \quad (1.28)$$

Тогда можно записать

$$\begin{aligned} J_0^T \cdot W \cdot [\Phi_0 + J_0 \cdot (B - B_0)] &= 0, \\ J_0^T \cdot W \cdot J_0 \cdot (B - B_0) &= -J_0^T \cdot W \cdot \Phi_0, \\ B &= B_0 - (J_0^T \cdot W \cdot J_0)^{-1} \cdot J_0^T \cdot W \cdot \Phi_0. \end{aligned} \quad (1.29)$$

Последнее выражение является рекуррентной формулой для уточнения значений вектора коэффициентов. Ее применяют до тех пор, пока евклидова длина (норма) вектора поправок $\|B - B_0\|$ не станет меньше заданной точности ε .

В ряде случаев решение нелинейной задачи МНК можно упростить, преобразовав уравнение модели к линейному виду путем соответствующей замены переменных. Найденные параметры преобразованных линейных характеристик используются для расчета коэффициентов исходных зависимостей. Некоторые наиболее типичные примеры преобразования переменных и пересчета коэффициентов показаны ниже.

$$\begin{aligned} y = \beta_1 \cdot \exp(\beta_2 \cdot x) &\rightarrow \begin{bmatrix} x' = x \\ y' = \ln(y) \end{bmatrix} \rightarrow \\ &\rightarrow y' = \alpha_1 + \alpha_2 \cdot x' \rightarrow \begin{bmatrix} \beta_1 = \exp(\alpha_1) \\ \beta_2 = \alpha_2 \end{bmatrix}, \end{aligned}$$

$$y = \beta_1 \cdot x^{\beta_2} \rightarrow \begin{bmatrix} x' = \ln(x) \\ y' = \ln(y) \end{bmatrix} \rightarrow y' = \alpha_1 + \alpha_2 \cdot x' \rightarrow \begin{bmatrix} \beta_1 = \exp(\alpha_1) \\ \beta_2 = \alpha_2 \end{bmatrix},$$

$$y = \beta_1 + \frac{\beta_2}{x} \rightarrow \begin{bmatrix} y' = y \\ x' = \frac{1}{x} \end{bmatrix} \rightarrow y' = \alpha_1 + \alpha_2 \cdot x' \rightarrow \begin{bmatrix} \beta_1 = \alpha_1 \\ \beta_2 = \alpha_2 \end{bmatrix},$$

$$y = \frac{\beta_1}{x + \beta_2} \rightarrow \begin{bmatrix} y' = \frac{1}{y} \\ x' = x \end{bmatrix} \rightarrow y' = \alpha_1 + \alpha_2 \cdot x' \rightarrow \begin{bmatrix} \beta_1 = \frac{1}{\alpha_2} \\ \beta_2 = \frac{\alpha_1}{\alpha_2} \end{bmatrix},$$

$$y = \frac{\beta_1 \cdot x + \beta_2}{x + \beta_3} \rightarrow \begin{bmatrix} x_1 = x \\ x_2 = -x \cdot y \\ y' = y \end{bmatrix} \rightarrow$$

$$\rightarrow y' = \alpha_0 + \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 \rightarrow \begin{bmatrix} \beta_1 = \frac{\alpha_1}{\alpha_2} \\ \beta_2 = \frac{\alpha_0}{\alpha_2} \\ \beta_3 = \frac{1}{\alpha_2} \end{bmatrix},$$

2 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

До выполнения работы необходимо самостоятельно изучить правила построения детерминированных статических моделей.

При выполнении работы необходимо решить две задачи.

1. По исходным данным построить полиномиальную модель

$$f(x) = \sum_{j=0}^m b_j \cdot x^j, \quad (1.30)$$

где значение m определяется заданием. Последовательно применяя процедуру исключения коэффициентов, в доверительный интервал которых

$$b_j - S_{bj} \cdot t_{1-\alpha/2}(n-m-1) \leq b_j \leq b_j + S_{bj} \cdot t_{1-\alpha/2}(n-m-1) \quad (1.31)$$

попадает нулевое значение, добиться полной значимости модели при уровне доверительной вероятности $\alpha = 0,95$. В первую очередь исключаются коэффициенты с наибольшей относительной

погрешностью. Оценить основные показатели качества модели: среднюю и максимальную погрешность приближения, а также показатель адекватности

$$R^2 = 1 - \frac{\sum_{i=1}^n \{f(x_i) - y_i\}^2}{\sum_{i=1}^n y_i^2}, \quad (1.32)$$

который должен быть как можно ближе к единице.

2. По тем же исходным данным построить нелинейную модель, форма которой определяется заданием. Рассчитать для нее те же показатели качества, что и для линейной модели.

Содержание отчета о лабораторной работе:

- общая постановка задачи;
- оценка параметров линейной модели;
- оценка параметров нелинейной модели;
- сравнение качества линейной и нелинейной моделей.

Варианты заданий приведены в приложении **Б**.

КОНТРОЛЬНЫЕ ВОПРОСЫ

7. Что понимается под аппроксимацией функций? Каков порядок процедуры аппроксимации?

8. Как строится линейная статическая модель с помощью процедуры МНК? Как определяются ее параметры?

9. Как оценивается качество линейной статической модели? Как можно повысить ее качество?

10. Как строится нелинейная статическая модель с помощью процедуры МНК? Как определяются ее параметры?

ЛИТЕРАТУРА

1. Бахвалов, Н. С. Численные методы [Текст] / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков. – 2-е изд. – М. : Физматлит, 2002. –

630 с. – (Технический университет. Математика). – ISBN 5-93208-043-4 (в пер.).

2. Лоусон, Ч., Хансон, Р. Численное решение задач метода наименьших квадратов [Текст] / Лоусон Ч., Хансон Р. – М. : Наука. Гл. ред. физ.-мат. лит., 1986. – 232 с.

3. Бард, Й. Нелинейное оценивание параметров [Текст] / Бард Й. – М. : Статистика, 1979. – 349 с. ; ил. – (Математико-статистические методы за рубежом).

4. Методы аппроксимации функций [Текст] : Методические указания к выполнению лабораторной работы по курсу «Основы численных методов» для специальности 071900 – Информационные системы и технологии и других инженерных специальностей. / Д.В. Болдырев. – Невинномысск : Изд-во НТИ СевКавГТУ, 2004. – 32 с. : ил.

ПРИЛОЖЕНИЕ А

Варианты заданий для выполнения лабораторных работ №1 и 2

Центральный фед.окр.				Южный фед. округ				Гюменская обл.	6,8	6,7	8,7
Белгородская обл.	5,6	6,0	5,9	Респ.Адыгея	13,7	12,9	17,6	Ханты-Манс. АО	6,1	7,8	9,7
Брянская обл.	6,8	6,7	8,8	Респ.Дагестан	22,3	22,3	27,2	Ямало-НенецкийАО	5,4	7,1	6,5
Владимирская обл.	10,9	9,0	9,1	Респ.Ингушетия	58,5	64,9	46,3	Челябинская обл.	5,1	5,4	5,2
Воронежская обл.	5,5	7,5	8,6	Кабардино-балк.респ.	20,7	23,4	25,7	Сибирский фед. окр			
Ивановская обл.	4,2	6,8	4,6	Респ.Калмыкия	16,7	18,0	21,7	Респ. Алтай	11,6	10,0	9,6
Калужская обл	5,6	5,7	6,3	Карачаево-Черк.респ.	19,4	13,6	16,9	Респ. Бурятия	13,4	14,2	15,3
Костромская обл.	5,0	4,8	5,9	Респ.Сев.Осетия	8,5	8,8	11,7	Респ. Тыва	20,5	21,8	19,7
Курская обл.	7,3	7,1	7,5	Краснодарский край	8,9	7,5	8,8	Респ.Хакасия	9,1	8,9	10,3
Липецкая обл.	4,9	8,0	4,2	Ставроп. край	7,9	7,0	9,6	Алтайский край	8,8	9,0	9,0
Московская обл.	3,0	3,3	3,7	Астраханская обл.	7,9	12,0	11,0	Красноярский край	9,9	9,0	9,2
Орловскя обл.	6,0	6,1	5,9	Волгоградская обл.	8,6	6,8	9,4	Гаймырский АО	10,0	14,0	9,1
Рязанская обл.ю	5,2	5,3	5,8	Ростовская обл.	8,0	8,6	8,6	Эвенкийский АО	3,9	4,4	3,2
Смоленская обл.	8,0	7,8	9,1					Иркутск ая обл	8,9	10,0	10,5
Тамбовская обл.	8,7	8,5	9,7	Приволжский фед. окр				Усть-Ордынский АО	12,6	12,8	9,8
Тверская обл.	4,5	5,9	5,3	Респ.Башкортостан	6,5	7,1	7,1	Кемеровская обл.	7,3	8,6	9,8
Тульская обл.	2,7	5,0	4,6	Респ.Марий Эл	10,2	9,9	9,1	Новосибирская обл.	7,4	7,8	9,0
Ярославская обл.	3,0	4,0	4,7	Респ.Мордовия	4,7	7,0	6,0	Омская обл.	9,3	8,6	9,1
Г.Москва	1,6	0,8	1,6	Респ.Татарстан	5,6	6,7	7,4	Гомская обл..	9,0	10,5	10,1
Северо-зап. фед. округ				Удмуртская респ	8,4	7,8	8,0	Читинская обл	8,8	11,1	12,4
Респ. Карелия	3,6	8,8	7,3	Чувашская респ.	8,6	11,4	9,9	Дальневост. фед.округ			
Респ. Коми	12,4	11,5	12,4	Пермский край	6,9	7,0	7,0	Респ.Саха	9,5	8,9	8,8
Архангельская обл.	5,9	5,5	7,2	Кировская обл.	7,9	7,1	8,7	Приморский край	8,0	8,0	9,6
Вологодская обл.	5,3	5,2	6,3	Нижегородская обл.	5,3	6,0	7,4	Хабаровский край	6,0	5,7	6,6
Калининградская обл.	4,5	6,6	6,5	Оренбургская обл.	6,5	9,4	10,8	Амурская обл.	8,2	10,3	11,2
Ленинградская обл.	5,9	7,4	6,8	Пензенская обл.	6,5	6,5	6,9	Камчатская обл.	9,1	9,5	11,0
Мурманская обл.	6,7	8,8	11,0	Самарская обл.	4,3	5,3	5,3	Корякский АО	6,4	7,8	11,1
Новгородская обл.	5,5	5,8	6,2	Саратовская обл.	8,2	9,1	9,9	Магаданская обл.	5,4	7,0	7,9
Псковская обл.	7,4	6,6	5,6	Ульяновская обл.	6,9	7,7	9,5	Сахалинская обл.	4,6	7,6	7,5
Г.Санкт-Петербург	2,4	2,2	2,7	Уральский фед. Округ				Еврейская АО	9,8	7,9	8,3
				Курганская обл.	12,4	11,3	12,5	Чукотский АО	3,7	4,4	3,5
				Свердловская обл.	7,0	6,7	7,3				

ПРИЛОЖЕНИЕ Б

Варианты заданий для выполнения лабораторных работ №3 и 4

№ варианта																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Начальное значение независимой переменной x																			
1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	1,9	1,8	1,7	1,6	1,5	1,4	1,3	1,2	1,1
Шаг изменения независимой переменной x																			
0,1	0,2	0,3	0,2	0,3	0,1	0,3	0,2	0,1	0,2	0,1	0,3	0,2	0,1	0,2	0,3	0,2	0,3	0,1	0,2
Значения зависимой переменной y																			
9,0	6,0	3,5	4,0	297	7,0	40,0	2,0	12,8	8,8	2,8	29,0	2,3	92,0	2,0	32,0	2,0	5,0	26,8	0,1
5,4	7,0	3,0	12,0	24,5	11,5	16,5	3,2	6,0	4,0	3,6	11,0	2,6	52,0	2,5	18,5	2,5	4,4	23,6	1,4
4,2	9,0	2,8	20,0	10,5	15,0	11,5	3,8	4,8	2,3	5,0	8,0	2,9	36,0	2,5	12,5	2,5	3,4	20,2	3,4
3,6	11,0	2,6	32,0	7,0	18,5	9,0	4,4	4,2	1,4	7,0	7,0	3,3	26,0	3,0	9,0	2,5	2,6	16,2	6,2
3,2	14,0	2,5	44,0	5,5	21,5	7,5	4,6	4,0	0,8	10,6	6,0	3,7	18,0	4,0	6,0	3,0	1,6	11,4	10,7
3,0	17,0	2,4	58,0	4,5	24,5	7,0	4,8	3,8	0,5	14,8	6,0	4,0	12,0	6,0	4,5	3,0	1,2	7,4	15,1
3,0	20,0	2,4	74,0	4,0	27,5	6,0	4,8	3,6	0,3	20,2	5,2	4,3	8,0	12,5	3,0	3,5	0,6	3,6	19,6
2,8	25,0	2,3	90,0	3,5	30,0	6,0	5,0	3,6	0,2	24,4	5,1	4,5	6,0	31,5	2,5	4,0	0,4	1,6	22,5
2,6	30,0	2,3	108	3,5	32,5	5,5	5,0	3,4	0,1	30,6	4,9	4,7	4,0	48,9	1,5	5,0	0,2	0,2	25,9
2,6	37,0	2,3	126	3,5	35,0	5,0	5,0	3,4	0,1	30,6	4,7	4,7	4,0	68,3	1,5	5,0	0,2	0,2	25,9
Степень аппроксимирующего полинома																			
5	6	6	5	5	6	6	5	5	6	6	5	5	6	6	5	5	6	6	5

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Невинномысский технологический институт

Кафедра информационных систем, электропривода и автоматики

КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

Методические указания к практическим занятиям
для направления подготовки 15.04.04 –
Автоматизация технологических процессов и производств

Невинномысск 2016

Настоящие методические указания предназначены для студентов направления 15.04.04 – Автоматизация технологических процессов и производств. Они разработаны в соответствии с федеральным государственным образовательным стандартом и основной образовательной программой направления.

В методических указаниях определены требования к содержанию практических занятий, приведены необходимые теоретические сведения, даны варианты заданий и приведен список рекомендуемых литературных источников.

Составитель *канд. техн. наук, доцент Д.В. Болдырев*

Отв. редактор *канд. техн. наук, доцент А.А. Евдокимов*

ВВЕДЕНИЕ

Методические указания подготовлены в соответствии с программой курса «Корреляционный и регрессионный анализ», в которых рассматриваются все основные методы и приемы расчета и анализа статистических показателей, содержатся задачи, решение которых необходимо для выработки практических навыков студентов при изучении курса. Задачи построены на условных данных. К каждому разделу показана последовательность выполнения расчетов, приведены примеры решения и набор задач для самостоятельного решения.

Цель практических занятий — овладение знаниями студентами методологических основ статистики, умение применять эти знания в анализе социально-экономических явлений, проводить статистические расчеты, привить студентам навыки проведения самостоятельной исследовательской работы с помощью статистических методов.

При проведении занятий студенты должны проявить умение самостоятельно применять статистическую методологию в анализе конкретных задач, работать с учебной и научной литературой, рассчитывать статистические показатели, делать на их основе аргументированные выводы.

1 СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ, СВОДКА И ГРУППИРОВКА

Статистическое наблюдение – первая стадия статистического исследования. Статистическое исследование начинается с учета состояния факторов и сбора первичного материала.

Учет факторов и сбор первичного материала осуществляется разными способами. Факты учитываются путем опроса человека при переписи населения, либо путем организации систематического учета на предприятиях и организациях. Сбор первичного материала систематического материала осуществляется путем текущей отчетности или специально организованным наблюдением.

Статистическое наблюдение проводится по разработанному плану, включающему программные и организационные вопросы.

В программе наблюдения содержатся вопросы об объекте наблюдения, единице совокупности, единице наблюдения, системе показателей, которые определяют содержание программы статистического наблюдения.

Объектом статистического наблюдения называется совокупность единиц изучаемого явления, о которых могут быть собраны сведения. Например, при переписи населения объектом наблюдения является совокупность жителей населенного пункта, а при учете хозяйственной деятельности предприятий — совокупность предприятий.

ЕДИНИЦА СОВОКУПНОСТИ — это элемент объекта статистического наблюдения, являющийся носителем признаков, подлежащих регистрации.

ЕДИНИЦА НАБЛЮДЕНИЯ — это первичное звено, от которого должны быть получены статистические сведения, т.е. завод, учебное заведение, рабочий, учащийся и т.д.

ЕДИНИЦА СОВОКУПНОСТИ - это то, что подвергается обследованию, а единица наблюдения - источник получаемых сведений. Единицей совокупности могут быть материальные ресурсы, отдельные виды оборудования, численность отдельных категорий работников.

В программе наблюдения содержатся перечень сформулированных вопросов, не которые долек быть получены ответы и показатели, составляющие содержание прогрера.

После того, как решены программные вопросы статистического наблюдения, решается вопрос об организации наблюдения. Далее необходимо изучить классификацию статистического наблюдения по полноте охвата единиц совокупности и непрерывности учета факторов во времени и способах их учета.

Пример решения задачи.

Исследовать успеваемость студентов второго курса экономического факультета ПГУ и факторы, влияющие на результаты зимней экзаменационной сессии. Установить

объект наблюдения, единицу совокупности, составить программу наблюдения.

Объектам наблюдения являются студенты второго курса экономического факультета ПГУ.

Единицей совокупности является отдельный студент. Единицей отчетности будет выступать деканат экономического факультета.

Программа статистического наблюдения содержит следующие вопросы:

- 1) Какие оценки подмени студентами по каждому предмету?
- 2) Для выявления факторов, влияющих на успеваемость, следует получить ответы на такие вопросы:
 - а) средний балл аттестата;
 - б) длительность перерыва в учебе после окончания учебного заведения;
 - в) посещаемость занятий;
 - г) своевременность выполнения письменных работ;
 - д) семейное положение;
 - е) бюджет времени.

Сводка представляет собой комплекс статистических операций по одобрению конкретных единичных данных, образующих совокупность в целях выявления типичных черт и закономерностей, присущих изучаемому явлению. Статистическая сводка состоит в том, что первичные материалы, полученные в результате наблюдения, заносятся в таблицы, подсчитываются итоговые показатели. В дальнейшем на основе свободных итогов вычисляют средние и относительные величины. Программ сводки могут быть разными.

Применяется простая сводка только для подсчета итоговых данных и группировка. При помощи группировок статистические материалы систематизируются, делятся на группы по существенным признакам. Группировка позволяет получить результаты, по которым можно изучить состав совокупности, характерные черты и свойства явлений, выявить закономерности и взаимосвязи. В процессе статистического исследования метод группировок применяется для решения многих задач. С помощью метода группировок решаются три

основные задачи: выделение социально-экономических типов, изучение структуры однотипных совокупностей, выявление существенных связей зависимости между признаками исследуемого явления. В соответствии с задачами группировок и подразделяются на три вида: типологические, структурные и аналитические. Иногда трудно установить границы между видами группировок. Например, типологическая группировка может служить основой изучения структуры совокупности, а на основе структурной группировки устанавливаются взаимосвязи между признаками.

Примерами типологических группировок могут служить группировки хозяйств по формам собственности: группировка населения по источникам средств существования и по общественным группам.

Структурные группировки применяются для анализа структуры совокупности. С помощью структурных группировок изучают население по полу, возрасту и другим признакам, состав предприятий по уровню рентабельности.

В основу типологических и структурных группировок могут быть положены атрибутивные количественные признаки.

Атрибутивный признак — это качественный признак группировки, который не имеет количественного выражения. Примером может служить группировка населения на городское и сельское. При группировке по полу выделяются две группы: мужчины и женщины.

При использовании количественных признаков в группировках возникают проблемы определения числа групп и величины интервала. При равных интервалах для определения числа групп используется формула Стерджесса:

$$n = 1 + 3,322 \lg N$$

где n — число групп; N — численность единиц совокупности.

Величина интервала определяется делением размаха вариации на число групп:

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

где X_{\max} и X_{\min} - максимальное и минимальное значения признака; n - число групп.

При статистическом исследовании иногда приходится производить вторичную группировку, т.е. перегруппировать статистический материал, уже сведенный в группы. К этому методу прибегают в тех случаях, когда в результате первоначальной группировки нечетко проявился характер распределения изучаемой совокупности, а также для приведения к сопоставимому виду группировок с различными интервалами с целью их сравнения.

Существует два способа вторичной группировки.

Первый - это укрупнение или, наоборот, уменьшение первоначальных интервалов.

Второй — это доленая перегруппировка, т.е. образование новых групп на основе закрепления за каждой группой определенной доли единиц совокупности.

Третьим видом группировок является метод аналитических группировок. Аналитические группировки — это исследование взаимосвязей между изучаемыми признаками. Взаимосвязанные признаки делятся на факторные и результативные. Взаимосвязь проявляется в том, что с изменением значения Факторного признака систематично возрастает или убывает значение результативного признака.

Разработка аналитической группировки производится с определения числа групп и величины интервала.

Далее составляется макет рабочей таблицы.

Групповые данные рабочей таблицы заносятся в свободную групповую таблицу, характеризующую зависимость между размером основных производственных фондов и объемом товарной продукции.

Обе таблицы должны быть оформлены статистически грамотно: иметь заглавие, подлежащее, сказуемое таблицы, единицы измерения показателей, итоговые общие и средние показатели. По результатам расчетов следует сделать выводы.

Порядок решения задачи.

С целью изучения зависимости между среднегодовой стоимостью основных фондов, производственных фондов и

выпуском продукции необходимо произвести группировку заводов по среднегодовой стоимости основных фондов, образовав четыре группы заводов с равными интервалами

По каждой группе и совокупности заводов определить:

1) число заводов;

2) среднегодовую стоимость основных фондов - всего и в среднем на один завод

3) стоимость выпущенной продукции - всего и в среднем на один завод;

4) стоимость выпущенной продукции на один рубль основных фондов (фондоотдачу).

Результаты представить в виде таблицы. Сделать краткие выводы.

Имеются следующие отчетные данные двадцати заводов одной из отраслей промышленности:

Номер завода	Среднегодовая стоимость основных производственных фондов, млн.руб.	Стоимость продукции, млн.руб.
1	6,9	10
2	8,9	12
3	3,0	3,5
4	5,7	4,5
5	3,7	3,4
6	5,6	8,8
7	4,5	3,5
8	7,1	9,6
9	2,0	2,6
10	10,0	13,9
11	6,5	6,8
12	7,5	9,9
13	8,3	10,8
14	5,6	8,9
15	4,5	7,0
16	6,1	8,0

17	3,0	2,5
18	6,9	9,2
19	6,5	6,9
20	4,1	4,3

Для образования групп заводов по стоимости основных производственных фондов определим величину интервала группировочного признака по формуле

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

где X_{\max} и X_{\min} — наименьшее и наибольшее значения признака совокупности; n - число групп.

Тогда $i = \frac{10 - 2}{5} = \frac{8}{5} = 1,6$ млн. руб.

Образуем группы:

- 1 2,0-3,6
- 2 3,6-5,2
- 3 5,2-6,8
- 4 6,8- 8,4
- 5 8,4 - 10,0

Составим рабочую таблицу:

Но мер группы	Группы заводов по стоимости основных производственных фондов, млн. руб.	Но мера за- вода	Среднегод овая стои- мость основ- ных производ- ственных фон- дов, млн. руб.	Стои- мость продукци и, млн. руб.
1	2-3,6	3	3,0	3,5
		9	2,0	2,6
		17	3,0	2,5
	итого	3	8,0	8,6
2	3,6-5,2	5	3,7	3,4
		7	4,5	3,5
		15	4,5	7
		20	4,1	4,3
	итого	4	16,8	18,2

3	5,2-6,8	4	5,7	4,5
		6	5,6	8,8
		11	6,5	6,8
		14	5,6	8,9
		16	6,1	8,0
		19	6,5	6,9
	Итого	6	36	43,9
4	6,8-8,4	1	6,9	10,0
		8	7,1	9,6
		12	7,5	9,9
		13	8,3	10,0
		18	6,9	9,2
	Итого	5	36,7	49,5
5	8,4-10,0	2	8,9	12
		10	10,0	13,9
	Итого	2	18,9	25,9

По данным рабочей таблицы составим аналитическую группировку.

