МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Логистика и организация производства»

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА В БИЗНЕСЕ

Методические рекомендации к практическим занятиям для студентов специальности 1-25 80 01 «Экономика» очной и заочной форм обучения

Могилев 2020

УДК 005.511
ББК 65.23
К 60

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Логистика и организация производства» «02» апреля 2020 г., протокол № 16

Составитель ст. преподаватель Т. А. Бородич

Рецензент канд. экон. наук, доц. А. В. Александров

В методических рекомендациях представлены материалы к практическим занятиям для студентов специальности 1-25 80 01 «Экономика» очной и заочной форм обучения.

Учебно-методическое издание

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА В БИЗНЕСЕ

Ответственный за выпуск	М. Н. Гриневич
Корректор	И. В. Голубцова
Компьютерная верстка	Н. П. Полевничая

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 16 экз. Заказ №

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 07.03.2019.
Пр-т Мира, 43, 212022, Могилев.

© Белорусско-Российский
университет, 2020

Содержание

1 Методы количественного анализа данных: графики и таблицы 4
2 Методы количественного анализа данных: обобщающие показатели 9
3 Вероятность и распределения вероятностей 12
4 Нормальные, биномиальные и пуассоновские распределения 14
5 Многошаговые процедуры принятия решений в условиях неопределенности 17
6 Выборки и распределения выборок 20
7 Оценка доверительных интервалов 24
8 Проверка гипотез 27
Список литературы 32

1 Методы количественного анализа данных: графики и таблицы

Теоретические сведения

Применение графических методов способствует усилению наглядности анализируемых объектов (их структуры, организованных связей, взаимодействия с другими объектами и т. д.), вскрытию недостатков и нахождению правильных решений.

В аналитической работе широко применяются таблицы и графики.

В таблицах представлена соответственно обработанная информация, что облегчает ее чтение, усвоение и анализ. Содержание таблицы сходно с грамматическим предложением, изложенным в виде чисел. Исходя из этого, в таблице выделятся подлежащее – то, что характеризуется (размещается в левой части таблицы и характеризует содержание ее строк), и сказуемое – то, что говорится о подлежащем (размещается в графах/столбцах).

Графики более наглядны, чем таблицы, особенно это обстоятельство характерно при проведении экономического анализа, где графики занимают особое место, так как большие объемы информации в графическом виде позволяют быстрее осмыслить результаты, облегчить восприятие и выявить закономерности.

Одним из наиболее мощных аналитических методов исследования является разделение («разбиение») данных на группы для сравнения структуры получившихся подмножеств. Эти методы широко применяются как в разведочном анализе данных, так и при проверке гипотез и известны под разными названиями (классификация, группировка, категоризация, разбиение, расслоение и пр.).

Для количественного описания различий между группами наблюдений разработаны многочисленные вычислительные методы, основанные на группировке данных (например, дисперсионный анализ). Однако графические средства (такие как гистограммы, диаграммы рассеяния, линейные графики, поверхности, тернарные диаграммы рассеяния и пр.) дают особые преимущества и позволяют выявить закономерности, которые трудно поддаются количественному описанию и которые весьма сложно обнаружить с помощью вычислительных процедур (например, сложные взаимосвязи, исключения или аномалии).

При проведении группировок важное значение имеет правильное решение о том, на какое число групп следует подразделять совокупность. Если признак атрибутивный, то число групп, на которое следует подразделять совокупность, определяется числом качественных градаций этого признака. В случае если группировочный признак количественного порядка, непрерывный или дискретный, с большим размахом вариации, то число групп может быть определено по формуле Стерджесса

$$n = 1+3{,}322 \lg N, \qquad (1.1)$$

где n – число групп;

N – численность совокупности.

Для характеристики групп должны быть образованы интервалы, которые могут быть равные и неравные, открытые и закрытые. При равных интервалах их величина определяется по формуле

$$i = (X_{\max} - X_{\min}) / n = R / n , \qquad (1.2)$$

где n – число групп;

$X_{\max}$ – максимальное значение;

$X_{\min}$ – минимальное значение;

R – размах вариации.

При построении равных интервалов определяются нижняя и верхняя границы каждого из них, причем считается, что пределы наблюдения могут входить «включительно» или «исключительно». На практике применяются оба метода, но все же предпочтительнее принцип «исключительно».