Зависимость выпуска продукции от размера среднегодовой стоимости основных производственных фондов:

Группы заводов по стоимости основных производственных фондов, мин. руб.	Число заводов	Стоимость основных производственных фондов млн. руб.		Выпуск продукции, млн. руб.		Выпуск продукции на один рубль основных производственных фондов, руб.
		Всего	Среднее на один завод	Всего	Среднее на один завод	
2,0-3,6	3	8	2,6	8	2,	1,08
3,6-5,2	4	,0	7	,6	87	1,08
5,2 - 6,3	6	1	4,2	1	4,	1,22

6,8- 8,4	5	6,8	0	8,2	55	1,35
8,4 - 10,0	2	3	6,0	4	7,	1,37
		6,0	7,3	3,9	32	
		3	4	4	9,	
		6,7	9,4	9,5	9	
		1	5	2	1	
		8,9		5,9	2,95	
Итого	2	1	5,8	1	7,	1,26
	0	16,4	2	46,	31	
				1		

Данные таблицы показывают, что с возрастанием стоимости основных производственных фондов на один завод увеличивается выпуск продукции.

Следовательно, между данными признаками существует прыгая зависимость. С увеличением стоимости основных производственных фондов на один завод, возрастает и фондоотдача (выпуск продукции па один рубль основных производственных фондов).

Задачи для самостоятельного решения.

Задача 1. За отчетный период имеются следующие данные о реализации товаров и издержек обращения по предприятиям торговли района, млн. руб.

№ п/п	Розничный товарооборот	Издержки обращения
1	2	3
1	511	30,0
2	560	34,0
3	800	46,0
4	463	30,9
5	230	15,9
6	392	25,2
7	640	42,0
8	404	27,0
9	200	16,4
10	425	34,8
11	570	37,0

12	472	28,0
13	250	18,7
14	665	39,0
15	650	36,0
16	620	36,0
17	383	25,0
18	550	38,5
19	750	44,0
20	660	37,0
21	452	27,0
22	565	35,0

1. Для выявления зависимости между объемом розничного товарооборота и издержками обращения сгруппируйте предприятия по объему розничного товарооборота, образовав 5 групп с равными интервалами. Каждую группу предприятия и совокупность их в целом охарактеризуйте показатели:

- числом предприятий;
- объемом товарооборота – всего и в среднем на одно предприятие;
- относительным уровнем издержек обращения (удельным весом издержек обращения в объеме розничного товарооборота).

Результаты группировки представьте в таблице. Дайте анализ показателей таблицы и сделайте выводы.

Задача 2. Имеются следующие данные 5%-й механической выборки отрасли промышленности, проведенной в регионе

№ п/п	Объем выполненных работ, млн. руб.	Накладные расходы, Млн. руб.
	1	2
1	6,0	2,0
2	7,0	2,2
3	5,1	1,4
4	6,5	2,5
5	7,4	2,1

6	11,0	3,6
7	11,5	3,8
8	5,3	1,4
9	7,6	2,3
10	9,2	3,0
11	13,0	4,4
12	11,8	4,2
13	9,0	2,8
14	8,0	2,6
15	9,5	3,5
16	6,4	2,0
17	5,5	1,8
18	8,4	4,0
19	9,6	3,1
20	12,0	3,8
21	9,4	3,0
22	5,0	1,5
23	5,8	1,8
24	7,2	2,0
25	12,7	4,6
26	14,7	4,6
27	15,0	5,0
28	14,1	4,8
29	10	3,4
30	10,5	3,6
31	8,6	3,2
32	7,5	2,4
33	10,6	3,2
34	7,8	2,2
35	10,4	3,2
36	8,5	3,0

Постройте аналитическую группировку предприятий для изучения зависимости между объемом выполненных работ и накладными расходами, образовав пять групп предприятий с разными интервалами, охарактеризовав каждую группу и совокупность в целом:

а) числом предприятий;

б) объемом накладных расходов — всего и в среднем на одно предприятие.

Результаты группировки представьте в таблице, сделайте выводы.

Задача 3. Для изучения капитальных вложений в производство из собственных средств предприятий в регионе проведена 5%-я механическая. выборка, в результате которой получены следующие данные (табл.)

№ п/п	Нераспределенная прибыль, млн. руб.	Инвестиции в основные средства, млн. руб.
	1	2
1	2,2	0,06
2	2,0	0,04
3	4,3	0,44
4	5,0	0,60
5	6,0	0,90
6	2,3	0,12
7	3,6	0,20
8	4,2	0,36
9	5,8	0,80
10	4,7	0,60
11	2,5	0,18
12	3,8	0,40
13	4,5	0,53
14	4,8	0,65
15	4,4	0,42
16	5,4	0,70
17	5,2	0,50
18	4,1	0,35
19	3,3	0,20
20	5,6	0,70
21	3,9	0,40
22	4,8	0,73
23	4,5	0,62
24	4,7	0,70

25	3,4	0,30
----	-----	------

1. Произведите аналитическую группировку предприятий для изучения зависимости между инвестируемой в денежные средства и нераспределенной прибылью, образовав четыре группы предприятий с равными интервалами. По каждой группе и совокупности предприятий исчислите:

- 1) число предприятий;
- 2) величину нераспределенной прибыли — всего и в среднем на одно предприятие;
- 3) величину инвестиций - всего и в среднем на одно предприятие.

Результаты представьте в таблице. Дайте анализ показателей и сделайте выводы.

Задача 4. имеются следующие выбранные данные о денежных вкладах в банке, в руб.:

500, 620, 1450, 1950, 2640, 3330, 4000, 756, 1420, 950, 1500, 1625, 1580, 2570, 2430, 2250, 2120, 2310, 2980, 715, 812, 890, 3260, 1320, 1250, 1890, 1000, 910, 1190, 1050.

Постройте интервальный ряд распределения денежных вкладов. Вычислите средний размер вклада, коэффициент вариации, относительные величины структуры. С вероятностью 0,954 определите границы, в которых будет находиться средний размер вклада в банке. Постройте график распределения. Сделайте выводы.

2 АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

Абсолютные статистические величины, характеризуя численность единиц изучаемой совокупности или объем присущих им признаков, всегда являются числами именованными. В зависимости от качественной особенности изучаемого явления и задач исследования эти величины выражаются в различных единицах измерения: *натуральных, трудовых и денежных*. При учете продукции и товаров в натуральном выражении часто применяются *условные единицы измерения*. Сущность применения условных единиц измерения состоит в том, что отдельные разновидности изучаемой совокупности выражаются в единицах одного признака, услов-

но принятого за единицу измерения. Поэтому основной вопрос применения условных единиц измерения состоит в выборе признака, по которому устанавливаются соответствующие коэффициенты пересчета.

Пример. За отчетный период предприятие произвело следующие виды мыла и моющих средств:

Виды мыла и моющих средств	Количество, кг
Мыло хозяйственное 60%-й жирности	500
Мыло хозяйственное 40%-й жирности	1250
Мыло туалетное 80%-й жирности	1500
Стиральный порошок 10%-й жирности	2500

Требуется определить общее количество выработанной предприятием продукции в условно-натуральных единицах измерения. За условную единицу измерения принимается мыло 40%-й жирности.

Решение: Для определения общего количества продукции, выработанной предприятием, необходимо исчислить коэффициенты перевода. Если условной единицей измерения является мыло 40%-й жирности, то это значение жирности принимается равным единице. Тогда коэффициенты перевода в условное мыло (40%-й жирности) исчисляем так: мыло хозяйственное 60%-й жирности: $60/40=1,5$; мыло туалетное 80%-й жирности: $80/40=2,0$; стиральный порошок 10%-й жирности: $10/40=0,25$.

Далее определим количество продукции в условно-натуральных единицах измерения.

Общий объем производства мыла и моющих средств по видам

Виды мыла и моющих средств	Количество, кг	Коэффициент перевода	Количество продукции в
			производства

			условно-натуральном исчислении, кг
Мыло хозяйственное 60%-й жирности	500	1,5	750
Мыло хозяйственное 40%-й жирности	1250	1,0	1250
Мыло туалетное 80%-й жирности	1500	2,0	3000
Стиральный порошок 10%-й жирности	2500	0,25	625
Итого	-	-	5625

Общий объем производства мыла и моющих средств в 40%-м исчислении составил 5625 кг.

Задачи для самостоятельного решения

Задача 1. В отчетном периоде на производственные нужды израсходованы следующие виды топлива: мазут топочный – 800 т, уголь донецкий – 460 т, газ природный – 940 тыс.м³.

На основе приведенных данных определите общий размер потребленного в отчетном периоде топлива в условных единицах измерения.

Примечание. Средние калорийные эквиваленты для перевода отдельных видов топлива в условное топливо:

Вид топлива
Калорийные эквиваленты

Уголь донецкий, т
0,9

Мазут топочный, т
1,37

Газ природный, тыс.м³
1,2

Задача 2. Имеются следующие данные о выпуске отдельных видов продукции, тыс. т:

Наименование	тыс. т	
Наименование	тыс. т	
Мыло хозяйственное:		
Мыло туалетное	40,0	
60% - ное		40,0 По-
рошок стиральный	28,0	
40% - ное		20,0

Определите общий объем этого производства путем выражения отдельных видов продукции в условных единицах.

Примечание. Для перерасчета отдельных видов продукции в условные единицы используются следующие соотношения:

Вид продукции	Коэффициент	Наименование
перерасчета	перерасчета	перерасчета
Мыло хозяйственное:		Мыло
туалетное	1,75	
60% - ное		1,75 По-
рошок стиральный	0,5	
40% - ное		1,0

Задача 3. Выплавка чугуна по цехам завода в отчетном периоде характеризуется следующими данными:

Вид чугуна	Тыс. т	Вид
чугуна	тыс. т	
Ванадиевый	52	Ли-
тейный	110	
Зеркальный	76	Пе-
редельный	126	
Ковкий и валкий	94	
Хромоникелевый	124	

Определите общий объем выплавки чугуна в натуральном выражении и в условном натуральном выражении (в переводе на передельный чугун). Объясните разницу в исчисленных показателях объема выпуска продукции.

Примечание. Для перевода отдельных видов чугуна в условные натуральные единицы измерения используются следующие коэффициенты:

Вид чугуна	Коэффициенты
Передельный	1,0
Литейный	1,15
Ванадиевый	1,35
Ковкий и валкий	1,15
Хромоникелевый	1,5
Зеркальный	1,5

Относительные величины характеризуют количественное соотношение сравниваемых абсолютных величин. Если база сравнения равна 1, то относительные величины выражаются в коэффициентах, если база сравнения равна 100, то относительная величина выражена в процентах (%), если база сравнения равна 1000 – в промилле (‰).

В зависимости от задач, содержания и познавательного значения выражаемых количественных соотношений различают следующие виды относительных показателей: 1) планового задания и выполнения плана; 2) динамики; 3) структуры; 4) интенсивности; 5) координации; 6) сравнения.

1. Относительные показатели планового задания – отношение уровня планового задания к уровню показателя, достигнутого в предыдущем периоде.

Пример. В II квартале выручка от реализации ОАО составила 120 млн. руб., в III квартале планируется объем выручки от реализации в 150 млн. руб.

Определить относительную величину планового задания.

Решение. $\frac{150}{120} * 100\% = 125\%$

Таким образом, в II квартале планируется увеличение выручки от реализации на 25 %.

Относительные показатели выполнения плана – отношение фактически достигнутого уровня в текущем периоде к уровню планируемого показателя на этот же период.

Пример. Выручка от реализации в IV квартале составил 200,0 млн. руб., при плане 150 млн. руб.

Определить степень выполнения плана товарооборота магазином в IV квартале.

Решение. $\frac{200,0}{150,0} * 100\% = 133,3\%$

План по товарообороту магазином выполнен на 133,3%, т.е. перевыполнение плана составило 33,3%.

Относительные показатели динамики характеризуют изменение уровня развития какого-либо явления во времени. Показатели этого вида получаются делением уровня признака за определенный период или момент времени на уровень этого же показателя в предыдущий период или момент. Относительные величины динамики иначе называют темпами роста. Они могут быть выражены в коэффициентах или процентах.

Относительные показатели структуры характеризуют состав изучаемой совокупности, доли, удельные веса элементов совокупности в общем итоге и представляют собой отношение части единиц совокупности (f_i) ко всему объему совокупности (f):

$$d = \frac{f_i}{\sum f_i} * 100\%$$

где d – удельный вес частей совокупности.

Пример. Имеются следующие данные о розничном товарообороте региона за 2001-2002 гг., млрд. руб.:

Год	1 кварт ал	11 кварт ал	111 кварт ал	1111 кварт ал	Всего за год
2001	173,7	182,4	190,3	206,9	753,3
2002	200,7	205,9	215,1	240,9	862,6

Исчислить относительные величины структуры розничного товарооборота по кварталам за каждый год.

Решение. Исчислим относительные величины структуры розничного товарооборота за 2002 и 2003гг.

2002

2003

$$d_I = \frac{173,7}{753,3} * 100\% = 23,0\%$$

$$d_{II} = \frac{184,2}{753,3} * 100\% = 24,2\%$$

$$d_I = \frac{200,7}{862,6} * 100\% = 23,3$$

$$d_{II} = \frac{205,9}{862,6} * 100\% = 23,9\%$$

Структура розничного товарооборота Российской Федерации по кварталам 2002-2003

Квартал	Удельный вес розничного товарооборота, %	
	2002	2003
I	23,0	23,3
II	24,2	23,9
III	25,3	24,9
IV	27,5	27,9
Итого	100	100

Данные таблицы свидетельствуют о том, что удельный вес розничного товарооборота возрастает с первого по четвертый кварталы.

Относительные показатели интенсивности характеризуют степень насыщенности или развития данного явления в определенной среде, являются именованными показателями и могут выражаться в кратных отношениях, процентах, промилле и других формах.

Пример. Среднегодовая численность населения в регионе составила 2,5 млн. человек, число родившихся – 22 тыс. человек.

Определить число родившихся на каждую 1000 человек населения (относительную величину интенсивности, характеризующую рождаемость).

Решение. Коэффициент рождаемости =
$$\frac{\text{число родившихся}}{\text{среднегодовая числ. населения}} * 1000 = \frac{22000}{2500000} * 1000 = 8,8 \text{ ‰}.$$

На каждую 1000 человек населения рождается 8,8 человека.

Относительные показатели координации характеризуют отношения частей изучаемой совокупности к одной из них, принятой за базу сравнения. Они показывают, во сколько раз одна часть совокупности больше другой или сколько единиц одной части приходится на 1, 10, 100, 1000 единиц другой части. Эти относительные величины могут быть исчислены как по абсолютным показателям, так и по показателям структуры.

Пример. Имеются следующие данные о численности экономически активного населения России млн. человек:

Экономически активное население	72,7
в то числе:	
занятые в экономике	65,9
безработные	6,8

Исчислить, сколько безработных приходится на 1000 занятых в экономике России.

Решение: $\frac{6,8}{65,9} * 1000 = 103,2$ человека.

Следовательно, на каждые 1000 занятых приходится 103,2 безработных.

Относительные показатели сравнения (ОПС) характеризуют отношения одноименных абсолютных показателей, соответствующих одному и тому же периоду или моменту времени, но к различным объектам или территориям.

Пример. В одном районе области размер инвестиций составил 300 млн. руб., в другом 600 млн. руб.

Исчислить относительную величину сравнения, приняв за базу сравнения размер инвестиций первого района.

Решение. ОПС = $600/300 = 2$ раза.

Следовательно, размер инвестиций во втором районе превышает инвестиции первого района в 2 раза.

Задачи для самостоятельного решения.

Задача 1. Имеются следующие данные о выпуске продукции производственным объединением во II квартале, млн. руб.:

Месяц	План на квартал	Фактическое выполнение
Апрель	-	198,0
Май	-	224,4
Июнь	-	237,6
Итого	600,0	660,0

Определите процент выполнения квартального плана нарастающим итогом за II квартал.

Задача 2. Имеются следующие данные о розничном товарообороте Российской Федерации за 2002-2003гг., млрд. руб.:

Показатель	2002	2003
Общий объём товарооборота	753,3	862,6
в том числе:		
продовольственные товары	361,9	420,7
непродовольственные товары	391,4	441,9

Исчислите относительные величины структуры розничного товарооборота. Дайте сравнительный анализ изменения структуры.

Задача 3. Среднегодовая численность населения в отчетном году составила 148,6 млн. человек, число родившихся – 1796,6 тыс. человек, число умерших – 1690,7 тыс. человек; в базисном году – соответственно 147,7 млн., 1304,6 тыс. и 2082,2 тыс. человек.

Определите относительные величины интенсивности показателей естественного движения населения Российской Федерации за отчетный и базисный периоды.

Задача 4. Имеются данные о розничном товарообороте области за 2003 г. млрд. руб.:

Розничный товарооборот -
 всего.....14403,3
 по формам собственности:
 государственная.....
 1445,7
 негосударственная.....12
 957,6
 в том числе частная.....
 10723,1

Определите показатели структуры розничного товарооборота по формам собственности и показатели координации.

3 СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАЦИИ

Прежде чем приступить к практическим занятиям, необходимо понять сущность средней величины, являющейся обобщающейся характеристикой совокупности однотипных явлений по изучаемому признаку, необходимо учесть, что средняя величина должна вычисляться с учетом экономического содержания определяемого показателя.

Средняя арифметическая простая (невзвешенная) равна сумме отдельных значений признака, деленной на число этих значений.

Отдельные значения признака называются вариантами и обозначаются через x ($x_1, x_2, x_3, \dots, x_n$), число единиц в совокупности обозначаются через n , среднее значение признака – через \bar{x} . Следовательно, средняя арифметическая простая равна:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n} = \frac{\sum x}{n}$$

Пример. Имеются следующие данные о производстве рабочими продукции А за смену:

Табельный номер рабочего шт.	Выпущено изделий за смену, шт.	Табельный номер рабочего	Выпущено изделий за смену,
1	16,0	1	17,0
2	17,0	2	18,0
3	18,0	3	20,0

4	17,0	4	21,0
5	16,0	5	18,0

В данном примере варьируется признак – выпуск продукции за смену. Численные значения признака (16,17 и т.д.) называют *вариантами*. Определим выработку продукции рабочими данной группы:

$$\bar{x} = \frac{\sum x}{n} = \frac{16,0+17,0+18,0++\dots+18}{10} = \frac{178}{10} = 17,8 \text{ шт.}$$

Задачи для самостоятельного решения

Задача 1. На механическом заводе рабочий обработал за каждый час деталей: за 1-й – 10 деталей, 2-й – 11, 3-й – 9, 4-й – 10, 5-й – 11, 6-й – 13, 7-й – 8, 8-й – 8. Определите среднюю выработку рабочего за час.

Задача 2. Имеются следующие данные о выпуске продукции по 23 предприятиям отрасли (млн. руб.):

Номер предприятия	Выпуск продукции	Номер предприятия	Выпуск продукции
1	2,8	13	3,4
2	9,4	14	1,3
3	1,9	15	3,4
4	2,5	16	5,0
5	3,5	17	4,9
6	3,2	18	3,6
7	2,3	19	6,0
8	2,5	20	3,2
9	8,6	21	2,9
10	1,5	22	5,6
11	3,2	23	5,4
12	4,2		

Исчислите средний размер продукции на один завод.

Средняя арифметическая взвешенная вычисляется по следующей формуле:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum x f}{\sum f}$$

где x – величина признака (варианта)

f – частоты (веса)

Полученная формула называется средней арифметической взвешенной. Из нее видно, что средняя зависит не только от значений признака, но и от их частот, т.е. от состава совокупности, от ее структуры. Изменим в условии задачи состав рабочих и исчислим среднюю в измененной структуре.

Пример. По цеху имеются данные о заработной плате рабочих:

Месячная заработная плата (x), руб.	Число рабочих f	$x \cdot f$
1100	2	2200
1300	4	5200
1600	8	12800
1900	20	38000
2200	16	35200
	50	93400

Средняя заработная плата одного рабочего составит:

$$\bar{x} = \frac{\sum x f}{\sum f} = \frac{93400}{50} = 1868 \text{ руб.}$$

Частотами (весами) могут быть относительные величины, взятые в процентах или коэффициентах. Метод расчета средней и конечный результат от этого не изменяется.

Пример. Представим данные о численности рабочих в относительных величинах:

Месячная заработная плата (x), руб.	Число рабочих в процентах к итогу (f)	$X \cdot f$	Число рабочих в коэффициентах (f')	$X \cdot f'$
-----------------------------------------	-------------------------------------------	-------------	----------------------------------------	--------------

1100	4	4400	0,04	4,4
1300	8	10400	0,08	10,4
1600	16	25600	0,16	25,6
1900	40	76000	0,40	76,0
2200	32	70400	0,32	70,4
	100,0	186800	1,00	186,8

Средняя заработная плата рабочего, взвешенная по процентным соотношениям, будет равна средней, полученной при решении типового примера 1:

$$\bar{x} = \frac{1100*4 + 1300*8 + 1600*16 + 1900*40 + 2200*32}{100} = \frac{18680}{100} = 186,8 \text{ руб.}$$

Если весами являются частоты, выраженные в коэффициентах, то вычисления упрощаются. Так как сумма коэффициентов всегда равна единице, то расчет средней сводится к определению суммы произведений вариантов на частоты.

Задачи для самостоятельного решения

Задача 1. По фирмам региона имеются следующие данные:

Показатель	№ фирмы				
	1	2	3	4	5
Количество выпущенных акций, тыс. ед.	100	90	110	105	120
Прибыль на одну акцию, руб.	9	8	12	16	14

Определите среднюю прибыль на одну акцию, полученную фирмами региона.

Задача 2. Имеются следующие данные о продаже картофеля на рынках города:

№ рын-ка	Средняя цена, руб.	Продано, тыс. кг.		Удельный вес прода-жи, %	
		март	май	март	май
1	3,5	35	28	25	20
2	3,0	42	42	30	30
3	2,5	63	70	45	50

Определите средние цены картофеля по трем рынкам города в марте и мае, используя в качестве частот: а) Абсолютные данные о продаже товара, б) удельный вес продажи товара.

Объясните, почему при неизменных ценах и общем объеме продажи средняя цена в мае изменилась.

Исчисление средней гармонической

Наряду со средней арифметической в статистике применяется средняя гармоническая величина, обратная средней арифметической из обратных значений признака. Как и средняя арифметическая она может быть простой и взвешенной.

Пример 1. Бригада токарей была занята обточкой одинаковых деталей в течение 8-часового рабочего дня. Первый токарь затрачивал на одну деталь 12 мин., второй — 15 мин., третий — 14, четвертый — 16 и пятый — 14 мин. Определите средняя время, необходимое на изготовление одной детали.

На первый взгляд кажется, что задача легко решается по формуле средней арифметической простой:

$$\bar{x} = \frac{\sum x}{n} = \frac{12+15+14+16+14}{5} = \frac{68}{5} = 13,6$$

Полученная средняя была бы правильной, если бы каждый рабочий сделал только по одной детали. Но в течение дня отдельными рабочими было изготовлено различное число деталей. Для определения числа деталей, изготовлен-

ных каждым рабочим, воспользуемся следующим соотношением

$$\frac{\text{Среднее время, затраченное на одну деталь}}{\text{время}} = \frac{\text{все затраченное}}{\text{число деталей}}$$

Число деталей, изготовленных каждым рабочим, определяется отношением всего времени работы к среднему времени, затраченному на одну деталь. Тогда среднее время, необходимое для изготовления одной детали, равно:

$$\bar{x} = \frac{8 \cdot 60 + 8 \cdot 60 + 8 \cdot 60 + 8 \cdot 60 + 8 \cdot 60}{\frac{8 \cdot 60}{12} + \frac{8 \cdot 60}{12} + \frac{8 \cdot 60}{12} + \frac{8 \cdot 60}{12} + \frac{8 \cdot 60}{12}} = \frac{2400}{40 + 32 + 43,6 + 30 + 34,2} = \frac{2400}{179,8} = 13,3$$

мин.

Это же решение можно представить иначе:

$$\bar{x} = \frac{n}{8 \cdot 60 \left(\frac{1}{12} + \frac{1}{15} + \frac{1}{11} + \frac{1}{16} + \frac{1}{14} \right)} = \frac{5}{0,3747} = 13,3$$

мин.

Таким образом, формула для расчета средней гармонической простой будет иметь вид:

$$\bar{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}}$$

Пример. Издержки производства и себестоимость единицы продукции А по трем заводам характеризуются следующими данными:

Номер завода	Издержки производства, тыс. руб.	Себестоимость единицы продукции, руб.
1	200	20
2	460	23
3	110	22

Исчислим среднюю себестоимость изделия по трем заводам. Как и прежде главным условием выбора формы средней является экономическое содержание показателя и исходные данные.

Средняя себестоимость
водства

Издержки произ-

единицы продукции (x) = количество
продукции =

$$= \frac{200+460+110}{\frac{200}{20,5} + \frac{460}{23,6} + \frac{110}{22,0}} = \frac{770}{35000} = 22,0$$

руб.

Таким образом, формулу для расчета средней гармонической взвешенной можно представить в общем виде:

$$x = \frac{w_1 + w_2 + w_3 \dots + w_n}{\frac{1}{x_1} w_1 + \frac{1}{x_2} w_2 + \frac{1}{x_3} w_3 + \dots + \frac{1}{x_n} w_n}$$

Задачи для самостоятельного решения

Задача 1. По семи цехам завода имеются данные о расходовании материала на производство продукции:

Номер цеха	Расход материала, м.		Номер цеха	Расход материала, м.	
	На одно изделие	На все изделия		На одно изделие	На все изделия
1	0,6	150	5	0,5	250
2	0,7	126	6	1,3	260
3	0,9	261	7	1,4	420
4	0,4	200			

Определите расход материала на одно изделие в среднем по заводу.

Задача 2. Имеются следующие данные о заработной плате рабочих завода:

Номер цеха	Базисный период		Отчетный период	
	Средняя заработная плата, руб.	Численность рабочих, чел.	Средняя заработная плата, руб.	Фонд заработной платы, руб.
1	110	300	110	27500

2	145	400	152	68400
3	160	200	170	40800
4	180	100	210	33600

Исчислите среднюю заработную плату рабочего в целом по заводу: 1) в базисном периоде; 2) в отчетном периоде. Сравните полученные данные. Укажите, какие виды средних необходимо применить в каждом случае.

Задача 3. выработка ткани по цехам фабрики характеризуется следующими данными:

Номер цеха	Базисный период		Отчетный период	
	Численность рабочих, чел.	Средняя выработка ткани за смену одним рабочим, м.	Выработано ткани всего, м.	Средняя выработка ткани за смену одним рабочим, м.
1	40	74	3555	79
2	60	85	5160	86
3	50	80	4565	83

Определите среднюю выработку ткани по заводу за смену одним рабочим: 1) в базисном периоде; 2) в отчетном периоде. Сравните полученные данные. Укажите, какие виды средних необходимо применить в каждом случае

Задача 4. Прибыль предприятий акционерного общества характеризуется показателями:

№ предприятия	Получено прибыли в отчетном периоде, тыс. руб.	Выполнение планового задания по прибыли к базисному периоду, %
1	231	105
2	117	90
3	136	85

Определите:

1) средний процент выполнения предприятиями плана по прибыли;

2) сколько недополучено прибыли АО в отчетном периоде по сравнению с базисным.

Задача 5. Имеются данные о финансовых показателях фирм за два периода:

№ группы	Базисный период		Отчетный период	
	Прибыль на одну акцию, руб.	Количество акций, тыс.	Прибыль на одну акцию, руб.	Сумма прибыли, тыс. руб.
1	8,0	60	9,0	810
2	4,0	40	8,0	480

Определите среднюю прибыль на одну акцию по двум фирмам в каждом периоде.

Исчисление моды и медианы

Характеристиками вариационных рядов наряду со средними являются мода и медиана. Мода — есть величина признака (варианта), наиболее часто повторяющаяся в изучаемой совокупности. Для дискретных рядов распределения модой будет значение варианты с наибольшей частотой.

Пример. Распределение проданной обуви по размерам характеризуется

следующими показателями:								
Размер обуви		36	37	38	39	40	41	42
43	44	45 и выше						
Число пар, в процентах		-	1	6	8	22	30	20
11	1	1						

В этом ряду распределения мода равна 41. Именно этот размер обуви пользовался наибольшим спросом покупателей.

Для интервальных рядов распределения с равными интервалами мода определяется по формуле:

$$M_0 = x_{m_0} + i_{m_0} * \frac{f_{m_0} - f_{m_{0-1}}}{(f_{m_0} - f_{m_{0-1}}) + (f_{m_0} - f_{m_{0+1}})}$$

где X_{m_0} , — начальное значение интервала, содержащего моду;

i_{m_0} , — величина модального интервала;

f_{m_0} , — частота модального интервала;

f_{m_0-1} — частота интервала, предшествующего модальному;

f_{m_0+1} — частота интервала, следующего за модальным.

Пример. Рассмотрим пример расчета моды. Распределение предприятий по численности промышленно-производственного персонала характеризуется следующими данными:

Группы предприятий по числу работающих, чел.	Число предприятий	Группы предприятий по числу работающих, чел.	Число предприятий
100-200	1	500-600	19
200-300	3	600-700	15
300-400	7	700-800	5
400-500	30		
			80

В этой задаче наибольшее число предприятий (30) имеет численность работающих от 400 до 500 человек. Следовательно, этот интервал является модальным интервалом ряда распределения. Введем следующие обозначения:

$$X_{m_0} = 400$$

$$i_{m_0} = 100$$

$$f_{m_0} = 30$$

$$f_{m_0-1} = 7$$

$$f_{m_0+1} = 19.$$

Подставим эти значения в формулу моды и произведем вычисления:

$$M_0 = X_{m_0} + i_{m_0} * \frac{f_{m_0} - f_{m_0-1}}{(f_{m_0} - f_{m_0-1}) + (f_{m_0} - f_{m_0+1})} = 400 + 100 \frac{30 - 7}{(30 - 7) + (30 - 19)} = 400 + 100 \frac{23}{23 + 11} = 467$$

человек.

Задачи для самостоятельного решения

Задача 1. По данным выборочного обследования семей области получено следующее распределение семей по размеру совокупного дохода на члена семьи:

Размер совокупного доход			
На члена семьи, руб.	650	800	1100
1300	1600	свыше 1600	
Число семей в процентах			
к итогу	5	12	42
19	10	12	

Определите моду среднедушевого дохода семей.

Задача 2. Себестоимость единицы одноименной продукции по предприятиям отрасли характеризуется следующими показателями:

Группы предприятий по себестоимости единицы продукции, руб.	Число предприятий
1,6-2,0	2
2,0-2,4	3
2,4-2,8	5
2,8-3,2	7
3,2-3,6	10
3,6-4,0	3
Итого	30

Определите моду себестоимости единицы продукции.

Расчет медианы

Медианой в статистике называется варианта, расположенная в середине вариационного ряда. Если ряд распределения дискретный и имеет нечетное число членов, то медианой будет варианта, находящаяся в середине упорядоченного ряда (упорядоченный ряд — это расположение еди-

ниц совокупности в возрастающем или убывающем порядке). Например, стаж пяти рабочих составил 2, 4, 7, 8 и 10 лет. В таком упорядоченном ряду медиана — 7 лет. По обе стороны от нее находится одинаковое число рабочих.

Если упорядоченный ряд состоит из четного числа членов, то медианой будет средняя арифметическая из двух вариантов, расположенных в середине ряда. Пусть теперь будет не пять человек в бригаде, а шесть, имеющих стаж работы 2, 4, 6, 7, 8 и 10 лет. В этом ряду имеются две варианты, стоящие в центре ряда. Это варианты 6 и 7.

Средняя арифметическая из этих значений и будет медианой ряда:

$$Me \frac{6+7}{2} = 6,5 \text{ лет}$$

Рассмотрим пример расчета медианы в дискретном ряду.

Пример. Определим медиану заработной платы рабочих.

Месячная заработная плата, руб.	Число рабочих	Сумма накопленных частот
1	2	3
1000	2	2
1200	6	8 (2+6)
1500	16	24 (8+16)
1700	12	-
2000	4	-
Итого 40		

Для определения медианы надо подсчитать сумму накопленных частот ряда (гр. 3 табл.). Нарращивание итога продолжается до получения накопленной суммы частот, превышающей половину. В нашем примере сумма частот составила 40, ее половина — 20. Накопленная сумма частот ряда получилась равной 24. Варианта, соответствующая этой сумме, т. е. 1500 руб., и есть медиана ряда.

Если же сумма накопленных частот против одной из вариантов равна точно половине суммы частот, то медиана

определяется как средняя арифметическая этой варианты и последующей.

Медиана интервального вариационного ряда распределения определяется по формуле

$$Me = x_{Me} + i_{Me} \frac{0,5 \sum f - S_{Me-1}}{f_{Me}}$$

где x_{Me} — начальное значение интервала, содержащего медиану;

i_{Me} — величина медианного интервала;

$\sum f$ — сумма частот ряда;

S_{Me-1} — сумма накопленных частот, предшествующих медианному интервалу;

f_{Me} — частота медианного интервала.

Пример. Рассчитаем медиану в интервальном вариационном ряду.

Группы предприятий по числу рабочих	Число предприятий	Сумма накопленных частот
100 — 200	1	1
200 — 300	3	4 (1+3)
300 — 400	7	11 (4+7)
400 — 500	30	41 (11+30)
500 — 600	19	-
600 — 700	15	-
700 — 800	5	-
Итого 80		

Определим прежде всего медианный интервал. В данной задаче сумма накопленных частот, превышающая половину всех значений (41), соответствует интервалу 400 — 500. Это и есть медианный интервал, в котором находится медиана.

Определим ее значение по приведенной выше формуле. Известно, что:

$$x_{Me} = 400;$$

$$i_{Me} = 100;$$

$$\sum f = 80;$$

$$S_{Me-1} = 11;$$

$$f_{Me} = 30.$$

Следовательно,

$$Me = x_{Me} + i_{Me} \frac{0,5 \sum f - S_{Me-1}}{f_{Me}} = 400 + 100 \frac{0,5 * 80 - 11}{30} = 400 + 100 \frac{40 - 11}{30} = 400 + 99,66 = 499,66$$

РАСЧЕТ ДИСПЕРСИИ И СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ ПО ИНДИВИДУАЛЬНЫМ ДАННЫМ В РЯДАХ РАСПРЕДЕЛЕНИЯ

Основными обобщающими показателями вариации в статистике являются дисперсия и среднее квадратическое отклонение.

Дисперсия — это средняя арифметическая квадратов отклонений каждого значения признака от общей средней. Дисперсия обычно называется средним квадратом отклонений и обозначается σ^2 . В зависимости от исходных данных дисперсия может вычисляться по средней арифметической простой или взвешенной:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} \quad \text{— дисперсия невзвешенная (простая);}$$

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} \quad \text{— дисперсия взвешенная.}$$

Среднее квадратическое отклонение представляет собой корень квадратный из дисперсии и обозначается σ :

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad \text{— среднее квадратическое отклонение невзвешенное;}$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} \quad \text{— среднее квадратическое отклонение взвешенное.}$$

Среднее квадратическое отклонение — это обобщающая характеристика абсолютных размеров вариации признака в совокупности. Выражается оно в тех же единицах измерения, что и признак (в метрах, тоннах, процентах, гектарах и т. д.). Вычислению среднего квадратического отклонения предшествует расчет дисперсии.

Пример. Исчислим дисперсию по данным задачи

Произведено продукции одним рабочим штук	Число рабочих	$x \cdot f$	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
8	7	56	-2	4	28
9	10	90	-1	1	10
10	15	150	0	0	0
11	12	132	1	1	12
12	6	72	2	4	24
Итого	50	500			74

Исчислим среднюю арифметическую взвешенную:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{500}{50} = 10 \quad \text{шт.}$$

Значения отклонений от средней и их квадратов представлены в таблице.

Определим дисперсию:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} = \frac{74}{50} = 1,48$$

Среднее квадратическое отклонение будет равно:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,48} = 1,216 \quad \text{шт.}$$

Если исходные данные представлены в виде интервального ряда распределения, то сначала надо определить дискретное значение признака, а далее применить тот же метод, что изложен выше.

РАСЧЕТ ГРУППОВОЙ, МЕЖГРУППОВОЙ И ОБЩЕЙ ДИСПЕРСИИ

(ПО ПРАВИЛУ СЛОЖЕНИЯ ДИСПЕРСИЙ)

Если совокупность разбита на группы (или части) по изучаемому признаку, то для такой совокупности могут быть исчислены следующие виды дисперсий: общая, групповые (частные), средняя из групповых (частных), межгрупповая.

Общая дисперсия равна среднему квадрату отклонений отдельных значений признака x от общей средней \bar{X} . Она может быть исчислена как простая средняя или как взвешенная соответственно по формулам:

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

Общая дисперсия отражает вариацию признака за счет всех условий и причин, действующих в совокупности.

Групповая (частная) дисперсия равна среднему квадрату отклонений отдельных значений признака внутри группы от средней арифметической этой группы (групповой средней). Она может быть исчислена как простая средняя или как взвешенная соответственно по формулам:

$$\sigma_{i^2} = \frac{\sum (x - \bar{x}_i)^2}{n} ; \quad \sigma_{i^2} = \frac{\sum (x - \bar{x}_i)^2 f}{\sum f}$$

Эта дисперсия отражает вариацию признака только за счет условий и причин, действующих внутри группы.

Средняя из групповых (частных) дисперсий — это средняя арифметическая, взвешенная из дисперсий групповых:

$$\bar{\sigma}_{i^2} = \frac{\sum \sigma_{i^2} f}{\sum f}$$

Межгрупповая дисперсия равна среднему квадрату отклонений групповых средних \bar{x}_i от общей средней \bar{x} :

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f}{\sum f}$$

Межгрупповая дисперсия характеризует вариацию результативного признака за счет группировочного признака.

Между указанными видами дисперсий существует определенное соотношение: общая дисперсия равна сумме средней из групповых дисперсий и межгрупповой дисперсии:

$$\sigma^2 = \bar{\sigma}_{i^2} + \delta^2$$

Это соотношение называют правилом сложения дисперсий. С его помощью, зная два вида дисперсий, можно определить третий:

$$\delta^2 = \sigma^2 - \overline{\sigma_{i^2}}; \overline{\sigma_{i^2}} = \sigma^2 - \delta^2$$

Пример. Имеются следующие данные о производительности ткачей за час работы

Табельный номер рабочего	Изготовлено продукции рабочим, штук	\bar{x} - x_1	$(x - x_1)^2$	Табельный номер рабочего	Изготовлено продукции рабочим, штук	\bar{x} - x_1	$(x - x_1)^2$
1	13	2	4	7	18	3	9
2	14	-	1	8	19	2	4
3	15	1	0	9	22	2	1
4	17	0	4	10	20	1	1
5	16	2	1	11	24	-	9
6	15	1	0	12	23	1	4
	90	0	10		126	3	28

Исчислим: 1) групповые дисперсии 2) среднюю из групповых дисперсий; 3) межгрупповую дисперсию; 4) общую дисперсию.

1. Для расчета групповых дисперсий исчислим средние по каждой группе:

$$\bar{x}_1 = \frac{90}{6} = 15; \bar{x}_2 = \frac{126}{6} = 21.$$

Расчет дисперсий по группам представлен в таблице. Подставив полученные значения в формулу, получим:

$$\sigma_{1^2} = \frac{\sum (x - \bar{x})^2}{n} = \frac{10}{6} = 1,666 \approx 1,67$$

$$\sigma_2^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{28}{6} = 4,66$$

2. Рассчитаем среднюю из групповых (частных) дисперсий:

$$\bar{\sigma}_{i^2} = \frac{\sum \sigma_{i^2} f}{\sum f} = \frac{1,67 * 6 + 4,66 * 6}{12} = \frac{10 + 28}{12} = \frac{38}{12} = 3,16$$

3. Исчислим межгрупповую дисперсию. Для этого предварительно определим общую среднюю как среднюю взвешенную из групповых средних:

$$\bar{x} = \frac{\sum \bar{x}_i f}{\sum f} = \frac{15 * 6 + 21 * 6}{12} = \frac{90 + 126}{12} = 18$$

Затем рассчитаем межгрупповую дисперсию:

4. Исчислим общую дисперсию по правилу сложения дисперсий:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f}{\sum f} = \frac{(15 - 18)^2 * 6 + (21 - 18)^2 * 6}{12} = \frac{9 * 6 + 9 * 6}{12} = \frac{108}{12} = 9.$$

Проверим полученный результат, исчислив общую дисперсию

обычным способом:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{(13-18)^2 + (14-18)^2 + (15-18)^2 + (17-18)^2 + (16-18)^2 + (15-18)^2 + (18-18)^2 + (19-18)^2 + (22-18)^2 + (20-18)^2 + (24-18)^2 + (23-18)^2}{12} = \frac{146}{12} = 12,16$$

Задачи для самостоятельного решения

Задача 1. Бригада рабочих механического цеха, состоящая из 10 человек, к концу месяца имела следующие показатели по выполнению норм выработки:

Группы рабочих по степени выполнения плана, %	Процент выполнении плана
до 100	90, 95, 85, 92
свыше 100	100, 102, 104, 103, 105, 104

Исчислите: 1) групповые дисперсии; 2) межгрупповую дисперсию; 3) общую дисперсию (обычным способом и по правилу сложения дисперсий).

Задача 2. Имеются показатели распределения основных фондов по заводам отрасли:

Группы заводов по стоимости основных фондов, млн. руб.	Число заводов	Основные фонды в среднем на завод. млн. руб.	Групповые дисперсии
1,2 — 2,7	9	1,8	0,17
2,7 — 4,2	11	3,2	0,09
4,2 — 5,7	7	4,8	0,25
5,7 — 7,2	3	6,9	0,14

Определите общую дисперсию основных фондов по совокупности заводов, применяя правило сложения дисперсий.

Задача 3. Имеются следующие данные о распределении рабочих по проценту допускаемого брака в процессе производства:

Процент брака	Число рабочих	Средний процент брака продукции на одного рабочего	Среднее квадратическое отклонение
До 1	7	0,8	0,67
1-3	20	2,3	0,65
3-5	15	3,7	0,51
5-7	5	5,9	0,48
Свыше 7	3	7,8	0,82

Исчислите общую дисперсию допускаемого рабочим брака продукции, применяя правило сложения дисперсий.

РАСЧЕТ ЭМПИРИЧЕСКОГО КОРРЕЛЯЦИОННОГО ОТНОШЕНИЯ

Правило сложения дисперсий используется в статистике для определения степени связи между изучаемыми признаками. Методом аналитической группировки, путем сравнения групповых средних показателей результативного признака,

изменяющегося под влиянием факторного, была установлена зависимость между размером основных фондов и выпуском продукции. Но групповые средние выпуска продукции на один завод являются результатом влияния на только факторного признака, но и других причин. поэтому наряду с выявлением взаимосвязи между интересующими нас признаками стоит задача количественного определения тесноты связи между ними. Метод аналитических группировок с применением правила сложения дисперсий позволяет определить тесноту связи между признаками с помощью эмпирического корреляционного отношения.

Первоначально исчислим коэффициент детерминации:

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

Он представляет отношение дисперсии групповых средних к общей дисперсии и показывает, какую часть общей вариации изучаемого признака составляет вариация межгрупповая, т.е. обусловленная группировочным признаком.

Корень из коэффициента детерминации называют *эмпирическим корреляционным отношением*

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

Пример. Покажем расчет эмпирического корреляционного отношения по данным аналитической группировки и определим степень связи между размером основных фондов и выпуском продукции на один завод, используя значения, приведенные в таблице.

Группы заводов по величине основных фон-	Число заводов f	Производство на один завод, млн . руб.	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^2$
			$y_i - 4,78$	$(y_i - 4,78)^2$	$(y_i - 4,78)^2$

ДОВ, млн. руб.					
1 —	3	1,87	— 2,9	8,41	25,230
2,2	9	2,95	—	3,348	0
2,2	5	4,60	1,83	9	30,140
— 3,4	3	5,30	—	0,032	1
3,4	4	10,9	0,18	4	0,1620
— 4,6		5	0,52	0,270	0,8112
4,6			6,17	4	152,27
— 5,8				38,07	56
5,8					
— 7,0					
Итог	24	4,78			208,61
о					89

Прежде всего определим средний выпуск продукции на один завод по отрасли в целом. Общая средняя будет равна:

$$\bar{x} = \frac{\sum y}{n} = \frac{114,8}{24} = 4,78 \text{ млн. руб.}$$

Определим дисперсию, характеризующую вариацию выпуска продукции за счет изменения основных фондов, т. е. межгрупповую дисперсию. Межгрупповая дисперсия равна:

$$\delta^2 = \frac{\sum (y_i - \bar{y})^2 f}{\sum f} = \frac{208,6189}{24} = 8,69$$

Исчислим общую дисперсию по индивидуальным данным, используя формулу

$$\sigma^2 = \overline{y^2} - (\bar{y})^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n} \right)^2$$

Номер завода	y	y ²	Номер завода	y	y ²	Номер завода	y	y ²
1	3,	1	9	2,	6,	1	3,	1
2	2	0,24	1	5	25	7	6	2,96
3	9,	9	0	3,	1	1	8,	64,0
4	6	2,16	1	5	2,25	8	0	6,25
5	1,	2,	1	2,	5,	1	2,	7,84

6	5	25	1	3	29	9	5	2,56
7	4,	1	2	1,	1,69	2	2,	166,4
8	2	7,64	1	3	1,	0	8	1
	6,	4	3	1,	96	2	1,	31,36
4		0,96	1	4	9,	1	6	19,39
	2,	7,	4	3,	0	2	1	
8		84	1	0	6,	2	2,9	
	9,	8	5	2,	25	2	5,	
4		8,36	1	5	6	3	6	
	1	1	6	7,	2,41	2	4,	
	1,9	41,61		9		4	4	
Итого							1	8
							14,8	16,90

Общая дисперсия равна:

$$\sigma^2 = \frac{816,9}{24} - \left(\frac{114}{24}\right)^2 = 34,04 - 22,85 = 11,19$$

Коэффициент детерминации равен:

$$\gamma_{1^2} = \frac{\delta^2}{\sigma^2} = \frac{8,69}{11,19} = 0,777$$

Корреляционное отношение равно:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{8,69}{11,19}} = \sqrt{0,777} = 0,882$$

Таким образом, можно сказать, что связь между размером основных фондов и выпуском продукции в рассмотренном примере высокая. Вариация выпуска продукции на 77,7% обусловлена вариацией стоимости основных фондов.

РАСЧЕТ КОЭФФИЦИЕНТА АСИММЕТРИИ

Одной из важнейших задач анализа вариационных рядов распределения является выявление закономерности распределения и определение характера. При симметричном распределении частот в вариационном ряду, обобщающие характеристики ряда — средняя арифметическая, мода и медиана — равны между собой.

Однако экономической статистике такое распределение встречается крайне редко. Чаще всего наблюдаются асимметрия распределения, которая может быть правосторонней или левосторонней в зависимости от расположения

частот. Если в ряду распределения преобладают варианты с меньшим, чем средняя арифметическая, значением признака, то вершина кривой распределения сдвинута влево и правая часть кривой оказывается длиннее. Такая асимметрия называется правосторонней (положительной). Если же преобладают варианты с большим, чем средняя арифметическая, значением признака — вершина кривой распределения сдвинута вправо, и, следовательно, левая часть кривой получается длиннее правой. Такая асимметрия называется левосторонней (отрицательной).