Для графического изображения дискретного ряда применяют полигон распределения. Для его построения на оси абсцисс откладывают значения признака, а на оси ординат – частоту.

Для графического изображения интервального вариационного ряда применяется гистограмма. При ее построении на оси абсцисс откладываются интервалы ряда, высота которых равна частотам, отложенным на оси ординат. Над осью абсцисс строятся прямоугольники, площадь которых соответствует величинам произведений интервалов на их частоты.

В практике экономической работы возникает потребность в преобразовании рядов распределения в кумулятивные ряды, строящиеся по накопленным частотам (с их помощью можно определить структурные средние, прослеживать за процессом концентрации изучаемого явления). Используя накопленные частоты, строят график в виде кумуляты (кривой сумм). При графическом изображении кумуляты накопленные частоты наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Длина этих линий равна величине накопленных частот в конкретном интервале. Соединяя затем эти перпендикуляры, получаем ломаную линию.

Вопросы для обсуждения

1 Виды графических методов и особенности их применения.

2 Аналитические методы исследования.

Задачи для решения

Задача 1. По исходным данным о предприятиях, представленным в таблице, произвести структурную группировку 20 предприятий по объему выполненных работ, образовав три группы с равными интервалами (таблица 1.1).

Таблица 1.1 − Характеристика предприятий

Номер предприятия	Объем выполненных работ, д. е.	Затраты, д. е.	Номер предприятия	Объем выполненных работ, д. е.	Затраты, д. е.
1	18,6	3,2	11	17,9	2,2
2	17,5	2,4	12	11,8	3,9
3	10,6	3,2	13	12,3	4,1
4	17,8	2,2	14	15,7	1,5
5	10,4	3,2	15	18,1	2,5
6	18,5	3,0	16	9,8	3,2
7	16,4	2,1	17	13,9	4,7
8	17,5	2,4	18	12,6	4,5
9	15,5	1,5	19	9,6	3,0
10	17,0	2,7	20	8,6	2,8

Охарактеризовать каждую группу и совокупность в целом: числом предприятий; объемом затрат − всего и в среднем на одно предприятие. Результаты группировки представить в виде таблицы.

Задача 2. Пользуясь формулой Стерджесса, определить интервал группировки сотрудников предприятия по уровню доходов, если общая численность сотрудников составляет 140 чел., а минимальный и максимальный доходы равны 200 и 1 600 р. соответственно.

Задача 3. Имеются следующие данные о количестве филиалов каждого из двадцати банков в городе. Количество филиалов в городе у разных банков: 2, 4, 3, 5, 4, 4, 6, 5, 4, 3, 4, 3, 4, 5, 3, 4, 6, 3, 5, 4. Построить ряд распределения по имеющимся данным. Графически изобразить ряд распределений.

Задача 4. Имеются следующие данные о размере прибыли двадцати предприятий. Прибыль: 4,7; 4,3; 6,9; 5,6; 5,1; 8,1; 4,6; 5,7; 6,4; 5,9; 5,2; 6,2; 6,8; 7,2; 7,9; 5,8; 4,9; 7,6; 7,1; 6,9 тыс. р. Построить ряд распределения по имеющимся данным. Графически изобразить ряд распределений.

Задача 5. По нижеприведенным данным о квалификации рабочих цеха требуется: построить дискретный ряд распределения; дать графическое изображения ряда; вычислить показатели центра распределения, показатели вариации и формы распределения. Тарифные разряды 26 рабочих цеха: 5; 3; 6; 4; 2; 2; 3; 5; 4; 4; 5; 2; 3; 4; 3; 5; 2; 3; 6; 5; 4; 2; 4; 3, 5, 6.

Задача 6. В таблице 1.2 имеются данные о распределении предприятий двух регионов по численности работников.

Построить вторичную группировку данных о распределении предприятий, пересчитав данные региона 2 в соответствии с группировкой региона 1.