Степень асимметрии может быть определена с помощью коэффициента

асимметрии:

$$K_A = \frac{\bar{x} - Mo}{\sigma}$$

где \bar{x} — средняя арифметическая ряда распределения;

Mo — мода;

σ — среднее квадратическое отклонение.

Задачи для самостоятельного решения

Задача 1. В результате обследования получены следующие данные о

распределении семей по размеру совокупного дохода:

Группы семей по размеру дохода, руб.	Число семей в процентах к итогу	Группы семей по размеру дохода, руб.	Число семей в процентах к итогу
До 1000	3,0	2500-3000	11,0
1000-1500	35,0	300-3500	14,0
1500-2000	20,0	Свыше	7,0
2000-2500	10,0	3500	
		Итого	100,0

Определите коэффициент асимметрии данного ряда распределения.

Задача 2. Имеются данные о размерах товарооборота магазинов государственной торговли за IV квартал отчетного года:

Группы торговых предприятий по размеру товарооборота, млн. руб.	Число торговых предприятий	Группы торговых предприятий по размеру товарооборота, млн. руб.	Число торговых предприятий
200 — 300	396	500 — 600	130
300 — 400	270	600 — 700	90
400 — 500	187	700 — 800	57
Итого			1130

Определите коэффициент асимметрии распределения магазинов по размеру товарооборота.

4 ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

При статистическом исследовании экономических явлений часто производится наблюдение не всех единиц совокупности, а лишь их части и по этой части судят о всей совокупности в целом. Это возможно при образовании выборочной совокупности в условиях стохастического (вероятностного) процесса. Отбор единиц из генеральной совокупности производится таким образом, чтобы выборочная совокупность была представлена (репрезентативна) и правильно характеризовала генеральную совокупность. Однако полной репрезентативности выборки достичь не удастся вследствие несоответствия состава выборочной совокупности составу генеральной совокупности. Поэтому необходима оценка надежности результатов выборки и возможности их распространения на генеральную совокупность. Надежность результатов выборки проверяется расчетом случайной ошибки выборки или ошибки репрезентативности. Ошибки выборки рассчитываются по формулам, разработанным в теории вероятностей для каждого вида выборки.

Решение типовых задач при собственно случайном типическом методе.

Средняя ошибка выборки для средней показывает среднюю величину всех возможных расхождений выборочной и генеральной средней.

При случайном повторном отборе средняя ошибка выборочной средней рассчитывается по формуле:

$$\mu = \sqrt{\frac{\sigma^2}{n}}$$

где μ — средняя ошибка выборочной средней; σ — дисперсия выборочной совокупности; n — численность выборки. При бесповторном отборе она рассчитывается по формуле

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

где N — численность генеральной совокупности.

Предельная ошибка выборки Δ рассчитывается по формуле $\Delta = \mu t$, где t — коэффициент доверия, зависит от значения вероятности P .

Значения t при заданной вероятности P определяются по таблице значений функции $\phi(t)$ которая выражается интегральной формулой Лапласа, и отражают зависимость между t и вероятностью P .

При механическом отборе средняя ошибка выборки рассчитывается по формуле собственно-случайного бесповторного отбора.

Пример. Методом случайной повторной выборки было взято для проверки на вес 200 шт. деталей. В результате был установлен средний вес детали 30 г при среднем квадратическом отклонении 4 г. С вероятностью 0,954 требуется определить пределы, в которых находится средний вес деталей в генеральной совокупности.

Генеральная средняя $\bar{\chi}$ отличается от выборочной средней $\tilde{\chi}$ на величину ошибки выборки Δ :

$$\bar{\chi} = \tilde{\chi} \pm \Delta_{\tilde{\chi}}$$

Чтобы определить границы генеральной средней с вероятностью 0,954, необходимо рассчитать предельную ошибку

ку выборочной средней. Предельная ошибка выборки для средней определяется по формуле при повторном отборе:

$$\Delta_{\bar{x}} = t \sqrt{\frac{\sigma^2}{n}}$$

С вероятностью 0,954 ошибка выборки не превысит двух средних ошибок, так как значение t при $P=0,954$ равно 2.

Подставим значения в формулу ошибки выборки:

$$\Delta_{\bar{x}} = t \sqrt{\frac{4^2}{200}} = 0,56 \text{ г.}$$

Определим верхнюю границу генеральной средней:

$$\bar{x} = \tilde{x} + \Delta_{\bar{x}} = 30 \text{ г.} + 0,56 \text{ г.} = 30,56 \text{ г.}$$

Определим нижнюю границу генеральной средней

$$\bar{x} = \tilde{x} - \Delta_{\bar{x}} = 30 - 0,56 = 29,44 \text{ г.}$$

С вероятностью 0,954 можно утверждать, что средний вес детали в генеральной совокупности находится в пределах $29,44 \leq \bar{x} \leq 30,56$ г.

Пример. В районе А проживает 2500 семей. Для установления среднего числа детей в семье была проведена 2%-ная случайная бесповторная выборка семей. В результате обследования были получены следующие данные:

Число детей в семье	0	1	2	3	4	5
Число семей	1 0	2 0	1 2	4	2	2

С вероятностью 0,997 требуется определить границы, в которых будет находиться среднее число детей в семье в генеральной совокупности (в городе А). Генеральная средняя

$$\bar{x} = \tilde{x} \pm \Delta_{\bar{x}}$$

Чтобы определить границы генеральной средней, необходимо рассчитать выборочную среднюю и ошибку выборочной средней. Рассчитаем среднее число детей в семье в выборочной совокупности, и дисперсию выборочной совокупности:

Число детей в семье x	Количество семей f	xf	$x - \tilde{x}$	$(x - \tilde{x})^2$	$(x - \tilde{x})^2 f$
0	10	0	-1,5	2,25	22,5
1	20	20	- 0,5	0,25	5,0
2	12	24	+ 0,5	0,25	3,0
3	4	12	+ 1,5	2,25	9,0
4	2	8	+ 2,5	6,25	12,5
5	2	10	+ 3,5	12,25	24,5
Итого	50	74	-	-	76,5

$$\tilde{x} = \frac{\sum xf}{\sum f} = \frac{74}{50} = 1,5 \text{ чел.}$$

$$\sigma^2 = \frac{\sum (x - \tilde{x})^2 f}{\sum f} = \frac{76,5}{50} = 1,53 \approx 1,5$$

Предельная ошибка выборочной средней при бесповторном случайном отборе рассчитывается по формуле.

$$\Delta_{\tilde{x}} = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

С вероятностью 0,997 наша ошибка выборки не превышает трех средних ошибок:

$$\Delta_{\tilde{x}} = 3 \sqrt{\frac{1,5}{50} \left(1 - \frac{50}{2500}\right)} = 0,5 \text{ чел.}$$

Определим пределы, в которых находится среднее число детей в семье в городе А:

$$\bar{x} = \tilde{x} \pm \Delta_{\tilde{x}} = 1,5 \pm 0,5$$

С вероятностью 0,997 можно утверждать, что среднее число детей в семье в городе А находится в пределах $1,0 \pm \bar{\chi} \leq 2,0$.

При повторном отборе средняя ошибка выборочной доли рассчитывается по формуле

$$\mu = \sqrt{\frac{w(1-w)}{n}}$$

где $\omega = \frac{m}{n}$ выборочная доля, доля единиц, обладающих изучаемым признаком;

m — число единиц, обладающих изучаемым признаком;

n — численность выборки.

Пример. При обследовании 100 образцов изделий, отобранных из партии в случайном порядке, оказалось 20 нестандартных. С вероятностью 0,954 определите пределы, в которых находится доля нестандартной продукции в партии.

Генеральная доля равна: $P = w \pm \Delta_w$. Чтобы определить границы генеральной доли, необходимо определить выборочную долю и ошибку выборочной доли.

Рассчитаем долю нестандартной продукции в выборочной совокупности:

$$\omega = \frac{m}{n} \qquad w = \frac{20}{100} = 0,2$$

Предельная ошибка выборочной доли с вероятностью 0,954 составит:

$$\Delta = t \sqrt{\frac{w(1-w)}{n}} = 2 \sqrt{\frac{0,2 * 0,8}{100}} = 0,08$$

Определим нижнюю границу генеральной доли $\omega - \Delta_\omega = 0,2 - 0,08 = 0,12$

Определим верхнюю границу генеральной доли $\omega + \Delta_\omega = 0,2 + 0,08 = 0,28$. С вероятностью 0,954 можно утверждать, что доля нестандартной продукции в партии товара находится в пределах $12\% \leq p \leq 28\%$

При бесповторном способе отбора средняя ошибка выборочной доли определяется по формуле:

$$\mu = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$$

Пример. В городе 500 тыс. жителей.. По материалам учета городского населения было обследовано 50 тыс. жителей методом случайного бесповторного отбора. В результате обследования установлено, что в городе 15% жителей старше 60 лет. С вероятностью 0,683 определите пределы, в которых находится доля жителей в городе в возрасте старше 60 лет.

Генеральная доля равна $P \pm \Delta_w$. Выборочная доля равна $\omega = 15\%$. С вероятностью 0,683 определим ошибку выборки для доли:

$$\Delta_w = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} = 1 * \sqrt{\frac{0,15 * 0,85}{50} * 0,9} = 0,048 \approx 0,05 \text{ или } 5\%$$

Определим верхнюю границу генеральной доли: $P_v = 0,15 + 0,05 = 0,20$ (или 20%).

Определим нижнюю границу генеральной доли: $P_n = 0,15 - 0,05 = 0,10$ (или 10%).

С вероятностью 0,683 можно утверждать, что доля жителей в возрасте старше 60 лет в городе А находится в пределах $10\% \leq p \leq 20\%$.

Решение типовых задач. Определение необходимой численности выборки.

В практике проектирования выборочного наблюдения возникает потребность определения численности выборки, которая необходима для обеспечения определенной точности расчета генеральных средних. Предельная ошибка выборки и ее вероятность при этом являются заданными.

При бесповторном случайном отборе необходимая численность выборки вычисляется по формуле

$$n = t^2 \sigma^2 N / N \Delta^2 + t^2 \sigma^2$$

Задача 1. В районе А проживает 2000 семей. В порядке случайной бесповторной выборки предполагается определить средний размер семьи при условии, что ошибка выборочной средней не должна превышать 0,8 человека с вероят-

ностью $P = 0,954$ и при среднем квадратическом отклонении 2,0 человека. Рассчитаем необходимую численность выборки:

$$N = 4 \cdot 4,0 \cdot 2000 / 2000 \cdot 0,64 + 4 \cdot 4,0 = 24 \text{ семьи}$$

При повторном случайном отборе численность выборки определяется по формуле:

$$N = t^2 \sigma^2 / \Delta^2$$

Задача 2. Для определения средней длины детали необходимо провести выборочное обследование методом случайного повторного отбора. Какое количество деталей надо отобрать, чтобы ошибка выборки не превышала 2 мм, с вероятностью 0,954 при среднем квадратическом отклонении 8 мм?

Рассчитаем необходимую численность выборки:

$$n = 2^2 \cdot 8^2 / 4 = 64 \text{ детали.}$$

При случайном бесповторном отборе для расчета необходимой численности выборки для определения доли с заданной точностью применяется следующая формула:

$$n = t^2 w(1-w)N / \Delta^2 N + t^2 w(1-w)$$

или по формуле

$$n = 0,25 t^2 N / \Delta^2 N + 0,25 t^2$$

если дисперсия доли неизвестна.

Задача 3. В городе А 10 тыс. семей. В порядке механической выборки предполагается определить долю семей в городе А с числом детей три и более. Какова должна быть численность выборки, чтобы с вероятностью 0,954 ошибка выборки не превышала 0,02 человека, если на основе предыдущих обследований известно, что дисперсия равна 0,2.

Рассчитаем необходимую численность выборки:

$$n = 4 \cdot 0,2 \cdot 10\,000 / (0,02)^2 \cdot 10\,000 + 4 \cdot 0,2 = 1\,666 \text{ семей}$$

При повторном способе отборки численность выборки рассчитывается по формуле:

$$n = w(1-w) t^2 / \Delta^2$$

При типической (районированной) выборке генеральная совокупность разбивается на однородные типические группы по какому-либо признаку или районы. Из каждой типической группы или района в случайном порядке отбираются единицы выборочной совокупности. Отбор единиц из типов может

производиться тремя методами: пропорционально численности единиц типических групп, непропорционально численности единиц типических групп, пропорционально колеблемости в группах.

Рассмотрим типическую выборку с пропорциональным отбором единиц из типических групп. Объем выборки из типической группы при отборе, пропорциональном численности единиц типических групп, определяется по формуле

$$n_i = n \cdot (N_i / N)$$

где n_i — объем выборки из типической группы;

n — общий объем выборки;

N_i — объем типической группы;

N — объем генеральной совокупности.

Средняя ошибка выборочной средней при бесповторном случайном и механическом способе отбора внутри типических групп рассчитывается по формуле

$$\mu = \sqrt{\frac{\bar{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)}$$

где $\bar{\sigma}^2$ — средняя из выборочных дисперсий типических групп.

Задача 4. В районе 10 тыс. семей. Из них 5 тыс. семей рабочих, 4 тыс. семей колхозников, 1 тыс. семей служащих. Для определения числа детей в семье была проведена 10%-ная типическая выборка с отбором единиц пропорционально численности единиц типических групп. Внутри групп применялся метод механического отбора. Результаты выборки представлены в таблице.

Таблица

Типы семей	Число семей в генеральной совокупности	Среднее число детей в семье, чел.	Среднее квадратическое отклонение, чел
Рабочие	5 000	2,3	1,2
Служащие	1 000	1,8	0,5

Колхоз-ники	4 000	2,8	2,5
-------------	-------	-----	-----

С вероятностью 0,997 определите пределы, в которых находится среднее число детей в семье в районе.

Рассчитаем объем выборки в каждой типической группе при условии, что численность выборочной совокупности равна 1000 семей:

$$n_1 = 1000 \cdot (5000 / 10\,000) = 500 \text{ семей}$$

$$n_2 = 1000 \cdot (1000 / 10\,000) = 100 \text{ семей}$$

$$n_3 = 1000 \cdot (4000 / 10\,000) = 400 \text{ семей}$$

Рассчитаем общую выборочную среднюю из групповых выборочных средних путем взвешивания последних по численности отобранных групп:

$$\bar{x} = \frac{\sum \bar{x}_i n_i}{\sum n_i} = \frac{2,3 \cdot 500 + 1,8 \cdot 100 + 2,8 \cdot 400}{1000} = 2,45 \text{ чел.}$$

Средний размер семьи в выборке равен 2,45 человека. Определим среднюю из внутригрупповых дисперсий:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i} ;$$

$$\bar{\sigma}^2 = \frac{1,44 \cdot 500 + 0,25 \cdot 100 + 6,25 \cdot 400}{1000} = 3,24$$

Рассчитаем среднюю ошибку выборочной средней типической выборки:

$$\mu = \sqrt{\frac{\bar{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)} ;$$

$$\mu = \sqrt{\frac{3,24}{1000} \left(1 - \frac{1000}{10000}\right)} = 0,053 \text{ чел.}$$

С вероятностью 0,997 определим предельную ошибку выборочной средней:

$$\Delta \hat{=} t_{\mu} = 3 \cdot 0,053 = 0,159 \hat{=} 0,16$$

С вероятностью 0,997 можно утверждать, что в районе среднее число детей в семье находится в пределах $2,3 \leq \bar{x} \leq 2,6$

Задачи для самостоятельного решения.

Задача 1. Имеются выборочные данные о распределении вкладчиков по размеру вкладов в Сбербанк города:

Определите с вероятностью 0,954:

1. Пределы среднего размера вклада в Сбербанк.
2. Пределы удельного веса вкладов с размером св. 10 тыс. руб.

№	Размер вклада, тыс. руб.	Число вкладов
I	До 4	32
II	4-6	56
III	6-8	92
IV	8-10	120
V	Св.10	100
	Итого	400

Задача 2. На заводе с числом рабочих 4000 человек было проведено 2%-ное выборочное обследование возраста рабочих, методом случайного бесповторного отбора. В результате обследования получены следующие данные:

Возраст рабочих, лет	До 30	30-40	40-50	50-60	Свыше 60
Число рабочих	18	42	10	6	4

С вероятностью 0,997 определите пределы, в которых находится средний возраст рабочих завода.

Задача 3. Для изучения производительности труда токарей на машиностроительном заводе было проведено 10%-ное выборочное обследование 100 рабочих методом случайного бесповторного отбора. В результате обследования получены данные о часовой выработке рабочих:

Часовая выработка, шт.	1	2	2	2	2	2
	8-20	0-22	2-24	4-26	6-28	8-30
Число рабочих	2	8	2	5	1	4

С вероятностью 0,997 определите пределы, в которых находится среднее время обработки одной детали токарями завода.

Задача 4. На машиностроительном заводе с числом рабочих 5000 человек было проведено 4%-ное выборочное обследование квалификации рабочих методом случайного бесповторного отбора.

В результате обследования получены следующие данные:

квалификация рабочих (тарифные разряды)	1	2	3	4	5	6
Число рабочих	10	30	40	70	30	20

С вероятностью 0,997 определите пределы, в которых находится средний тарифный разряд рабочих завода.

Задача 5. В районе 2000 семей. С целью определения среднего размера семьи района было проведено 3%-ное выборочное обследование семей методом случайного бесповторного отбора. В результате обследования получены следующие данные:

размер семьи, чел.	0	1	2	3	4	5	6	7	8
Число семей	4	8	1	1	8	4	3	2	1
		4	6						

С вероятностью 0,997 определите пределы, в которых находится средний размер семьи в районе.

Задача 6. Для определения средней дневной заработной платы рабочих магазина произведена 20%-ная типическая выборка с отбором единиц пропорционально численности типических групп (внутри типов применялся -метод случайного бесповторного отбора). Результаты выборки представлены в таблице.

Таблица

Тип магазина	Средняя заработная плата, руб.	Среднее квадратическое отклонение	Число продавцов, чел

		ние, руб.	
I	2 000	50	150
II	2 100	200	500
III	2 150	100	350

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

Задача 7. В целях изучения производительности четырех типов станков, производящих одни и те же операции, была произведена 10%-ная типическая выборка с отбором единиц пропорционально численности типических групп (внутри групп применялся метод случайного бесповторного отбора).

Т ип стан- ка	Число отобранных станков	Среднее число деталей, изготовлен- ных на станке за час работы, шт.	Среднее квадратическое отклонение, шт.
I	15	400	40
II	30	520	20
III	45	700	50
IV	10	610	70

С вероятностью 0,997 определите предел, в котором находится среднее число деталей, производимых на одном станке за 1 ч работы для всей совокупности станков.

5 РЯДЫ ДИНАМИКИ.

При изучении ряда динамики возникает необходимость получения обобщающей величины его абсолютных уровней. Для этого определяют средний уровень ряда как среднюю величину из совокупности абсолютных уровней ряда динамики за те или иные периоды времени. Такая обобщающая величина в рядах динамики называется средней хронологической. Методы ее расчета зависят от вида ряда динамики и способов получения статистических данных.

В интервальном ряду динамики расчет среднего уровня ряда производится по методу средней арифметической простой (невзвешенной) :

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

где y — абсолютные уровни ряда;

n — число уровней.

Пример. Имеются следующие данные о производстве товара «А» области, млн. руб.:

2000	2001	2002	2003	2004
7,6	9,1	7,8	8,4	9,6

Необходимо определить средний уровень валового сбора овощей за 2000 -2004 годы.

Определим средний уровень данного интервального ряда:

$$\bar{y} = \frac{\sum y}{n} = \frac{7,6+9,1+7,8+8,4+9,6}{5} = \frac{42,5}{5} = 8,5 \text{ млн. руб.}$$

В моментном ряду динамики с равноотстоящими датами определение среднего уровня ряда производится по формуле средней хронологической моментного ряда динамики:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \dots + y_{n-1} + \frac{1}{2}y_n}{n-1}$$

Пример. По следующим данным о товарных запасах в розничной сети торгующих организаций города определить величину среднеквартального запаса за 2002 г., млн. руб.:

1/I 2002 г.	64,1
1/IV 2002 г.	57,8
1/VII 2002 г.	60,0
1/X 2002 г.	63,2
1/I 2003 г.	72,3

Определим средний уровень данного моментного ряда динамики:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + y_4 + \frac{1}{2}y_5}{5-1} = \frac{\frac{64,2}{2} + 57,8 + 60,0 + 63,2 + \frac{72,3}{2}}{5-1} = \frac{249,2}{4} = 62,3 \text{ млн. руб.}$$

В практике экономической работы часто приходится определять средние уровни ряда моментных величин с неравно-

отстоящими датами времени. Например, списочная численность работников предприятия за январь (чел.): на 1/1 — 842, на 5/1 — 838, на 12/1 — 843, на 26/1 — 845.

Расчет среднего уровня в таких рядах динамики производится по методу средней арифметической взвешенной:

$$\bar{y} = \frac{\sum yt}{\sum t}$$

где y — уровни, сохраняющиеся без изменения в течение промежутков (интервалов) времени t .

Пример. За январь 2003 г. произошли следующие изменения в списочном составе работников предприятий, чел.:

Состояло по списку на 1/1 2003 г.	842
Выбыло с 5/1	4
Зачислено с 12/1	5
Зачислено с 26/1	2

Необходимо определить среднедневную списочную численность работников предприятия за январь.

Для расчета средней численности работников определим продолжительность t каждого календарного периода с постоянной численностью работающих и общее число человеко-дней.

Т а б л и ц а

Календарные периоды января	Число работников y	Длина периода, дней t	Число человеко-дней ty
1-4	842	4	3 368
5-11	838	7	5 866
12-25	843	14	11 802
26-31	845	6	5 070
Итого	-	31	26 106

Отсюда среднедневная списочная численность работников в январе:

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{26106}{31} = 842 \text{ чел.}$$

Задачи для самостоятельной работы

Задача 1. Списочная численность рабочих составляла на начало месяца, чел.:

1/I 2002г.	400	1/VII 2002г.	496
1/II 2002г.	420	1/VIII 2002г.	450
1/III 2002г.	405	1/IX 2002г.	412
1/IV 2002г.	436	1/X 2002г.	318
1/V 2002г.	450	1/XI 2002г.	231
1/VI 2002г.	472	1/XII 2002г.	235
		1/I 2003г.	210

Определите: 1) вид ряда динамики; 2) среднемесячные уровни ряда в первом и втором полугодиях; 3) изменение списочной численности рабочих во втором полугодии по сравнению с первым.

Задача 2. Имеются следующие данные об остатках материалов в первом полугодии 2002г. по месяцам года, тыс. руб.:

	I	II	III	IV	V	VI	VII	VIII
Остатки на начало периода	8	7	6	5	4	4	4	3
	20	26	18	16	13	11	90	

Определите: 1) вид ряда динамики; 2) среднемесячные остатки материалов за I и II кварталы года; 3) изменение остатка материалов во II квартале по сравнению с I.

Задача 3. Остатки вкладов населения в сберегательных кассах города характеризуются следующими данными на дату, тыс. руб.:

1	1	1	1	1	1	1
/I	/II	/III	/IV	/V	/VI	/VII
3	3	3	3	3	3	3
00,2	12,4	23,3	14,8	16,5	19,3	24,6

Определите: 1) вид ряда динамики; 2) среднемесячные уровни остатка вкладов населения на I и II кварталы; 3) изменение остатка вкладов населения во II квартале по сравнению с I.

Задача 4. Получены следующие данные о товарных запасах торговой организации по товарным группам в 2002г. на дату (в сопоставимых ценах; тыс. руб.):

Товары	1	1	1	1	1/1
	/I	/IV	/VII	/X	2003г.
Продовольственные	106	135	156	190	220
Непродовольственные	610	650	520	670	540

Определите: 1) вид рядов динамики; 2) среднеквартальные запасы по продовольственным и непродовольственным товарам и по обеим группам в целом; 3) относительные величины структуры товарных запасов: а) на начало года, б) на конец года.

Задача 5. Имеются следующие данные об изменениях в списочном составе работников фабрики за I и II кварталы 2002г., чел.:

Состояло по списку на 1/I	634
Зачислено с 12/I	10
Выбыло с 26/I	12
Зачислено с 23/III	6
Зачислено с 1/IV	18
Выбыло с 1/VI	8

Определите:

1) среднесписочную численность работников фабрики: а) за I квартал, б) за II квартал;

2) абсолютный прирост численности работников фабрики во II квартале по сравнению с I.

Задача 6. За январь и февраль 2002 г. произошли следующие изменения в списочном составе работников цеха завода, чел.:

Состояло по списку на 1/1 2002г.	126
Выбыло с 5/1	3
Выбыло с 12/1	4
Выбыло с 26/1	5
Зачислено с 9/II	8
Выбыло с 16/II	3
Зачислено с 25/II	6

Определите:

- 1) среднесписочную численность работников цеха завода: а) за январь, б) за февраль;
- 2) абсолютный прирост численности работников цеха завода в феврале по сравнению с январем.

Расчета в ряду динамики темпов роста

Одним из важнейших направлений анализа рядов динамики является изучение особенностей развития явления за отдельные периоды времени.

В статистике для выявления специфики развития изучаемых явлений за отдельные периоды времени определяют относительные и абсолютные показатели изменения ряда динамики: темпы роста, абсолютные и относительные приросты, абсолютное значение одного процента прироста.

Пример. Выпуск продукции предприятием характеризуется следующими данными (в сопоставимых ценах; млн. руб.):

1	1	1	2	2	2
997	998	999	000	001	002
1	1	1	1	1	1
2,3	3,4	4,8	6,4	7,8	9,9

Требуется произвести анализ динамики выпуска продукции предприятием за 1999-2004 годы.

Важнейшим показателем изменения абсолютных уровней ряда динамики по отдельным периодам времени является темп роста K . Величина этого показателя определяется из сопостав-

ления уровня изучаемого периода y_i с уровнем, который принимается за базу сравнения. Выражаются темпы роста в процентах или в виде коэффициентов. Выбор базы для расчета в ряду динамики темпов роста определяется задачей исследования.

Если задачей изучения ряда динамики является, например, контроль хода выполнения плана, то сравнение уровней ряда динамики производится по отношению к году, который принимался за базу сравнения.

$$K_b = y_i / y_0$$

Определим базисные темпы роста:

$$2000 - 13,4 / 12,3 = 1,089$$

$$2001 - 14,8 / 12,3 = 1,203$$

$$2002 - 16,4 / 12,3 = 1,333$$

$$2003 - 17,8 / 12,3 = 1,447$$

$$2004 - 19,9 / 12,3 = 1,618$$

Из полученных базисных темпов роста следует, что по годам происходило систематическое возрастание темпов роста выпуска продукции, %:

$$108,9 < 120,3 < 133,3 < 144,7 < 161,8.$$

Если же при изучении данного ряда динамики ставится задача определить изменения выпуска продукции в каждом последующем периоде по сравнению с предыдущим, то определяются цепные (погодные) темпы роста K_c , когда за базу сравнения отдельных уровней ряда y_i , - каждый раз принимается предыдущий уровень y_{i-1} :

$$K_c = y_i / y_{i-1}$$

Определим цепные темпы роста:

$$2000 - 13,4 / 12,3 = 1,089;$$

$$2001 - 14,8 / 12,3 = 1,104;$$

$$2002 - 16,4 / 14,8 = 1,108;$$

$$2003 - 17,8 / 16,4 = 1,085;$$

$$2004 - 19,9 / 17,8 = 1,118;$$

Из полученных цепных темпов роста видно, что в 2000, 2001 и 2002 гг. происходил рост выпуска продукции из года в год, %:

$$108,9 < 110,4 < 110,8.$$

В 2003 г. имело место некоторое замедление роста выпуска продукции, %:

$$110,8 > 108,5 < 111,8.$$

Эта специфика развития данного явления в 2003г. из определенных выше показателей базисных темпов роста не была видна.

Для выражения изменений уровней ряда динамики в абсолютных величинах исчисляют статистический показатель абсолютного прироста Δy . Величина этого показателя определяется как разность между уровнем изучаемого периода y_i и уровнем, принимаемым за базу сравнения. При определении накопленных (базисных) абсолютных приростов Δy_b за базу сравнения принимается постоянный уровень.

$$\Delta y_b = y_i - y_0$$

$$2000 - 13,4 - 12,3 = 1,1$$

$$2001 - 14,8 - 12,3 = 2,5$$

$$2002 - 16,4 - 12,3 = 4,1$$

$$2003 - 17,8 - 12,3 = 5,5$$

$$2004 - 19,9 - 12,3 = 7,6$$

Из полученных значений накопленных абсолютных приростов видно, что по годам происходило систематическое возрастание абсолютных приростов выпуска продукции, млн. руб.:

$$1,1 < 2,5 < 4,1 < 5,5 < 7,6.$$

При определении цепных абсолютных приростов Δy_c базой сравнения каждый раз выступает уровень предыдущего периода y_{i-1} , и расчет абсолютных приростов (млн. руб.) производится по формуле

$$\Delta y_c = y_i - y_{i-1}$$

Следовательно

$$2000 - 13,4 - 12,3 = 1,1$$

$$2001 - 14,8 - 13,4 = 1,4$$

$$2002 - 16,4 - 14,8 = 1,6$$

$$2003 - 17,8 - 16,4 = 1,4$$

$$2004 - 19,9 - 17,8 = 2,1$$

Из полученных годовых (цепных) абсолютных приростов видно сокращение абсолютного прироста выпуска продукции на данном предприятии в 2003 г., млн. руб.:

$$1,1 < 1,4 < 1,6 > 1,4 < 2,1$$

Для выражения изменения величины абсолютного прироста уровней ряда динамики в относительных величинах определяется статистический показатель — темп прироста ΔK . Величина темпа прироста ΔK определяется из сравнения в отношении абсолютного прироста Δy с уровнем, принимаемым при анализе за базу сравнения. При исчислении базисных темпов прироста ΔK_6 в качестве базы сравнения берется постоянный уровень y_0 .

$$\Delta K_6 = \Delta y_6 / y_0$$

Определим базисные темпы прироста, %:

$$1998 - (1,1 / 12,3) \cdot 100 = 8,9$$

$$1999 - (2,5 / 12,3) \cdot 100 = 20,3$$

$$2000 - (4,1 / 12,3) \cdot 100 = 33,3$$

$$2001 - (5,5 / 12,3) \cdot 100 = 44,7$$

$$2002 - (7,6 / 12,3) \cdot 100 = 61,8$$

Базисные темпы прироста, подобно базисным темпам роста, показывают систематическое возрастание по годам пятилетки выпуска продукции данным предприятием. Но в отличие от темпов роста, показывающих, во сколько раз (или процентов) происходит рост выпуска продукции, базисные темпы прироста показывают, насколько произошло приращение (в относительных величинах) абсолютных уровней ряда динамики.

При определении цепных темпов прироста $\Delta K_{ц}$ в качестве базы сравнения выступает уровень предшествующего периода y_{i-1} , и расчет осуществляется по формуле

$$\Delta K_{ц} = \Delta y_{ц} / y_{i-1}$$

Определяем эти показатели, %:

$$2000 - (1,1 / 12,3) \cdot 100 = 8,9$$

$$2001 - (1,4 / 13,4) \cdot 100 = 10,4$$

$$2002 - (1,6 / 14,8) \cdot 100 = 10,8$$

$$2003 - (1,4 / 16,4) \cdot 100 = 8,5$$

$$2004 - (2,1 / 17,8) \cdot 100 = 11,8$$

Если же при определении темпов прироста ΔK предварительно были исчислены темпы роста K , то расчет темпов прироста

производят так: $\Delta K = K-1$, или $\Delta K = K - 100\%$ (если темп роста выражен в процентах).

Показатель абсолютного значения одного процента прироста ($A\%$) определяется путем отношения (а каждом периоде) абсолютного прироста $\Delta y_{ц}$ к темпу прироста $\Delta K_{ц}$. Расчет этого показателя имеет экономический смысл только на цепной основе:

$$A\% = \Delta y_{ц} / \Delta K_{ц} \%$$

Определяем эти показатели, тыс. руб.:

$$2000 - 1,1 / 8,9 = 123$$

$$2001 - 1,4 / 10,4 = 134$$

$$2002 - 1,6 / 10,8 = 148$$

$$2003 - 1,4 / 8,5 = 164$$

$$2004 - 2,1 / 11,8 = 178$$

При анализе годовых уровней ряда динамики расчет абсолютного значения одного процента прироста можно произвести по схеме:

$$A\% = 0,01 y_{i-1}$$

Например, для 2003г. $A\% = 0,01 y (2002г.) = 0,01 \times 16,4 = 0,164$ млн. руб.; для 2004г. $A\% = 0,01 y (2003 г.) = 0,01 \cdot 17,8 = 0,178$ млн. руб.

Для суждения о среднем изменении абсолютных приростов исчисляется показатель среднего абсолютного прироста $\overline{\Delta y}$. Способы расчета среднего абсолютного прироста зависят от характера исходных данных.

Обычно определение среднего абсолютного прироста производят по цепным абсолютным приростам $\Delta y_{ц}$ по формуле

$$\overline{\Delta y} = \frac{\sum \Delta y_{ц}}{n}$$

Так, для данного примера этот расчет будет следующим:

$$\overline{\Delta y} = \frac{1,1+1,4+1,6+1,4+2,1}{5} = \frac{7,6}{5} = 1,52 \text{ млн. руб.}$$

Средний абсолютный прирост можно исчислить и непосредственно по абсолютным уровням ряда динамики y по формуле

$$\overline{\Delta y} = \frac{y_n - y_0}{m - 1}$$

где t — число учетных единиц времени в ряду динамики. Для данного примера имеет:

$$\overline{\Delta y} = \frac{19,9 - 12,3}{6 - 1} = \frac{7,6}{5} = 1,52 \text{ млн. руб.}$$

И наконец, в тех случаях, когда в качестве исходных материалов даны накопленные (базисные) абсолютные приросты Δy_b , то расчет среднего абсолютного прироста производится по формуле

$$\overline{\Delta y} = \frac{\Delta y_b}{m - 1}$$

Для рассматриваемой задачи накопленный (базисный) абсолютный прирост равен 7,6 млн. руб. По этому значению расчет среднегодового абсолютного прироста составляет:

$$\overline{\Delta y} = \frac{7,6}{6 - 1} = 1,52 \text{ млн. руб.}$$

Задачи для самостоятельного решения

Задача 1. Получены следующие данные о производстве продукции промышленным предприятием за 1997-2002 гг. (в сопоставимых ценах; млн.руб.):

	1	1	1	2	2	2
99	99	99	00	00	00	
7	8	9	0	1	2	
	2	2	2	2	2	3
3,3	4,9	6,6	7,6	9,0	2,3	

Для анализа ряда динамики: определите: 1) показатели, характеризующие рост производства продукции (по годам и по отношению к базисному 1997г.): а) темпы роста, б) абсолютные приросты, в) темпы прироста; 2) определите абсолютное значение одного процента прироста (для каждого года); 3) представьте полученные данные в табличной форме; 4) определите средний абсолютный прирост за пятилетие в целом.

Задача 2. Имеются, следующие данные по заводу о выпуске продукции за 1997 - 2002 гг. (в сопоставимых ценах; млн. руб.):

	1	1	1	2	2	2
	997	998	999	000	001	002
	4	5	5	5	6	6
	6,8	0,9	5,3	8,7	2,4	6,2

Для анализа ряда динамики: 1) определите показатели, характеризующие рост выпуска продукции (по годам и по отношению к базисному 1997г.): а) темпы роста, б) абсолютные приросты, в) темпы прироста; 2) определите для каждого года абсолютное значение одного процента прироста; 3) представьте полученные данные в табличной форме; 4) определите за пятилетие в целом средний абсолютный прирост.

Задача 3. Имеются следующие данные о выпуске шерстяных тканей производственным объединением, млн. пог. м:

	1	1	1	1	1	2	2	2
	995	996	997	998	999	000	001	002
	6	6	7	7	8	9	1	1
	5,8	9,3	5,6	9,0	4,2	1,6	01,5	08,7

На основе приведенных данных: 1) установите начальный, конечный и базисный уровни ряда динамики для определения за годы десятой пятилетки: а) среднего уровня ряда, б) цепных и базисных темпов роста, в) среднегодового абсолютного прироста; 2) и числите указанные в п. 1 показатели анализа ряда динамики.

Задача 4. Имеются следующие данные о перевозках грузов автомобильным транспортом области, млн. т:

	1	1	1	1	1	2	2	2
995	996	997	998	999	000	001	002	
	1	1	1	2	2	2	2	2
92	95	98	02	19	30	58	84	

На основе приведенных данных: 1) установите начальный, [конечный и базисный уровни ряда динамики для определения за годы десятой пятилетки: а) среднего уровня ряда, б) цепных и базисных темпов роста, в) среднегодового абсолютного прироста; 2) и считайте указанные в п. 1 показатели анализа ряда динамики.

ОПРЕДЕЛЕНИЕ В РЯДАХ ДИНАМИКИ СРЕДНЕГО ТЕМПА РОСТА И ПРИРОСТА

При изучении рядов динамики возникает необходимость определения среднего темпа роста явления за отдельные периоды его развития. Для рядов динамики годовых уровней чаще всего производят расчет среднегодовых темпов роста за отдельные этапы экономического развития страны (например, пятилетки). В рядах внутригодовой динамики исчисляют среднемесячные и среднеквартальные темпы роста. Для определения среднего темпа роста обычно используют метод средней геометрической. Применительно к рядам динамики формула средней геометрической в общем I виде записывается так:

$$\bar{K} = \sqrt[p]{PK}$$

где PK — произведение цепных темпов роста (в коэффициентах);

p — число K .

В зависимости от характера исходных данных формула средней геометрической видоизменяется.

Рассмотрим основные случаи определения средних темпов динамики:

1. Исчисление среднего темпа по цепным темпам роста. Если в качестве исходных данных выступают цепные

темпы роста, то для расчета среднего темпа динамики используется формула

$$\bar{K} = \sqrt[n]{K_1 * K_2 * \dots * K_n}$$

где K — цепные темпы роста (в коэффициентах);
 n — число темпов.

Пример. Выпуск продукции предприятием характеризуется следующими данными по кварталам 1980 г. в коэффициентах к предыдущему кварталу:

	I	I	II	I
		I	I	V
	1	1	1	1
	,109	,24	,225	,197
	5	0	8	4

Подставив в формулу $\bar{K} = \sqrt[4]{K_1 * K_2 * K_3 * K_4}$ приведенные выше данные, получаем:

$$\bar{K} = \sqrt[4]{1,1095 * 1,24 * 1,2258 * 1,1974} = 2,167$$

2. Расчет среднего темпа динамики по базисным темпам роста или прироста.

Если в качестве исходных данных выступают базисные темпы роста или прироста, то на основе зависимости между цепными и базисными коэффициентами динамики расчет среднего темпа производят по формуле:

$$\bar{K} = \sqrt[m-1]{K_6}$$

где K_6 — базисный темп роста (в коэффициентах);
 m — число учетных единиц времени в изучаемом периоде.

Пример. На 1998-2002 гг. предусматривается увеличение производства продукции (в сопоставимых ценах) в 1,3 раза, в том числе товаров входящих в специализацию предприятия - на 22%.

Требуется определить, какие должны быть у предприятия среднегодовые темпы роста производства товарной продукции.

Заданный коэффициент (1,3) представляет темп развития производства продукции в 2002г. по сравнению с базой— 1997 годом, т. е. базисный темп роста:

$$K_{\frac{2002}{1997}} = 1,3 = K_6$$

Поэтому для определения среднегодового темпа применяется формула

$$\bar{K} = \sqrt[m-1]{K_6}$$

При $t = 6$ (число годовых периодов в изучаемом отрезке времени):

$$\bar{K} = \sqrt[6-1]{1,3} \approx 1,053, \text{ или } 105,3\%$$

Задание по росту выпуска товаров входящих в специализацию предприятия выражено базисным темпом прироста. Для определения среднегодового темпа динамики заданный темп прироста ($\Delta K_6 = 22\%$) необходимо из процентной формы выражения перевести в коэффициент ($\Delta K_6 = 0,22$) и далее в темп роста:

$$K_6 = \Delta K_6 + 1 = 0,22 + 1 = 1,22$$

Определим значение среднегодового темпа роста:

$$\bar{K} = \sqrt[6-1]{1,22} = 1,04, \text{ или } 104\%.$$

3. Расчет среднего темпа роста по абсолютным уровням ряда динамики.

Из зависимости $\bar{K} = \sqrt[n]{K_1 * K_2 * \dots * K_n} = \sqrt[m-1]{K_6}$ где $K_6 = y_n : y_0$ расчет среднего темпа роста производится по исходным абсолютным уровням ряда динамики по формуле:

$$\bar{K} = \sqrt[m-1]{y_n y_0}$$

где y_n — конечный уровень ряда динамики;
 y_0 — базисный уровень ряда динамики;

t — число учетных единиц времени в изучаемом периоде.

Задачи для самостоятельного решения

Задача 1. Темпы роста объема продукции добывающей промышленности в экономическом районе характеризуются следующими данными (в процентах к предыдущему году):

	1	1	1	1	1	1	1	1	1	1	1
971	972	973	974	975	976	977	978	979	980		
	1	1	1	1	1	1	1	1	1	1	1
04	06	07	05	06	05	05	05	04	06		

Определите среднегодовые темпы роста объема продукции добывающей промышленности: 1) в девятой пятилетке; 2) в десятой пятилетке; 3) в целом за период с 1970 по 1980 г.

ОПРЕДЕЛЕНИЕ В РЯДАХ ДИНАМИКИ ОБЩЕЙ ТЕНДЕНЦИИ РАЗВИТИЯ

Важной задачей статистики при анализе рядов динамики является определение общей тенденции развития. На развитие явления во времени могут оказывать влияние различные по своему характеру и силе воздействия факторы. Одни из них оказывают более или менее постоянное воздействие и формируют в рядах динамики определенную тенденцию развития. Воздействие же других факторов может быть кратковременным.

При изучении в рядах динамики общей тенденции развития явления применяются различные приемы и методы. Одним из наиболее элементарных способов изучения общей тенденции в ряду динамики является *укрупнение интервалов*. Этот способ основан на укрупнении периодов, к которым относятся уровни ряда динамики.

Пример. Имеются следующие данные о выпуске продукции группой предприятий по месяцам 1980 г., млн. руб.:

январь	23,2	июль	28,4
февраль	19,1	август	24,1

март	22,3	сентябрь	26,3
апрель	25,1	октябрь	29,1
май	24,5	ноябрь	30,3
июнь	27,3	декабрь	26,5

Для выявления общей тенденции роста выпуска продукции произведем укрупнение интервалов. Для этой цели исходные (месячные) данные о выработке продукции объединяем в квартальные и получаем показатели выпуска продукции группой предприятий по кварталам 1980 г., млн. руб.:

	I	II	III	IV
	6	7	7	8
	4,5	6,9	8,8	5,9

В результате укрупнения интервалов общая тенденция роста выпуска продукции данной группой заводов выступает отчетливо:

$$64,6 < 76,9 < 78,8 < 85,9.$$

Выявление общей тенденции ряда динамики можно произвести путем сглаживания ряда динамики с помощью подвижной (скользящей) средней. Сущность этого приема состоит в том, что по исходным уровням ряда (эмпирическим данным) определяют расчетные (теоретические) уровни. При этом посредством осреднения эмпирических данных индивидуальные колебания погашаются, и общая тенденция развития явления выражается в виде некоторой плавной линии (теоретические уровни).

Пример. По городу имеются следующие данные в торговле продуктами потребительской кооперации (среднедневная выручка; тыс. руб.):

К	1	1	1	1
варта	97	97	97	97
л	6	7	8	9
I	1	2	4	4

	75	47	20	26
II	2	2	4	4
	63	98	41	49
III	3	3	4	4
	26	66	53	82
IV	2	3	3	4
	97	41	99	60

Специфический для данного явления характер колебаний уровней ряда можно видеть из графического представления исходных (эмпирических) данных.

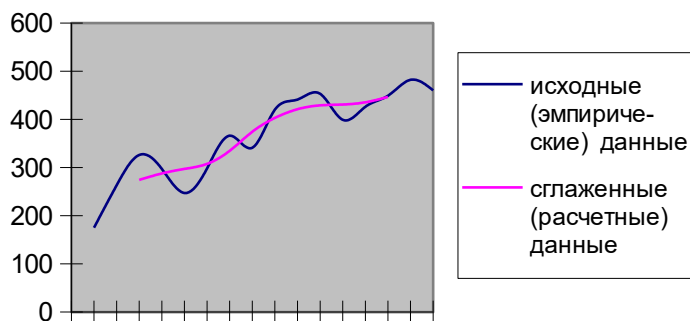


График. Комиссионная продажа продуктов потребительской кооперации города по кварталам в 2000-2003 гг.

Увеличение уровней объема реализации сельскохозяйственных продуктов во II и III кварталах и относительное их снижение в IV квартале характерно для каждого из представленных годовых периодов. Для выражения общей тенденции развития явления методом сглаживания рядов динамики необходимо прежде всего определить по эмпирическим данным подвижные (скользящие) средние. Основное условие применения этого метода состоит в вычислении звеньев подвижной (скользящей) средней из такого числа уровней ряда, которое соответствует длительности наблюдаемых в ряду динамики циклов. Для ряда внутригодовой динамики с сезонными циклами развития явления по одноименным кварталам года применяют четырехчленные скользящие средние. Расчет состоит в определении средних величин из четырех уровней ряда с отбрасыванием при вычислении каждой новой средней одного уровня ряда слева и присоединением одного уровня справа:

первая средняя $\bar{y}_1 = \frac{y_1 + y_2 + y_3 + y_4}{4}$

вторая средняя $\bar{y}_2 = \frac{y_2 + y_3 + y_4 + y_5}{4}$

третья средняя $\bar{y}_3 = \frac{y_3 + y_4 + y_5 + y_6}{4} \text{ и т. д.}$

Применительно к исходным данным получаем тринадцать средних::

Первая $\bar{y}_1 = \frac{175 + 263 + 326 + 297}{4} = 265,25$

Вторая $\bar{y}_2 = \frac{263 + 326 + 297 + 247}{4} = 283,25$

Третья $\bar{y}_3 = \frac{326 + 297 + 247 + 298}{4} = 292$

.....

Тринадцатая $\bar{y}_{13} = \frac{426 + 449 + 482 + 460}{4} = 454,25$

Особенность сглаживания по четному числу уровней состоит в том, что каждая из исчисленных четырехчленных средних относится к соответствующим промежуткам между двумя смежными кварталами. Так, первая средняя ($\bar{y}_1 = 265,25$) относится к промежутку между II и III кварталами 2000 г., вторая средняя ($\bar{y}_2 = 283,25$)—к промежутку между III и IV кварталами 2000г. и т. д.

Для получения значений сглаженных уровней соответствующих кварталов необходимо провести центрирование расчетных средних.

Так, для определения сглаженного среднего уровня III квартала 2000г. произведем центрирование первой средней

\bar{y}_1 и второй средней \bar{y}_2 :

$$\bar{y}_{сIIIкв} = \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{265,25 + 283,26}{2} = 274,25$$

Для определения сглаженного среднего уровня IV квартала 1976 г. произведем центрирование второй средней \bar{y}_2 и третьей средней \bar{y}_3

$$\bar{y}_{сI/кв.} = \frac{\bar{y}_2 + \bar{y}_3}{2} = \frac{283,25 + 292}{2} = 287,6 \quad \text{и т.д.}$$

Ход расчета необходимых данных для получения средних (теоретических) уровней представим в таблице сглаживания ряда динамики по четырехчленной переменной (скользящей) средней:

Таблица

Период	Исходные уровни (y_i)	Средняя из суммы четырех уровней ряда	Сглаженный средний уровень (с центрированием \bar{y}_c)
1	2	3	4
2 000	I	175	-
	II	263	1061: 4 = 265,25
	III	326	1133: 4 = 283,25
	IV	297	1168: 4 = 292
2 001	I	247	1208: 4 = 302
	II	298	1252: 4 = 313
	III	366	1425: 4 = 356,25

	IV	341	$1568:4 = 392$	374,1
2 002	I	420	$1655:4 = 413,75$	402,9
	II	441	$1713:4 = 428,25$	421,0
	III	453	$1719:4 = 429,75$	429
	IV	399	$1727:4 = 431,75$	430,75
2 003	I	426	$1756:4 = 439,0$	435,37
	II	449	$1817:4 = 454,25$	446,62
	III	482		-
	IV	460		-

Из нанесенных на график пунктирной линией данных видна довольно отчетливая тенденция роста объема торговли.