Таблица 1.2 − Исходные данные

Регион 1			Регион 2		
Группы предприятий по численности работников, чел.	Число предприятий, %	Численность работников, %	Группы предприятий по численности работников, чел.	Число предприятий, %	Численность работников, %
До 50	30	1	До 100	32	1
51…100	25	2	101…250	28	6
101 … 500	25	3	251 … 500	12	8
501 … 1000	9	9	501 … 1000	10	10
1001 … 2000	5	10	1001 … 2500	9	12
2001 … 5000	3	15	2501 … 7500	5	15
5001 … 8000	2	22	7501 … 10000	3	23
8001 и более	1	38	10001 и более	1	25
Итого	100	100	Итого	100	100

Задача 7. Имеются следующие данные о количестве заявок на автомобиль технической помощи по дням: 12; 8; 5; 14; 7; 2; 8; 10; 2; 6; 10; 8; 3; 13; 11; 8; 8; 12; 9; 8; 5; 14; 15; 10; 12; 6; 8; 2; 8; 7; 9; 7; 8; 4; 6; 13; 5; 3; 12; 2; 2; 7; 9; 8; 5; 8; 6; 13; 11; 6. Построить интервальный ряд распределения, определив величину интервала группировки по формуле Стерджесса.

Задача 8. Имеются следующие данные о возрастном составе работников отдела (лет): 24; 38; 20; 29; 26; 38; 34; 22; 28; 30; 22; 23; 30; 33; 27; 24; 30; 32; 28; 25; 29; 26; 31; 21; 29; 24; 32; 25; 29; 29, 32, 44.

Для анализа распределения работников отдела по возрасту требуется: построить интервальный ряд распределения (количество интервалов определить по формуле (1.2)); дать графическое изображение ряда; исчислить показатели центра распределения, показатели вариации и формы распределения. Сформулировать вывод.

Задача 9. По предприятию имеются следующие данные о проверке 100 партий детской обуви, передаваемых в торговую сеть (таблица 1.3).

Таблица 1.3 – Результаты проверки качества детской обуви

Число бракованных пар обуви	0	1	2	3
Число партий, содержащих данное число бракованных пар обуви	62	33	4	1

Дать графическое изображение представленных данных.

Задача 10. По цеху имеются следующие данные о распределении рабочих по стажу работы (таблица 1.4).

Таблица 1.4 – Исходные данные

Группы рабочих по стажу работы, лет	0…2	2…4	4…6	6…8	8…10	10…12	12…14
Число рабочих	7	9	12	24	17	9	5

Дать графическое изображение представленных данных.

Задача 11. Имеются следующие данные о величине межремонтного пробега автомобилей (таблица 1.5).

Таблица 1.5 – Исходные данные

Величина межремонтного пробега, тыс. км	90…100	100…120	120…140	140…160	160…180
Число автомобилей	14	66	100	30	16

Дать графическое изображение ряда в виде гистограммы и кумуляты.

2 Методы количественного анализа данных: обобщающие показатели

Теоретические сведения

Статистические показатели являются одной из важнейших категорий статистики. Они используются для описания исследуемых массовых явлений и процессов, выступают инструментом их познания.

По степени агрегированности показатели подразделяются на индивидуальные, характеризующие единичные процессы, и обобщающие, отражающие совокупность в целом или её части.

Абсолютные показатели характеризуют численность совокупности либо объём изучаемого явления в конкретных границах пространства и времени, т. е. отражают уровень развития явления, его размер.

Относительные показатели – это цифровые обобщающие показатели, они есть результат сопоставления двух статистических величин. По своей природе относительные величины производны от деления текущего (сравниваемого) абсолютного показателя на базисный показатель.

Относительная величина структуры (*ОВС*) характеризует структуру совокупности, определяет долю (удельный вес) части в общем объеме совокупности. ОВС рассчитывают как отношение объема части совокупности к абсолютной величине всей совокупности, определяя тем самым удельный вес части в общем объеме совокупности (%):

$$ОВС = \frac{m_i}{M} \cdot 100\ \%, \qquad (2.1)$$

где m_i – объем исследуемой части совокупности;

M – общий объем исследуемой совокупности.

Относительная величина координации *ОВК* характеризует соотношение между двумя частями исследуемой совокупности, одна из которых выступает как база сравнения (%):

$$ОВК = \frac{m_i}{m_б} \cdot 100\ \%, \qquad (2.2)$$

где m_i – одна из частей исследуемой совокупности;

$m_б$ – часть совокупности, которая является базой сравнения.