Спецификой способа сглаживания рядов динамики является то, что полученные средние не дают теоретических рядов, в основе которых лежала бы математически выраженная закономерность.

Более совершенным приемом изучения общей тенденции в рядах динамики является аналитическое выравнивание. При изучении общей тенденции методом аналитического выравнивания исходят из того, что изменения уровней ряда динамики могут быть с той или иной степенью приближения выражены определенными математическими функциями. На основе теоретического анализа выявляется характер развития явления во времени и на этой основе выбирается то или иное математическое выражение типа изменения явления: по прямой, параболе второго порядка, показательной (логарифмической) кривой и т. д.

Рассмотрим применение метода аналитического выравнивания по прямой для выражения общей тенденции на следующем примере.

Пример. Имеются данные о выпуске продукции предприятиями легкой промышленности района за 1995—2003 гг. (в сопоставимых ценах; млн. руб.):

1995	221	2000	320
1996	235	2001	360
1997	272	2002	371
1998	285	2003	395
1999	304		

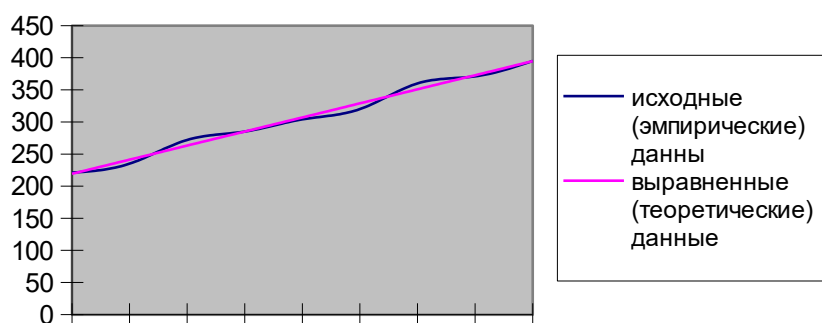


График.
Выпуск продукции предприятиями района в 1995-2003 гг.

Общее представление о характере тенденции изменения изучаемого явления можно получить из графического изображения ряда динамики. Из графика видно, что для изучаемого периода времени (1995-2003 гг.) уравнение прямой достаточно полно отображает общую тенденцию развития явления.

Для выравнивания ряда динамики по прямой используют уравнение

$$y_t = a_0 + a_1 t$$

Способ наименьших квадратов дает систему двух нормальных уравнений для нахождения параметров a_0 и a_1 :

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

где y — исходные (эмпирические) уровни ряда динамики;

n — число членов ряда;

t — время.

Решение системы уравнений позволяет получить выражения для параметров a_0 и a_1 :

$$a_0 = (\sum t^2 \sum y - \sum t \sum ty) / (n \sum t^2 - \sum t \cdot \sum t)$$

$$a_1 = (n \sum ty - \sum t \sum y) / (n \sum t^2 - \sum t \sum t)$$

В рядах динамики техника расчета параметров уравнения упрощается. Для этой цели показателям времени t придают такие значения, чтобы их сумма была равна нулю, т. е. $\sum t=0$.

Применительно к данному примеру, в котором число исходных (эмпирических) уровней ряда — нечетное ($n=9$), это выполнимо

1995	1996	1997	1998	1999	2000	2001	2002	2003
-4	-3	-2	-1	0	+1	+2	+3	+4

При условии что $\sum t=0$, исходные нормальные уравнения принимают вид

$$\begin{cases} a_0 n = \sum y_i \\ a_1 \sum t = \sum y_i t \end{cases}$$

Откуда

$$a_0 = \frac{\sum y_i}{n} = \bar{y}$$

$$a_1 = \frac{\sum y_i t}{\sum t^2}$$

Произведем расчет необходимых значений в таблице.

Год	Эмпирические уровни ряда (y_i)	Условные обозначения времени (t)	t^2	$y_i t$	y_t
1995	221	-4	16	-884	219,32
1996	235	-3	9	-705	241,24
1997	272	-2	4	-544	263,16
1998	285	-1	1	-285	285,08
1999	304	0	0	0	307,00
2000	320	+1	1	320	328,92
2001	360	+2	4	720	350,84
2002	371	+3	9	1113	372,76

				113	76
2003	395	+4	16	1 580	394, 68
Итог о	2 763	0	60	1 315	2 763

По итоговым данным определяем параметры уравнения:

$$a_0 = 2\,763 / 9 = 307;$$

$$a_1 = 1\,315 / 60 = 21,9167.$$

Значение Σt^2 можно вычислить и другим путем. Для случая нечетного числа уровней ряда динамики используется формула

$$\Sigma t^2 = ((n-1) n (n+1)) / 12 = (8 \cdot 9 \cdot 10) / 12 = 60$$

В результате получаем следующее уравнение общей тенденции ряда динамики:

$$y_t = 307 + 21,92t$$

Заметим, что при упрощенном способе расчета ($\Sigma t = 0$) параметр $a_0 = 307$ характеризует величину центрального выравненного уровня ряда, который был принят за $t = 0$. В рассматриваемом примере это уровень 1999г. Подставляя в уравнение $y_t = 307 + 21,92t$ принятые обозначения t , вычислим выравненные (теоретические) уровни ряда динамики:

$$1995 \quad y_t = 307 + 21,92 (-4) \approx 219,32$$

$$1996 \quad y_t = 307 + 21,92 (-3) \approx 241,24 \text{ и т.д.}$$

Для проверки расчета значений y_t используется формула

$$\Sigma y_i = \Sigma y_t$$

В нашем примере $\Sigma y_i = 2\,763 = \Sigma y_t$; следовательно, значения y_t определены верно.

Полученные величины теоретических уровней ряда y_i нанесем пунктирной линией на график с эмпирическими данными (см. рис. 8.3). Эта линия и есть графический образ общей тенденции выпуска продукции предприятиями легкой промышленности района в 1995—2003 гг.

Задачи для самостоятельного решения

Задача 1. Имеются следующие данные о реализации продукции в магазинах городов по месяцам 2000 — 2003гг. (тыс. т):

Месяц	2000г.	2001г.	2002г.	2003г.
-------	--------	--------	--------	--------

Январь	5,3	5,3	8,3	10,4
Февраль	5,2	5,0	7,6	10,2
Март	8,0	8,8	11,0	11,8
Апрель	8,2	9,8		14,1
Май	9,8	15,4		17,8
Июнь	14,9	18,3	24,8	27,6
Июль	11,8	17,1	23,8	25,0
Август	10,3	15,4	19,4	19,8
Сентябрь	8,0	12,9	15,7	17,4
Октябрь	6,5	9,5	11,8	12,7
Ноябрь	5,4	9,0	10,2	11,0
Декабрь	5,6	7,5	10,1	8,6

Для изучения общей тенденции реализации данной продукции: 1) произведите преобразование исходных данных путем укрупнения периодов времени: а) в квартальные уровни, б) в годовые уровни; 2) нанесите на линейный график полученные квартальные уровни; 3) произведите сглаживание квартальных уровней с применением четырехчленной скользящей средней; 4) нанесите полученные при сглаживании данные на график с квартальными уровнями; 5) сделайте выводы о характере общей тенденции изучаемого явления.

Задача 2. По городу имеются данные о реализации яиц в магазинах по месяцам 2000-2003 (млн. шт.):

Месяц	2000г.	2001г.	2002г.	2003г.
Январь	12,7	16,3	23,6	30,8
Февраль	11,5	17,4	14,0	24,1
Март	12,0	18,4	17,4	21,2
Апрель	45,6	78,9	94,8	73,1
Май	40,4	67,3	76,7	60,9
Июнь	60,0	66,6	51,4	77,6
Июль	42,0	42,7	21,2	43,6
Август	23,4	39,9	14,2	40,7
Сентябрь	14,1	28,9	13,5	70,0
Октябрь	14,6	25,2	23,2	40,7
Ноябрь	16,3	27,9	30,4	32,7
Декабрь	18,0	30,5	21,9	33,0

Для изучения общей тенденции реализации данной продукции: 1) произведите преобразование исходных данных путем укрупнения периодов времени: а) в квартальные уровни, б) в годовые уровни; 2) нанесите на линейный график полученные квартальные уровни; 3) произведите сглаживание квартальных уровней с применением четырехчленной скользящей средней; 4) нанесите полученные при сглаживании данные на график с квартальными уровнями; 5) сделайте выводы о характере общей тенденции изучаемого явления.

Задача 3. Реализация кондитерских изделий в магазинах продовольственного торгового города по месяцам 2000 - 2003гг. характеризуется следующими данными (т):

Месяц	2000г.	2001г.	2002г.	2003г.
Январь	32,1	31,8	37,3	36,9
Февраль	30,8	27,9	32,2	33,4
Март	33,0	38,9	42,0	36,9
Апрель	34,4	39,5	40,9	46,6
Май	29,1	36,2	36,8	37,5
Июнь	30,9	41,4	40,3	39,2
Июль	31,4	39,4	35,3	38,3
Август	29,3	34,3	34,0	36,3
Сентябрь	32,5	34,7	33,7	36,8
Октябрь	32,5	34,6	34,9	35,4
Ноябрь	34,7	36,3	35,3	39,2
Декабрь	38,2	41,6	42,7	48,1

Для изучения общей тенденции реализации данной продукции: 1) произведите преобразование исходных данных путем укрупнения периодов времени: а) в квартальные уровни, б) в годовые уровни; 2) нанесите на линейный график полученные квартальные уровни; 3) произведите сглаживание квартальных уровней с применением четырехчленной скользящей средней; 4) нанесите полученные при сглаживании данные на график с квартальными уровнями; 5) сделайте выводы о характере общей тенденции изучаемого явления.

Задача 4. Имеются следующие данные о реализации сахара в продовольственных магазинах города по месяцам 2000 - 2003гг. (т) :

Месяц	2000г.	2001г.	2002г.	2003г.
Январь	78,9	108,6	129,1	150,7
Февраль	78,1	107,9	128,6	149,6
Март	86,0	106,8	130,7	153,6
Апрель	97,5	132,1	152,8	174,4
Май	83,3	113,0	139,8	153,7
Июнь	86,0	111,8	147,4	158,6
Июль	90,6	124,4	163,8	199,2
Август	86,1	114,1	146,3	164,3
Сентябрь	81,3	108,4	137,8	135,5
Октябрь	105,1	124,0	152,2	159,3
Ноябрь	97,2	118,4	143,2	155,5
Декабрь	102,1	136,3	156,5	158,2

Для изучения общей тенденции реализации данной продукции: 1) произведите преобразование исходных данных путем укрупнения периодов времени: а) в квартальные уровни, б) в годовые уровни; 2) нанесите на линейный график полученные- квартальные уровни; 3) произведите сглаживание квартальных уровней с применением четырех-членной скользящей средней; 4) нанесите полученные при сглаживании данные на график с квартальными уровнями; 5) сделайте выводы о характере общей тенденции изучаемого явления.

6 ИНДЕКСЫ

Индексы — обобщающие показатели сравнения во времени и в пространстве не только однотипных (одноименных) явлений, но и совокупностей, состоящих из несоизмеримых элементов.

Динамика одноименных явлении изучается с помощью индивидуальных индексов (I), которые представляют собой известные относительные величины сравнения, динамики или выполнения плана (обязательств):

$$i_q = q_1 / q_0; \quad i_p = p_1 / p_0; \quad i_{pq} = p_1 q_1 / p_0 q_0;$$

где подстрочное обозначение «0» соответствует уровню базисного периода (с которым сравнивают) или момента времени. «1» уровню отчетного (сравниваемого) периода или момента времени.

Изменения совокупностей, состоящих из элементов, непосредственно не сопоставимых (например, различных видов продукции), изучают с помощью групповых, или общих, индексов (I). Последние по методам построения подразделяются на агрегатные индексы и средневзвешенные из индивидуальных индексов.

Формулы агрегатных индексов:

1) физического объема:

$$i_q = \Sigma q_1 p_0 / \Sigma q_0 p_0$$

где q — индексируемая величина; p₀ — соизмеритель, или вес, который фиксируется на уровне одного и того же периода. В случае индексов объемных показателей весами являются качественные показатели (цена, себестоимость и др.), зафиксированные на уровне базисного периода.

Разница между числителем и знаменателем индекса

$$\Delta^q p q = \Sigma q_1 p_0 - \Sigma q_0 p_0$$

в данном случае означает абсолютное изменение товарооборота (прирост или снижение) за счет изменения физического объема;

2) цен и других качественных показателей:

$$I_p = \Sigma p_1 q_1 / \Sigma p_0 q_1 \quad (\text{формула Пааше}),$$

$$I_p = \Sigma p_1 q_0 / \Sigma p_0 q_0 \quad (\text{формула Ласпейреса}),$$

где q — объемы (количества) являются весами, взятыми на одинаковом уровне (отчетном или базисном).

Разница между числителем и знаменателем индексов

$\Delta^p r q = \sum p_1 q_1 - \sum p_0 q_1$, или $\Delta^p r q = \sum p_1 q_0 - \sum p_0 q_0$
означает:

- в первом случае — абсолютный прирост товарооборота (выручки от продаж) в результате среднего изменения цен или экономии (перерасход) денежных средств населения в результате среднего снижения (повышения) цен;
- во втором случае — условный абсолютный прирост товарооборота если бы объемы продаж в отчетном периоде совпали с объемами продаж в базисном периоде;

3) товарооборота (выручки от реализации или продаж):

$$I_{pq} = \sum p_1 q_1 / \sum p_0 q_0$$

где pq — индексируемое сложное явление, в состав которого входят соизмеримые элементы совокупности. Разница между числителем и знаменателем индекса $\Delta r q = \sum p_1 q_1 / \sum p_0 q_0$ составляет абсолютное изменение товарооборота за счет совместного действия обоих факторов: цен на продукцию и ее количества.

Формулы средних индексов из индивидуальных:

1) физического объема:

$\sum i_q q_0 p_0 / \sum q_0 p_0 = \sum i_q d_{pq}^0$ - средний арифметический индекс,

где d_{pq}^0 — доля товарооборота отдельных видов продукции в общем товарообороте базисного периода;

2) цен:

$I_p = \sum p_1 q_1 / \sum (p_1 q_1 / i_p) = (\sum (d_{pq}^1 / i_p))^{-1}$ - средний гармонический индекс (Пааше),

где d_{pq}^1 — доля товарооборота отдельных видов продукции в общем товарообороте отчетного периода;

$I_p = \sum i_p p_0 q_0 / \sum p_0 q_0 = \sum i_p d_{pq}^0$ - средний арифметический индекс (Ласпейреса)

Если индексы качественных показателей построены на основе весов, взятых на уровне отчетного периода (например, по формуле Пааше), то рассмотренные выше агрегатные индексы, а также их элементы взаимосвязаны между собой:

$I_{pq} = I_p I_q$ (так называемая мультипликативная модель);

$\Delta r q = \Delta^p r q + \Delta^q r q$ (так называемая аддитивная модель).

Участие каждого фактора в формировании общего прироста товарооборота в относительном выражении может быть определено так:

$$d^p_{\Delta pq} = \Delta^p pq / \Delta pq = (\sum p_1 q_1 - \sum p_0 q_1) / (\sum p_1 q_1 - \sum p_0 q_0) = (I_{pq} - I_q) / (I_{pq} - 1) \quad (\text{фактор цен})$$

$$d^q_{\Delta pq} = \Delta^q pq / \Delta pq = (\sum p_0 q_1 - \sum p_0 q_0) / (\sum p_1 q_1 - \sum p_0 q_0) = (I_q - 1) / (I_{pq} - 1) \quad (\text{фактор объема})$$

При этом $d^p_{\Delta pq} + d^q_{\Delta pq} = 1$, или 100%.

Если сравнивают друг с другом не два периода (момента), а более, то выделяют **ценную и базисную** системы индексов.

Цепные и базисные индивидуальные индексы взаимосвязаны между собой:

- произведение цепных индексов равно конечному базисному;
- частное от деления двух смежных базисных индексов равно промежуточному цепному.

Между ценными и базисными общими индексами, построенными на основе постоянных весов, существует взаимосвязь, аналогичная взаимосвязи между индивидуальными индексами.

Индексы, построенные на основе переменных весов, непосредственно перемножать и делить нельзя.

Пример. Имеются следующие данные о проданных товарах:

Товары	Единица измерения	Количество, тыс. Ед.		Цена, руб.	
		Базисный период	Отчетный период	Базисный период	Отчетный период
А	Кг	1000	750	15	20
Б	л	2000	1800	5	6

Определить:

- 1) индивидуальные индексы объемов продаж в натуральном выражении, цен и товарооборота;
- 2) агрегатные индексы физического объема;
- 3) агрегатные индексы цен по формулам Пааше и Ласпейреса;
- 4) общий индекс товарооборота;

5) абсолютные приросты товарооборота за счет изменения объемов продаж, цен и за счет совместного действия обоих факторов.

Показать взаимосвязь между общими индексами и между абсолютными приростами товарооборота.

Решение.

1. По товару «А»:

$$i_q = q_1 / q_0 = 750 / 1000 = 0,75, \text{ или } 75\% \text{ (снижение на } 25\%);$$

$$i_p = p_1 / p_0 = 20 / 15 = 1,333, \text{ или } 133,3\% \text{ (рост на } 33,3\%);$$

$$i_{pq} = p_1 q_1 / p_0 q_0 = 20 \cdot 750 / 15 \cdot 1000 = 1,0, \text{ или } 100\% \text{ (без изменения)}$$

При этом

$$i_{pq} = i_p \cdot i_q \Rightarrow 1,000 \approx 1,333 \cdot 0,75.$$

По товару «Б»:

$$i_q = q_1 / q_0 = 1800 / 2000 = 0,90, \text{ или } 90\% \text{ (снижение на } 10\%);$$

$$i_p = p_1 / p_0 = 6 / 5 = 1,2, \text{ или } 120\% \text{ (рост на } 20\%);$$

$$i_{pq} = (p_1 q_1) / (p_0 q_0) = (6 \cdot 1800) / (5 \cdot 2000) = 1,08, \text{ или } 108\% \text{ (рост на } 8\%)$$

При этом

$$i_{pq} = i_p \cdot i_q \Rightarrow 1,08 = 1,2 \cdot 0,9.$$

$$2. I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{750 \cdot 15 + 1800 \cdot 5}{1000 \cdot 15 + 2000 \cdot 5} = \frac{20\ 250}{25\ 000} = 0,81, \text{ или } 81\%$$

(количество проданных товаров по двум видам в среднем снизилось на 19%).

Индексный метод широко применяется также для изучения динамики средних величин и выявления факторов, влияющих на динамику средних. С этой целью исчисляется система взаимосвязанных индексов: переменного, постоянного состава и структурных сдвигов.

Индекс переменного состава представляет собой отношение двух взвешенных средних величин с переменными весами, характеризующее изменение индексируемого (осредняемого) показателя.

Индекс переменного состава для любых качественных показателей имеет следующий вид:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}$$

Величина этого индекса характеризует изменение средне-взвешенной средней за счет влияния двух факторов: осредняемого показателя у отдельных единиц совокупности и структуры изучаемой совокупности.

Индекс постоянного (фиксированного) состава представляет собой отношение средних взвешенных с одними и теми же весами (при постоянной структуре). Индекс постоянного состава учитывает изменение только индексируемой величины и показывает средний размер изменения изучаемого показателя (\bar{x}) у единиц совокупности. В общем виде он может быть записан следующим образом:

$$I_x = (\sum x_1 f_1 / \sum f_1) / (\sum x_0 f_1 / \sum f_1)$$

Для расчета индекса постоянного состава можно использовать агрегатную форму индекса:

$$I_x = \sum x_1 f_1 / \sum x_0 f_1$$

Индекс структурных сдвигов характеризует влияние изменения структуры изучаемого явления на динамику среднего уровня индексируемого показателя и рассчитывается по формуле

$$I_{стр} = (\sum x_0 f_1 / \sum f_1) / (\sum x_0 f_0 / \sum f_0)$$

Под структурными изменениями понимается изменение доли отдельных групп единиц совокупности в общей их численности (d). Система взаимосвязанных индексов при анализе динамики среднего уровня качественного показателя имеет вид

$$I_{\bar{x}} = I_x I_{стр}$$

В индексах средних уровней в качестве весов могут быть взяты удельные веса единиц совокупности ($d = f / \sum f$), которые отражают изменения в структуре изучаемой совокупности. Тогда систему взаимосвязанных индексов можно записать в следующем виде:

$$\sum x_1 d_1 / \sum x_0 d_0 = (\sum x_1 d_1 / \sum x_0 d_1) (\sum x_0 d_1 / \sum x_0 d_0)$$

Или

Индекс средней величины (переменного состава) = индекс постоянного состава \times индекс структурных сдвигов.

Аналогично приведенным формулам строятся индексы средних уровней: цен, себестоимости продукции, фондоотдачи, производительности труда, оплаты труда и др.

Пример. Имеются данные о выпуске однородной продукции по предприятиям АО:

№ предприятия АО	Выпуск продукции				Себестоимость единицы продукции, руб.		Индивидуальный индекс стоимости
	I квартал		II квартал		I кварт ал	II кварт ал	
	Тыс . ед. q_0	% d_0	Тыс . ед. q_1	% d_1	Z_0	Z_1	
1	40	40	36	30	7,0	8,0	1,143
2	60	60	84	70	6,0	6,5	1,083
Итого	100	100	120	100	6,4	6,95	1,086

Определить для двух предприятий:

- 1) среднюю себестоимость единицы продукции;
- 2) индекс средней себестоимости продукции;
- 3) среднее изменение себестоимости продукции;
- 4) индекс структурных сдвигов.

Решение. 1. Средняя себестоимость единицы данного вида продукции по двум предприятиям АО определяется как средняя арифметическая взвешенная:

$$\bar{z}_0 = \frac{\sum z_0 q_0}{\sum q_0} = \frac{7 \cdot 40 + 6 \cdot 60}{100} = \frac{640}{100} = 6,4 \text{ руб.}$$

$$\bar{z}_1 = \frac{\sum z_1 q_1}{\sum q_1} = \frac{8 \cdot 36 + 6,5 \cdot 84}{120} = \frac{834}{120} = 6,95 \text{ руб.}$$

2. Индекс себестоимости продукции переменного состава равен

$$I_z = \frac{\bar{z}_1}{\bar{z}_0} = \frac{6,95}{6,4} = 1,086, \quad \text{или } 108,6\%$$

(средняя себестоимость единицы продукции по двум предприятиям возросла на 8,6%).

$$3. \quad I_z = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_1}{\sum q_1} = \frac{834}{120} : \frac{756}{120} = \frac{6,95}{6,3} = 1,103 \quad \text{или } 110,3\%$$

Это означает, что в среднем по двум предприятиям себестоимость единицы продукции повысилась на 10,3%.

4. Индекс структурных сдвигов может быть рассчитан с помощью взаимосвязи индексов:

$$I_{стр.} = I_z / I_z = \frac{1,086}{1,103} = 0,985, \quad \text{или } 98,5\%.$$

Средняя себестоимость единицы продукции по двум предприятиям снизилась на 1,5% за счет изменения удельного веса отдельных предприятий в общем выпуске продукции.

Если в качестве весов взять удельные веса отдельных предприятий в общем объеме выпуска продукции ($d = q / \sum q$), то получим следующую систему взаимосвязанных индексов:

$$\begin{aligned} & (8 \cdot 0,3 + 6,5 \cdot 0,7) / (7 \cdot 0,4 + 6,0 \cdot 0,6) = (8 \cdot 0,3 + 6,5 \cdot 0,7) \\ & / (7 \cdot 0,3 + 6,0 \cdot 0,6) \quad (7 \cdot 0,3 + 6,0 \cdot 0,6) / (7 \cdot 0,4 + 6,0 \cdot 0,6); \\ & 6,95 / 6,4 = (6,95 / 6,3) \cdot (6,3 / 6,4); \\ & 1,086 = 1,103 \cdot 0,985 \end{aligned}$$

Общий вывод: если бы происшедшие изменения себестоимости продукции не сопровождались структурными перераспределениями в ее выпуске, то средняя себестоимость продукции по двум предприятиям возросла бы на 10,3%. Изменение структуры выпуска продукции отдельных предприятий в общем объеме выпуска вызвало снижение себестоимости на 1,5%. Одновремен-

ное воздействие двух факторов увеличило среднюю себестоимость продукции по двум предприятиям на 8,6%.

Задачи для самостоятельного решения.

Задача 1. Имеются данные о продаже товаров в розничной торговле города

Товар	Продано, тыс. кг		Средняя цена продажи 1 кг товара в базисном периоде, руб.	Индекс цен в отчетном периоде к базисному, %
	Базисный период	Отчетный период		
А	20	20,4	100	115
Б	46	43,7	50	120
В	6	4,2	120	150

Определите:

- 1) индексы физического объема продаж по каждому товару;
- 2) сводные индексы: физического объема товарооборота, цен и товарооборота;
- 3) абсолютный прирост товарооборота за счет изменения: а) цен; б) объема продажи товаров. •

Задача 2. Затраты на производство продукции на предприятии характеризуется данными:

Вид продукции	Затраты на производство продукции в отчетном ме-	Индексы себестоимости единицы
А	770	110
Б	490	98

Определите по двум видам продукции:

1. Общий индекс себестоимости продукции и абсолютный прирост (снижение) затрат в связи с изменением себестоимости единицы продукции.
2. Общий индекс физического объема продукции, если известно, что затраты на продукцию за прошедший период возросли на 15,5%.

Задача 3. Имеются данные о продаже товаров на рынках города.

Т	Продано товара, тыс. кг	Средняя цена товара, руб.
о-		

вар	Базис- ный период	Отчет- ный период	Базис- ный период	Отчет- ный период
А	20	25	10	12
Б	60	50	30	45

Определите:

- 1) индивидуальные индексы цен и объема продажи товаров;
- 2) общие (сводные) индексы:
 - а) товарооборота;
 - б) цен;
 - в) физического объема товарооборота;
- 3) абсолютное изменение товарооборота по двум товарам вследствие изменения: а) цен и б) объема продажи.

Задача 4. Имеются данные о продаже товаров в районе:

Товарные группы	Продано товаров в 2000 г. млн	Прирост количества продажи в 2001 г. к 2000 г.
Ткани	700	+10
Обувь	1000	+20
Трикотажные изделия	800	Без изменения

Определите: 1) общий индекс физического объема товарооборота; 2) абсолютный прирост стоимости проданных товаров за счет изменения количества продажи товаров.

Задача 5. Розничный товарооборот по трем районам характеризуется следующими данными:

Район	Товарооборот в отчетном году, млн. руб.	Индексы товарооборота в отчетном году по сравнению с базисным	
		В фактических ценах, число раз	В сопоставимых ценах, %
1	2040	1,26	105
2	1500	1,20	80
3	2 492	1,19	85

Определите

- 1) среднее изменение цен на товары по каждому району;
- 2) общий индекс цен по трем районам;
- 3) общий индекс физического объема товарооборота;
- 4) общий индекс товарооборота в фактических ценах;
- 5) абсолютный прирост товарооборота вследствие изменения цен, объема продажи по каждому району и в целом по трем районам.

Задача 6. Динамика объема продукции и затрат на ее производство характеризуется данными:

Вид продукции	Фактические затраты на производство, тыс. руб.		Индекс физического объема продукции II кв. к I кв., %
	I	II квартал	
А	300	360	105
Б	200	250	90

Определите общие индексы:

- 1) физического объема продукции;
- 2) себестоимости продукции;
- 3) абсолютный прирост (снижение) затрат за счет изменения объема продукции, себестоимости продукции.

Задача 7. Имеются данные о продаже продуктов в городе:

	Товарооборот,		Индекс цен в отчетном периоде к базисному, %
	Базисный период	Отчетный период	
Овощи	160	272	80
Молочные продукты	240	228	120

Определите:

- 1) общий индекс товарооборота;
- 2) общий индекс цен: Г. Пааше и Э. Ласпейреса, поясните их результаты;
- 3) общий индекс физического объема товарооборота;
- 4) покажите взаимосвязь исчисленных индексов.

Задача 8. Имеются данные о продаже товаров в розничной торговле города:

Товар	Удельный вес продажи товара в базисном периоде	Индексы объема продажи
А	30	120
Б	50	80
В	20	70
Итого	100	-

Определите общий индекс объема продажи товаров.

Задача 9. Имеются следующие данные о выпуске одноименной продукции «А» и ее себестоимости по двум предприятиям:

Предприятия	Производство		Себестоимость 1	
	I	II	1	II
1	80	90	70	65
2	70	10	58	50

Вычислите:

1. Индекс себестоимости переменного состава;
2. Индекс себестоимости постоянного состава;
3. Индекс структурных сдвигов.

Покажите взаимосвязь между исчисленными индексами. Поясните полученные результаты индексов.

Задача 10. Имеются следующие данные о заработной плате работников предприятий по отраслям экономики:

От- расль	Средняя заработная плата, тыс. руб.		Удельный вес среднесписоч- ной численности работников, %	
	Период		Период	
	ба- за	отчетный	базисный	отчетный
1	4,0	6,0	30	45
2	3,5	4,2	50	35
3	2,5	4,0	20	20

За прошедший период цены на потребительские товары и услуги по трем районам в среднем возросли на 32%.

Определите:

1. По каждому району абсолютный и относительный прирост средней заработной платы.

2. По трем районам вместе: .

Индексы средней заработной платы:

- а) переменного состава; .
- б) постоянного состава;
- в) структурных сдвигов.

3. Абсолютный прирост средней заработной платы вследствие изменения;

- а) заработной платы в каждом районе;
- б) структуры среднесписочной численности работников.

4. Индекс покупательной способности рубля.

5. Индекс реальной заработной платы.

Задача 11. Динамика производственных показателей двух предприятий АО, выпускающих одноименную продукцию «А», характеризуется данными:

Предприятие	Выпуск продукции, тыс.		Себестоимость единицы продукции, тыс. руб.	
	I квартал	II квартал	I квартал	II квартал
1	8	10	10	12
2	12	15	9	8

Определите для двух предприятий:

1. Индекс средней себестоимости (индекс переменного состава).
2. Среднее изменение себестоимости продукции (индекс постоянного состава).
3. Индекс структурных сдвигов.
4. Абсолютный прирост средней себестоимости за счет изменения: а) себестоимости продукции на каждом предприятии; б) структуры произведенной продукции.

Задача 12. Имеются данные о производстве одноименной продукции и ее себестоимости по двум АО:

АО	Объем продукции, %		Себестоимость единицы продукции, тыс.	
	I кв.	II кв.	I кв.	II кв.
1	58	40	380	400
2	42	60	200	300

Определите:

- 1) индексы себестоимости продукции по каждому АО;
- 2) общие индексы себестоимости продукции для двух АО (переменного, постоянного состава и структурных сдвигов). Покажите взаимосвязь между исчисленными индексами. Поясните результаты индексов.

7 СТАТИСТИЧЕСКИЕ ИЗУЧЕНИЯ СВЯЗИ

тия										
1	12,0	462	11,0	462	1	1	2	1	1	1
2	18,8	939	12,0	506	2	2	5	6	-1	1
3	11,0	506	15,4	765	3	3	1	2	-1	1
4	29,0	110	17,5	804	4	4	9	9	0	0
5	0	8	18,8	872	5	5	4	5	-1	1
6	17,5	872	8	939	6	6	7	3	4	16
7	23,4	136	20,8	998	7	7	10	10	0	0
8	35,6	8	23,4	100	8	8	3	8	-5	25
9	15,4	100	4	2	9	9	8	7	1	1
10	4	2	26,1	8	10	10	6	4	2	4
	26,1	998	29,0	136						Σ=5
	20,7	804	35,6	8						3

$$\rho = 1 - \frac{6 \cdot 53}{10 \cdot 99} = 1 - \frac{318}{990} = 0,679$$

Следовательно, связь между исследуемыми показателями достаточно тесная.

Для определения тесноты связи между тремя и более признаками применяется ранговый коэффициент согласия — коэффициент конкордации, который вычисляется по формуле

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

где m — количество факторов; n — число наблюдений; S — сумма квадратов отклонений рангов.

Пример. Одновременно с приведенной выше задачей определялась также теснота связи между объемом реализованной продукции (млн. руб.), суммой накладных расходов на реализацию (тыс. руб.), себестоимостью единицы продукции (коп.) и средней заработной платой рабочих (руб.). Зависимость между исследуемыми факторами представлена таблицей:

Номер предприятия	Реализация продукции, млн. руб. (y)	Накладные расходы на реализацию, тыс. руб. (x)	Себестоимость единицы продукции, коп. (z)	Средняя месячная заработная плата рабочих, руб. (U)
1	12,0	462	68,8	168,5
2	18,8	939	70,2	158,7
3	11,0	506	71,4	171,7
4	29,0	1 108	78,5	183,9
5	17,5	872	66,9	160,4
6	23,9	765	69,7	165,2
7	35,6	1 368	72,3	175,0
8	15,4	1002	77,5	170,4
9	26,1	998	65,2	162,7
10	20,7	804	70,7	163,0

Тесноту связи рассчитаем с помощью коэффициента конкордации. Для этого присваиваем ранги исходным данным:

R _y	R _x	R _z	R _u	Сумма строк	Квадраты сумм
2	1	3	6	12	144
5	6	5	1	17	289
1	2	7	8	18	324
9	9	10	10	38	1
7	5	2	2	13	444
10	3	4	5	19	169
3	10	8	9	37	361
8	8	9	7	27	1
6	7	16	3	19	369
	4		4	20	729
				Σ=	361
				220	400
					Σ=5
					600

$$S = 5600 - \frac{(220)^2}{10} = 5600 - 4840 = 760$$

Величина коэффициента конкордации показывает, что между исследуемыми величинами имеется тесная зависимость. Эта зависимость объясняется тем, что все рассматриваемые величины являются показателями эффективности работы предприятий.

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

Для определения тесноты связи двух качественных признаков, каждый из которых состоит только из двух групп, применяются *коэффициенты ассоциации и контингенции*. Для их вычисления строится таблица, которая показывает связь между двумя явлениями, каждое из которых должно быть альтернативным, т. е. состоящим из двух качественно отличных друг от друга значений признака (например, хороший, плохой).

Коэффициенты вычисляются по формулам:

$$A = \frac{ad - bc}{ad + bc} \quad \text{ассоциация;}$$

$$K = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}} \quad \text{контингенция.}$$

Коэффициент контингенции всегда меньше коэффициента ассоциации. Связь считается подтвержденной, если $A \geq 0,5$, или $K \geq 0,3$.

Пример. Исследовалась связь между выполнением норм выработки молодыми рабочими и уровнем образования. Результаты обследования характеризуются следующими данными:

Группы рабочих	Выполняющие нормы	Не выполняющие нормы	Всего
Окончившие среднюю школу	78	22	100
Не окончившие среднюю школу	32	68	100

	110	90	200
--	-----	----	-----

$$A = \frac{78*68 - 32*22}{78*68 + 32*22} = \frac{4600}{6008} = 0,766$$

;

$$K = \frac{78*68 - 32*22}{\sqrt{(78+22)(22+68)(78+32)(32+68)}} = \frac{5304 - 704}{\sqrt{99000000}} = 0,46$$

Традиционные методы корреляционного анализа позволяют не только оценить тесноту связи, но и выразить эту связь аналитически. Применению корреляционного анализа должен предшествовать качественный, теоретический анализ исследуемого социально-экономического явления или процесса.

Связь между двумя факторами аналитически выражается уравнениями:

прямой: $\bar{y}_x = a_0 + a_1 x$;

гиперболы: $\bar{y}_x = a_0 + \frac{a_1}{x}$

параболы: $\bar{y}_x = a_0 + a_1 x + a_2 x^2$ (или другой ее степени);

степенной функции: $\bar{y}_x = a_0 x^{a_1}$

Параметр a_0 показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов. Параметр a — коэффициент регрессии, показывает, насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу. На основе этого параметра вычисляются *коэффицици-*

енты эластичности, которые показывают изменение результативного признака в процентах в зависимости от изменения факторного признака на 1%:

$$\varepsilon = a_1 \frac{\bar{x}}{\bar{y}}$$

Для определения параметров уравнений используется метод наименьших квадратов, на основании которого строится соответствующая система уравнений.

Теснота связи при линейной зависимости измеряется с помощью линейного коэффициента корреляции:

$$r = \frac{x\bar{y} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

а при криволинейной зависимости с помощью корреляционного отношения:

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

Пример. По данным о стоимости основных производственных фондов и объеме валовой продукции нужно определить уравнение связи и тесноту связи. Связь предполагается линейной. Принимая для этой связи выше написанное уравнение прямой линии, определим его параметры на основе метода наименьших квадратов, решив следующую систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ \dots \end{cases}$$

Расчеты указанных в системе уравнений сумм произведем в табличной форме:

Стоимость основных производственных фондов, млн. руб. (x)	объем валовой продукции, млн. руб. (y)	xy	x ²	y ²	\bar{y}_x
1	20				19,
2	25				4
3	31			400	25,
4	31			625	0
5	40			961	30,
6	56	20	1	961	6
7	52	50	4	160	36,
8	60	93	9	0	2
9	60	124	16	313	41,
10	70	200	25	6	8
		336	36	270	47,
		364	49	4	4
		480	64	360	53,
		540	81	0	0
		700	10	360	58,
			0	0	6
				490	64,
				0	2
					69,
					8
$\Sigma = 55$	$\Sigma = 446$	$\Sigma = 2907$	$\Sigma = 38$	$\Sigma = 22487$	$\Sigma = 44$

			5		6
--	--	--	---	--	---

$$10a_0 + 55a_1 = 445$$

$$a_0 = 13,8 \quad a_1 = 5,6$$

$$\bar{y}_x = 13,8 + 5,6x$$

Следовательно, с увеличением стоимости основных производственных фондов на 1 млн. руб. объем валовой продукции увеличивается в среднем на 5,6 млн. руб., или с увеличением стоимости основных производственных фондов на 1 % объем валовой продукции увеличивается на 0,69%.

Определим коэффициент эластичности:

$$\varepsilon = 5,6 * 5,5 / 44,5 = 30,8 / 44,5 = 0,6921$$

Рассчитаем величину линейного коэффициента корреляции:

$$r = \frac{\frac{2970}{10} - \frac{55}{10} \cdot \frac{445}{10}}{\sqrt{\left[\frac{22487}{10} - \left(\frac{445}{10} \right)^2 \right] * \left[\frac{385}{10} - \left(\frac{55}{10} \right)^2 \right]}} = 0,97$$

Расчет коэффициентов регрессии несколько осложняется, если ряды по исследуемым факторам сгруппированы, а связь криволинейная.

Пример. Связь между выпуском изделий и затратами на упаковку характеризуется следующими данными по 50 однотипным заводам:

Стоимость	Выпуск изделия, тыс. шт. (x)	Количество
-----------	------------------------------	------------

упаков- ки, тыс. руб. (y)					заводов
10				2	2
8		1	6	7	14
6	3	9	7	1	20
4	6	5	2		13
2	1				1
Количе- ство заводов	10	15	15	10	50

Считая, что зависимость описывается уравнением параболы второго порядка, определим коэффициенты a_0 , a_1 , a_2 , а также рассчитаем показатель тесноты связи.

Составим систему нормальных уравнений для определения коэффициентов:

$$\begin{cases} na_0 + a_1 \sum x + a_2 \sum x^2 = \sum y \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum xy \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum x^2 y \end{cases}$$

x \ y	1	2	3	4	f_y	yf_y	xyf_y y	x^2yf_y xy
10				2	2	20	80	320
8		1	6	7	14	11	37	136
6	3	9	7	1	20	2	2	0
4	6	5	2		13	12	27	708
2	1				1	0	6	176
f_x	10	15	15	10	50	52	88	2

xf_x	10	30	45	40	125	2	2	256
x^2f	10	60	135	160	365	30	81	6
x	10	12	405	640	117	6	8	
x^3f	10	0	121	256	5			
x	4,4	24	5	0	402			
x^4f	0	0	6,5	8,2	5			
x		5,4	3	0				
\bar{y}_i		7						

$$\begin{cases} 50a_0 + 125a_1 + 365a_2 = 306; \\ 125a_0 + 365a_1 + 1175a_2 = 818; \end{cases}$$

Преобразуем систему путем деления соответствующих значений на коэффициенты при первых членах:

$$a_0 + 2,50a_1 + 7,30a_2 = 6,12;$$

$$a_0 + 2,92a_1 + 9,56a_2 = 6,54;$$

$$a_0 + 3,27a_1 + 11,03a_2 = 7,03,$$

$$\text{отсюда } a_0 = 10,9; a_1 = 6,00; a_2 = 1,4.$$

Следовательно,

$$\bar{y} = 10,9 - 6,0x + 1,4x^2.$$

Имея $\bar{y} = 6,12$ и $\sigma_y^2 = 2,63$, можно рассчитать δ^2 . Для этого построим вспомогательную таблицу:

\bar{y}_i	$\bar{y}_i - \bar{y}$	$(\bar{y}_i - \bar{y})^2$	f_x	$(\bar{y}_i - \bar{y})^2 f_x$
4,40	—	2,9584	10	29,5
5,47	1,72	0,4225	15	840
6,53	—	0,1681	15	6,33
8,20	0,65	4,3264	10	75
	0,41			2,52
	2,08			16
				43,3
				264

			50	81,7694
--	--	--	----	---------

Следовательно, $\delta^2 = 1,64$.

Вычисляем эмпирическое корреляционное отношение

$$\eta = \sqrt{\frac{\delta^2}{\sigma_y^2}}; \eta = \sqrt{\frac{1,64}{2,63}} = 0,78.$$

Если зависимость между двумя факторами выражается уравнением гиперболы

$$\bar{y}_x = a_0 + \frac{a_1}{x}$$

то система уравнений для определения параметров a_0 и a_1 такова:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y \end{cases}$$

Для определения параметров уравнения регрессии, выраженного степенной функцией $\bar{y} = a_0 x^{a_1}$ приводят функцию к линейному виду:

$\lg \bar{y}_x = \lg a_0 + a_1 \lg x$, отсюда система уравнений для определения параметров запишется:

$$\begin{cases} n \lg a_0 + a_1 \sum \lg x = \sum \lg y \end{cases}$$

Зависимость между тремя и более факторами называется множественной или многофакторной корреляционной зависимостью.

Линейная связь между тремя факторами выражается уравнением

$$\bar{y}_{12} = a_0 + a_1 x + a_2 z$$

а система нормальных уравнений для определения неизвестных параметров a_0, a_1, a_2 будет следующей:

$$\begin{cases} \sum y_i = n a_0 + a_1 \sum x_i + a_2 \sum z_i \\ \sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i z_i \\ \sum z_i y_i = a_0 \sum z_i + a_1 \sum x_i z_i + a_2 \sum z_i^2 \end{cases}$$

Теснота связи между тремя факторами измеряется с помощью множественного (совокупного) коэффициента корреляции

$$R = \sqrt{\frac{r^2_{xy} + r^2_{yz} - 2r_{yx} * r_{yz} * r_{xz}}{1 - r^2_{xz}}}$$

где r_{ij} — парные коэффициенты корреляции между соответствующими факторами.

Для более углубленного анализа вычисляются частные коэффициенты корреляции.

Пример. Взаимосвязь между среднегодовой стоимостью основных производственных фондов, относительным уровнем затрат на реализацию продукции (в процентах) и стоимостью реализованной продукции характеризуется следующими данными:

Номер предприятия	Стоимость основных производственных фондов. млн. руб.	Уровень затрат на реализацию (в % к стоимости ре-	Объем реализованной продукции млн. руб.

		лизован- ной продукц ии)	
1	3	4	20
2	3	3	25
3	5	3	20
4	6	5	30
5	7	10	32
6	6	12	25
7	8	12	29
8	9	11	37
9	9	15	36
10	10	15	40

Считая зависимость между этими показателями линейной, определим параметры уравнений регрессии a_e , $a_{,}$, a_a и вычислим множественный и частные коэффициенты корреляции. Для этого составим вспомогательную таблицу для расчета коэффициентов регрессии и корреляции:

Номер предпр иятия	x	z	y	x^2	z^2	xz	xy	zy	y^2
1	3	4	2	9	16	1	60	80	40
2	3	3	0	9	9	2	75	75	0
3	5	3	2	25	9	9	10	60	62
4	6	5	5	36	25	1	0	15	5
5	7	1	2	49	1	5	18	0	40
6	6	0	0	36	0	3	0	32	0
7	8	1	3	64	1	0	22	0	90
8	9	2	0	81	4	7	4	30	0
9	9	1	3	81	1	0	15	0	10

0	1	1	2	2	6	12	7	0	34	24
		0	1	2	6	1	2	23	8	62
			1	5	4	22	8	2	40	5
			1	2	8	5	4	33	7	84
			5	9	1	22	9	3	54	1
			1	3	9	5	9	32	0	13
			5	7	1		1	4	60	69
				3	1		3	40	0	12
				6	0		5	0		96
				4	0		1			16
				0			5			00
							0			
Итого	6	9	2	4	10	6	20	28	90	
о	6	0	9	9	18	7	78	80	80	
			4	0		6				

$$\begin{cases} 10a_0 + 66a_1 + 90a_2 = 294; \\ 66a_0 + 490a_1 + 676a_2 = 2078; \\ 66a_0 + 490a_1 + 676a_2 = 2078; \end{cases}$$

Отсюда $a_0 = 12,88$; $a_1 = 2,08$; $a_2 = 0,31$.

Следовательно,

$$\bar{y}_{xz} = 12,88 + 2,08x + 0,31z;$$

$$r_{xy} = 0,88; \quad r_{xz} = 0,79; \quad r_{yz} = 0,76.$$

Множественный коэффициент корреляции равен:

$$R = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{yx}r_{yz}r_{xz}}{1 - r_{xz}^2}} = \sqrt{\frac{0,7744 + 0,5776 - 2 * 0,88 * 0,79 * 0,76}{1 - 0,6241}} = 0,78.$$

Рассчитаем частные коэффициенты:

$$r_{yx(z)} = \frac{r_{xy} - r_{yz}r_{xz}}{\sqrt{(1 - r_{yz}^2)(1 - r_{xz}^2)}} = \frac{0,88 - 0,79 * 0,76}{\sqrt{(1 - 0,5776) * (1 - 0,6241)}} = 0,69;$$

$$r_{yz(x)} = \frac{r_{xy} - r_{yx} * r_{xz}}{\sqrt{(1 - r_{yx}^2)(1 - r_{xz}^2)}} = \frac{0,76 - 0,88 * 0,79}{\sqrt{(1 - 0,7744) * (1 - 0,6241)}} = \frac{0,0648}{\sqrt{0,0840}} = \frac{0,0648}{0,29} = 0,22$$

Задачи для самостоятельного решения.

1. Задача. Установите направление и характер связи между среднегодовой стоимостью производственных основных фондов и объемом валовой продукции по 10 однотипным предприятиям, применив для этого метод приведения параллельных данных:

Номер предприятия	Стоимость основных производственных фондов, млн. руб.	Объем валовой продукции, млн. руб.
1	3,8	5,4
2	9,0	10,9
3	2,0	6,8
4	4,9	8,5
5	4,5	9,3
6	5,2	9,8
7	3,4	6,5
8	12,0	15,6
9	10,2	12,9
10	12,5	16,4

2. Задача. С помощью метода приведения параллельных данных установите направление и характер связи между тремя факторами по 12 совхозам одной специализации:

Номер предприятия	Объем реализованной	Балансовая прибыль,
-------------------	---------------------	---------------------

	Продукции, млн. руб.	тыс. руб.
1	1,7	20
2	2,2	75
3	8,6	41
4	1,3	82
5	3,4	106
6	3,9	129
7	4,7	145
8	5,8	180
9	3,6	210
10	6,4	250
11	7,2	262
12	9,1	360
13	10,5	420
14	12,6	300
15	10,8	480
16	13,3	550
17	11,4	600

Определите вид корреляционной зависимости, постройте уравнение регрессии, рассчитайте параметры уравнения. Вычислите тесноту связи. Объясните полученные статистические характеристики.

3 Задача. По 16 предприятиям, изготавливающим одноименные запасные стали к сельскохозяйственным машинам, были получены следующие данные об объеме производства деталей и затратах на изготовление одной тысячи деталей:

Номер предприятия	Объем производства, тыс. шт.	Затраты на изготовление 1 тыс. шт., тыс. руб.	Номер предприятия	Объем производства, тыс. шт.	Затраты на изготовление 1 тыс. шт., тыс. руб.
-------------------	------------------------------	-----------------------------------------------	-------------------	------------------------------	-----------------------------------------------

1	1,2	12	9	4,5	4
2	1,8	10	10	5	4
3	2,0	14	11	6	3
4	2,5	8	12	7	3
5	3,0	6	13	8	2
6	3,2	7	14	9	2
7	3,5	5	15	10	1
8	4,0	5	16	12	1

Определите вид корреляционной зависимости, постройте уравнение регрессии, рассчитайте параметры уравнения. Вычислите показатель тесноты связи. Дайте анализ полученным данным.

4 Задача. Ниже приводятся данные об этажности домов и расходах на эксплуатацию 1 м² в год оплачиваемой площади:

Расходы на эксплуатацию 1 м ² оплачиваемой площади, руб.	Этажность жилых домов				Итого
	1-3	4-5	6-8	9-11	
1,2 —					
1,4					
1,4 —			9		24
1,6			4		16
1,6 —	5	12	2	15	10
1,8	20	3			20
1,8 —					
2,0					
	25	15	15	15	70

Определите вид корреляционной зависимости, постройте уравнение регрессии, рассчитайте параметры. Вычислите показатель тесноты связи.

Задача 5. Вычислите корреляционное отношение, если известно, что общая дисперсия равна 35,68, групповые дисперсии $\sigma_1^2 = 12$, $\sigma_2^2 = 6$, $\sigma_3^2 = 18$, а численность групп соответственно 36, 43, 21 единиц.

Задача 6. Общая дисперсия $\sigma^2 = 8,4$, общая средняя $\bar{x} = 13,0$, групповые средние $\bar{x}_1 = 10$, $\bar{x}_2 = 15$, $\bar{x}_3 = 12,0$, численности групп соответственно равны 32, 53, 45. Определите величину корреляционного отношения.

Задача 7. Надой молока и себестоимость 1 кг его по 12 колхозам характеризуются следующими данными:

Себестоимость 1 кг, коп.	31	30	26	27	26	24	23	21
20	18	16	13					
Надой молока, тыс. ц.	0,8	1,4	1,9	2,4	3,1	4,0	5,8	7,3
8,0	9,7	11,4	14,0					

Постройте уравнение гиперболы, вычислите параметры, рассчитайте корреляционное отношение.

Задача 8. По данным об уровне рентабельности и уровне издержек обращения по 40 магазинам определите вид корреляционной зависимости, рассчитайте параметры уравнения регрессии и вычислите показатель тесноты связи; Проанализируйте полученные результаты.

Уровень издержек обращения, %	Уровень рентабельности, %				Итого
2,5 — 3,0				9	9
3,0 — 3,5			6	1	7
3,5 — 4,0		3	4		7

4,0 — 4,5	2	7			9
4,5 — 5,0	8				8
Итого	10	10	10	10	40

По этим данным постройте многофакторную регрессионную модель, отражающую влияние приведенных факторов на величину рентабельности. Проведите предварительный качественный анализ исходных данных, определите вид уравнения регрессии, отберите факторы, имеющие наибольшее влияние на результативный признак. Рассчитайте множественный, частные и парные коэффициенты корреляции. Определите эластичность факторов. Проанализируйте полученные результаты.

ЛИТЕРАТУРА

1. Вентцель, Е.С., Овчаров, Л.А. Теория случайных процессов и ее инженерные приложения [Текст] : Учеб. пособие для втузов. / Вентцель Е.С., Овчаров Л.А. – 2-е изд., стер. – М. : Высш. шк., 2000. – 383 с. : ил. – (Высшая математика для втузов). – ISBN 5-06-003831-9.
2. Случайные процессы [Текст] : Учеб. для вузов. / Волков И.К., Зуев С.М., Цветкова Г.М. ; Под ред. В.С. Зарубина, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 448 с. – (Математика в техническом университете. Вып. XVIII). – ISBN 5-7038-1267-4 (Вып. XVIII) ; ISBN 5-7038-1270-4.
3. Вентцель, Е.С., Овчаров, Л.А. Теория вероятностей и ее инженерные приложения [Текст] : Учеб. пособие для втузов. / Вентцель Е.С., Овчаров Л.А. – 2-е изд., стер. – М. : Высш. шк., 2000. – 480 с. : ил. – (Высшая математика для втузов). – ISBN 5-06-003830-0.
4. Теория вероятностей [Текст] : Учеб. для вузов. / А.В. Печин-

кин, О.И. Тескин, Г.М. Цветкова и др. ; Под ред. В.С. Зарубина, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 456 с. : ил. – (Математика в техническом университете. Вып. XVI). – ISBN 5-7038-1307-7 (Вып. XVI) ; ISBN 5-7038-1270-4.

5. Дрейпер, Н., Смит, Г. Прикладной регрессионный анализ [Текст] : монография : в 2 т. / Дрейпер Н., Смит Г. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 1986-1987. – (Математико-статистические методы за рубежом).

Т. 1 : – М. : Финансы и статистика, 1986. – 366 с. : ил.

Т. 2 : – М. : Финансы и статистика, 1987. – 351 с. : ил.

6. Химмельблау, Д. Анализ процессов статистическими методами [Текст] / Д. Химмельблау. – М. : Мир, 1973. – [957 с.] : ил.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
**Федеральное государственное автономное образовательное учреждение
высшего образования**
**«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
НЕВИННОМЫССКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ
(ФИЛИАЛ)»**

**Методические указания к самостоятельной работе
для студентов направления
15.03.04 «Обработка экспериментальных данных»
по дисциплине
«ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»**

Невинномысск, 2024

Методические указания разработаны в соответствии с требованиями ФГОС ВО в части содержания и уровня подготовки выпускников направления подготовки 15.03.04 «Автоматизация технологических процессов и производств».

Методические указания содержат рекомендации по организации самостоятельной работы студента при изучении дисциплины «Обработка экспериментальных данных».

Составитель доцент кафедры ИСЭА Д.В. Болдырев

Ответственный редактор доцент кафедры ИСЭА А.А. Евдокимов

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. ОБЩАЯ ХАРАКТЕРИСТИКА САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУ- ЧАЮЩИХСЯ ПРИ ИЗУЧЕНИИ ДИСЦИПЛИНЫ «ОБРАБОТКА ЭКС- ПЕРИМЕНТАЛЬНЫХ ДАННЫХ».....	6
1.1. Подготовка к лекциям.....	8
1.2. Подготовка к лабораторным занятиям.....	9
1.3. Самостоятельное изучение материала тем.....	10
2. СРЕДСТВА ОЦЕНИВАНИЯ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕ- ТЕНЦИЙ ОБУЧАЮЩИХСЯ ПРИ ИЗУЧЕНИИ ДИСЦИПЛИНЫ «ОБРА- БОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ».....	13
3. ОТЧЕТНОСТЬ ПО ДИСЦИПЛИНЕ.....	15
РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА.....	17

ВВЕДЕНИЕ

Дисциплина «Обработка экспериментальных данных» ставит своей целью формирование следующих компетенций будущего бакалавра по направлению подготовки 15.03.04 — Обработка экспериментальных данных.

Код, формулировка компетенции	Код, формулировка индикатора	Планируемые результаты обучения по дисциплине (модулю), характеризующие этапы формирования компетенций, индикаторов
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИД-1 _{УК-1} Выделяет проблемную ситуацию, осуществляет ее анализ и диагностику на основе системного подхода	Применяет системный подход при анализе проблемной ситуации
	ИД-2 _{УК-1} Осуществляет поиск, отбор и систематизацию информации для определения альтернативных вариантов стратегических решений в проблемной ситуации	Определяет альтернативные варианты решений проблемы на основе отобранной и систематизированной информации
	ИД-3 _{УК-1} Определяет и оценивает риски возможных вариантов решений проблемной ситуации, выбирает оптимальный вариант её решения	Оценивает риски возможных решений проблемы, выбирает оптимальный вариант ее решения
ПК-1. Способен исследовать автоматизируемый объект и подготавливать рекомендации по его автоматизации с учетом современного уровня развития профессиональной сферы	ИД-1 _{ПК-1} Планирует экспериментальные исследования, составляет научные отчеты и внедряет результаты исследований и разработок в области автоматизации технологических процессов и производств	Разрабатывает планы эксперимента в области автоматизации технологических процессов и производств, составляет научные отчеты в соответствии с нормативными требованиями и организует внедрение результатов исследова-

		ний в практическую деятельность промышленных предприятий и организаций
	ИД-2 _{ПК-1} Проводит эксперименты по заданным методикам, обрабатывает и анализирует их результаты	Знает методики проведения эксперимента; проводит эксперимент по заданным методикам; проводит математическую и статистическую обработку опытных данных о характеристиках средств и систем автоматизации.
	ИД-3 _{ПК-1} Анализирует научно-техническую информацию, отечественный и зарубежный опыт по направлению исследований в области автоматизации технологических процессов и производств	Выполняет сбор и анализ научно-технической информации, отечественного и зарубежного опыта в области автоматизации технологических процессов и производств

Главными задачами дисциплины являются:

- формирование инженерного мышления, позволяющего понимать важность обработки и систематизации научно-технической информации о современных средствах автоматизации на научно-техническое развитие промышленности, на рациональное использование сырьевых, энергетических и других видов ресурсов, на развитие современных систем и технологий;
- формирование навыков обработки опытных данных о характеристиках средств и систем автоматизации.

В результате освоения дисциплины студент должен:

- уметь применять системный подход при анализе проблемной ситуации;
- выполнять сбор и анализ научно-технической информации;
- проводить эксперимент по заданным методикам;

- уметь проводить математическую и статистическую обработку опытных данных о характеристиках средств и систем автоматизации;
- владеть современными методами и средствами обработки экспериментальных данных.

Методические указания предназначены для выполнения самостоятельной работы по дисциплине «Обработка экспериментальных данных с учетом требований ФГОС ВО для направления подготовки 15.03.04 — Автоматизация технологических процессов и производств. Они способствуют лучшему усвоению студентами теоретических положений и обеспечивает приобретение практических навыков по исследованию элементов и систем автоматического регулирования и управления.

1. ОБЩАЯ ХАРАКТЕРИСТИКА САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПРИ ИЗУЧЕНИИ ДИСЦИПЛИНЫ «ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Самостоятельная работа студентов (далее — СРС) является неотъемлемой составляющей образовательного процесса в Университете и является обязательной для каждого студента. Основная цель СРС — освоение в полном объеме образовательной программы и последовательное формирование компетенций эффективной самостоятельной профессиональной (практической и научно-теоретической) деятельности. Самостоятельная работа конкретна по своей предметной направленности и сопровождается непрерывным контролем и оценкой ее результатов.

Количество часов, отводимое на самостоятельную работу, определяется учебным планом направления подготовки 15.03.04.

Содержательно самостоятельная работа студентов определяется ФГОС ВО направления подготовки 15.03.04, программой и учебно-методическим комплексом дисциплины «Обработка экспериментальных данных».

Методика организации самостоятельной работы студентов зависит от структуры, характера и особенностей дисциплины «Обработка экспериментальных данных», объема часов на ее изучение, вида заданий для СРС, индивидуальных возможностей студентов и условий учебной деятельности.

Формы самостоятельной работы студентов определяются содержанием дисциплины «Обработка экспериментальных данных», степенью подготовленности студентов. Они могут быть тесно связаны с теоретическим курсом и иметь учебный или учебно-исследовательский характер. Форму самостоятельной работы студентов определяют кафедра ИСЭА при разработке программы дисциплины «Обработка экспериментальных данных».

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений студентов.

СРС, не предусмотренная образовательной программой, учебным планом и учебно-методическими материалами, раскрывающими и конкретизирующими их содержание, осуществляется студентами инициативно, с целью реализации собственных учебных и научных интересов.

В учебном процессе выделяют аудиторную и внеаудиторную самостоятельную работу.

Аудиторная самостоятельная работа по дисциплине «Обработка экспериментальных данных» выполняется на учебных занятиях (лекциях, практических, лабораторных занятиях и консультациях) под руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа студентов выполняется во внеаудиторное время по заданию и при методическом руководстве и контроле преподавателя, но без его непосредственного участия. СРС включает в себя:

- подготовку к аудиторным занятиям (лекционным и практическим) и выполнение соответствующих заданий;
- работу над отдельными темами учебных дисциплин (модулей) в соответствии с учебно-тематическими планами;

- выполнение контрольных работ;
- подготовку ко всем видам промежуточных и итоговых контрольных испытаний.

1.1. Подготовка к лекциям

Главное в период подготовки к лекционным занятиям — научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин.

Каждому студенту следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Слушание и запись лекций — сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Же-

лательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях.

Конспект лекций лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось просить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

1.2. Подготовка к лабораторным занятиям

Для того чтобы лабораторные занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по рассмотренному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться студентом на лабораторных занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит

несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

1.3. Самостоятельное изучение материала тем

Конспект — наиболее совершенная и наиболее сложная форма записи. Слово «конспект» происходит от латинского «conspicere», что означает «обзор, изложение». В правильно составленном конспекте обычно выделено самое основное в изучаемом тексте, сосредоточено внимание на наиболее существенном, в кратких и четких формулировках обобщены важные теоретические положения.

Конспект представляет собой относительно подробное, последовательное изложение содержания прочитанного. На первых порах целесообразно в записях ближе держаться тексту, прибегая зачастую к прямому цитированию автора. В дальнейшем, по мере выработки навыков конспектирования, записи будут носить более свободный и сжатый характер.

Конспект книги обычно ведется в тетради. В самом начале конспекта указывается фамилия автора, полное название произведения, издательство, год и место издания. При цитировании обязательная ссылка на страницу

книги. Если цитата взята из собрания сочинений, то необходимо указать соответствующий том. Следует помнить, что четкая ссылка на источник — неременное правило конспектирования. Если конспектируется статья, то указывается, где и когда она была напечатана.

Конспект подразделяется на части в соответствии с заранее продуманным планом. Пункты плана записываются в тексте или на полях конспекта. Писать его рекомендуется четко и разборчиво, так как небрежная запись с течением времени становится малопонятной для ее автора. Существует правило: конспект, составленный для себя, должен быть по возможности написан так, чтобы его легко прочитал и кто-либо другой.

Формы конспекта могут быть разными и зависят от его целевого назначения (изучение материала в целом или под определенным углом зрения, подготовка к докладу, выступлению на занятии и т.д.), а также от характера произведения (монография, статья, документ и т.п.). Если речь идет просто об изложении содержания работы, текст конспекта может быть сплошным, с выделением особо важных положений подчеркиванием или различными значками.

В случае, когда не ограничиваются переложением содержания, а фиксируют в конспекте и свои собственные суждения по данному вопросу или дополняют конспект соответствующими материалами их других источников, следует отводить место для такого рода записей. Рекомендуется разделить страницы тетради пополам по вертикали и в левой части вести конспект произведения, а в правой свои дополнительные записи, совмещая их по содержанию.

Конспектирование в большей мере, чем другие виды записей, помогает вырабатывать навыки правильного изложения в письменной форме важные теоретических и практических вопросов, умение четко их формулировать и ясно излагать своими словами.

Таким образом, составление конспекта требует вдумчивой работы, затраты времени и труда. Зато во время конспектирования приобретаются знания, создается фонд записей.

Конспект может быть текстуальным или тематическим. В текстуальном конспекте сохраняется логика и структура изучаемого произведения, а запись ведется в соответствии с расположением материала в книге. За основу тематического конспекта берется не план произведения, а содержание какой-либо темы или проблемы.

Текстуальный конспект желательно начинать после того, как вся книга прочитана и продумана, но это, к сожалению, не всегда возможно. В первую очередь необходимо составить план произведения письменно или мысленно, поскольку в соответствии с этим планом строится дальнейшая работа. Конспект включает в себя тезисы, которые составляют его основу. Но, в отличие от тезисов, конспект содержит краткую запись не только выводов, но и доказательств, вплоть до фактического материала. Иначе говоря, конспект — это расширенные тезисы, дополненные рассуждениями и доказательствами, мыслями и соображениями составителя записи.

Как правило, конспект включает в себя и выписки, но в него могут войти отдельные места, цитируемые дословно, а также факты, примеры, цифры, таблицы и схемы, взятые из книги. Следует помнить, что работа над конспектом только тогда будет творческой, когда она не ограничена текстом изучаемого произведения. Нужно дополнять конспект данными из других источников.

В конспекте необходимо выделять отдельные места текста в зависимости от их значимости. Можно пользоваться различными способами: подчеркиваниями, вопросительными и восклицательными знаками, репликами, краткими оценками, писать на полях своих конспектов слова: «важно», «очень важно», «верно», «характерно».

В конспект могут помещаться диаграммы, схемы, таблицы, которые придадут ему наглядность.

Составлению тематического конспекта предшествует тщательное изучение всей литературы, подобранной для раскрытия данной темы. Бывает, что какая-либо тема рассматривается в нескольких главах или в разных местах книги. А в конспекте весь материал, относящийся к теме, будет сосредоточен в одном месте. В плане конспекта рекомендуется делать пометки, к каким источникам (вплоть до страницы) придется обратиться для раскрытия вопросов. Тематический конспект составляется обычно для того, чтобы глубже изучить определенный вопрос, подготовиться к докладу, лекции или выступлению на семинарском занятии. Такой конспект по содержанию приближается к реферату, докладу по избранной теме, особенно если включает и собственный вклад в изучение проблемы.

2. СРЕДСТВА ОЦЕНИВАНИЯ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ ОБУЧАЮЩИХСЯ ПРИ ИЗУЧЕНИИ ДИСЦИПЛИНЫ «ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Вопросы для собеседования

1. Что понимается под понятием «данные»?
2. Что такое анализ данных?
3. Каковы этапы анализа данных?
4. Чем качественный этап анализа данных отличается от количественного?
5. Какие основные задачи решаются при качественном анализе данных?
6. Что понимается под понятием «вероятность»?
7. Чем относительная частота отличается от вероятности?
8. Каковы основные свойства распределения вероятности?
9. Что понимается под понятием «случайная переменная»?
10. Что такое доверительный интервал?

11. Из чего состоит алгоритм проверки гипотез?
12. Что такое нулевая гипотеза?
13. Что понимается под понятием «регрессия».
14. Что такое функция регрессии?
15. Каковы основные причины использования регрессионного анализа?
16. На основе какого метода вычисляются коэффициенты уравнения регрессии?
17. Из чего состоит проверка достоверности модели регрессии?
18. Какие критерии адекватности используются при оценке регрессии?
19. Что понимается под понятием «корреляция»?
20. Что такое матрица корреляции?
21. Как проверить результаты множественной линейной регрессии?
22. Что такое пошаговая регрессия?
23. Какие правила остановки в пошаговой регрессии?
24. Что такое дисперсионный анализ?
25. Каково назначение переменных-индикаторов?
26. Какая связь между регрессионным и дисперсионным анализом?
27. Каковы основные компоненты дисперсионного анализа?
28. В чем заключается метод наименьших квадратов?
29. В чем заключается метод наименьших модулей?
30. В чем заключается метод минимакса?
31. В чем заключается метод наибольшего правдоподобия?

Критерии оценивания компетенций

Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал монографической литературы, правильно обосно-

вызывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.

Оценка «**хорошо**» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.

Оценка «**удовлетворительно**» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.

Оценка «**неудовлетворительно**» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы.

3. ОТЧЕТНОСТЬ ПО ДИСЦИПЛИНЕ

В рамках рейтинговой системы успеваемость студентов по дисциплине оценивается в ходе текущего контроля успеваемости и промежуточной аттестации.

Максимально возможный балл за весь текущий контроль устанавливается равным 55. Текущее контрольное мероприятие считается сданным, если студент получил за него не менее 60% от установленного для этого контроля максимального балла. Рейтинговый балл, выставляемый студенту за текущее контрольное мероприятие, сданное студентом в установленные графиком контрольных мероприятий сроки, определяется следующим образом:

Уровень выполнения контрольного задания	Рейтинговый балл (в % от максимального балла за контрольное за-
-----------------------------------------	-----------------------------------------------------------------

	дание)
Отличный	100
Хороший	80
Удовлетворительный	60
Неудовлетворительный	0

Промежуточная аттестация

Процедура зачета с оценкой как отдельное контрольное мероприятие не проводится, оценивание знаний обучающегося происходит по результатам текущего контроля.

Шкала пересчета рейтингового балла по дисциплине в оценку по 5-балльной системе

Рейтинговый балл по дисциплине	Оценка по 5-балльной системе
88-100	Отлично
72-87	Хорошо
53-71	Удовлетворительно
менее 53	Неудовлетворительно

Максимальная сумма баллов по **курсовому проекту** устанавливается в **100** баллов и переводится в оценку по 5-балльной системе в соответствии со шкалой.

Шкала соответствия рейтингового балла 5-балльной системе

Рейтинговый балл	Оценка по 5-балльной системе
88-100	Отлично
72-87	Хорошо
53-71	Удовлетворительно
<53	Неудовлетворительно

Для студентов заочной формы обучения рейтинговая оценка знаний не предусмотрена

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Перечень основной литературы:

1. Тагиева Р.Ф. Обработка экспериментальных данных. В 2 частях. Ч.1 : учебное пособие / Тагиева Р.Ф., Титов А.Н.. — Казань : Казанский национальный исследовательский технологический университет, 2017. — 96 с. — ISBN 978-5-7882-2260-8, 978-5-7882-2261-5 (ч.1). — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/100572.html> (дата обращения: 13.01.2024). — Режим доступа: для авторизир. пользователей
2. Деревнин Д.А. Статистическая обработка экспериментальных данных : учебно-методическое пособие / Деревнин Д.А., Ситников В.Н.. — Тюмень : Тюменский индустриальный университет, 2019. — 50 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/101430.html> (дата обращения: 13.01.2024). — Режим доступа: для авторизир. пользователей.

Перечень дополнительной литературы:

1. Третьяк Л.Н. Основы теории и практики обработки экспериментальных данных : учебное пособие / Третьяк Л.Н., Воробьев А.Л.. — Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2015. — 216 с. — ISBN 978-5-7410-1282-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/61387.html> (дата обращения: 13.01.2024). — Режим доступа: для авторизир. пользователей.
2. Гребенникова И.В. Методы математической обработки экспериментальных данных : учебно-методическое пособие / Гребенникова И.В.. — Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2015. — 124 с. — ISBN 978-5-7996-1456-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/66551.html> (дата обращения: 13.01.2024). — Режим доступа: для авторизир. пользователей.