Относительная величина динамики *ОВД* характеризует изменение объема одного и того же явления во времени в зависимости от принятого базового уровня. *ОВД* рассчитывают как отношение уровня анализируемого явления или процесса в текущий момент времени к уровню этого явления или процесса за прошедший период времени. В результате получаем коэффициент роста, который выражается кратным отношением. При исчислении этой величины в про-

центах (результат умножается на 100) получаем темп роста.

Темпы роста можно просчитывать как с постоянным базовым уровнем (базисные темпы роста – $ОВД_б$), так и с переменным базовым уровнем (цепные темпы роста – $ОВД_ц$):

$$ОВД_б = \frac{P_т}{P_б} \cdot 100\ \%, \qquad (2.3)$$

где $P_т$ – уровень текущий;

$P_б$ – уровень базисный;

$$ОВД_ц = \frac{P_т}{P_{т-1}} \cdot 100\ \%, \qquad (2.4)$$

где $P_т$ – уровень текущий;

$P_{т-1}$ – уровень, предшествующий текущему.

В статистической практике относительные величины интенсивности применяются при исследовании степени объемности явления по отношению к объему среды, в которой происходит распространение этого явления. *ОВИ* здесь показывает, сколько единиц одной совокупности (числитель) приходится на одну, на десять, на сто единиц другой совокупности (знаменатель). *ОВИ* рассчитывается по формуле

$$ОВИ = \frac{A}{B_A} \cdot 100\ \%, \qquad (2.5)$$

где A – распространение явления;

B_A – среда распространения явления А.

Вопросы для обсуждения

1 Относительные величины интенсивности.
2 Относительные величины динамики.
3 Относительная величина структуры.

Задачи для решения

Задача 1. Производство электроэнергии в области составило 27,4 млрд кВтч при среднегодовой численности населения 9,3 млн чел. Определить относительную величину интенсивности, характеризующую производство электроэнергии на душу населения.

Задача 2. Производство автомобилей характеризуется следующими данными (таблица 2.1).

Таблица 2.1 – Исходные данные

Показатель	Произведено, тыс. шт.
Всего	1 195
В том числе:	
грузовые	173
легковые	1 022

Вычислить относительные показатели структуры и координации. Сформулировать выводы по результатам расчётов.

Задача 3. В таблице 2.2 имеются данные об урожайности пшеницы. Рассчитать относительные показатели сравнения.

Таблица 2.2 − Исходные данные

Страна	Беларусь	Россия	США	Германия
Урожайность пшеницы, ц/га	13,7	16,0	18,8	20,8

Задача 4. В таблице 2.3 имеются данные о внешнеторговом обороте условной страны за два года. Рассчитать: относительные величины структуры, характеризующие доли экспорта и импорта во внешнеторговом обороте страны; относительные величины координации, характеризующие соотношение экспорта и импорта.

Таблица 2.3− Исходные данные

Период	Внешнеторговый оборот, млрд долл.	В том числе	
		Экспорт	Импорт
Первый год			
I квартал	38,7	21,1	17,6
II квартал	39,9	20,4	19,5
III квартал	41,4	21,6	19,8
IV квартал	47,9	25,1	22,8
Второй год			
I квартал	37,7	19,4	18,3
II квартал	38,4	20,7	17,7
III квартал	32,5	18,8	13,7
IV квартал	29,7	20,3	9,4

3 Вероятность и распределения вероятностей

Теоретические сведения

Случайная переменная – это величина, которая может принимать любое из набора взаимоисключающих значений с определенной вероятностью.

Вероятность – это положительное число, которое находится в интервале между нулём и единицей.

Если вероятность равна нулю, то события быть не может.

Если вероятность равна единице, то событие обязательно должно произойти.

Распределение вероятности показывает вероятности всех возможных значений случайной переменной. Это теоретическое распределение, которое выражено математически и имеет среднее и дисперсию – аналоги среднего и дисперсии в эмпирическом распределении.

Распределение вероятностей какой-либо действительной случайной величины x, возможные значения x_1, x_2, ..., x_n, ... которой образуют конечную или бесконечную последовательность, задается указанием этих значений и соответствующих им вероятностей $P\{x = x_n\}$: p_1, p_2, ..., p_n, ... (p_n должны быть положительными и в сумме составлять единицу).

Распределения вероятностей указанного типа называются дискретными. Примерами дискретных распределений вероятностей служат:

– биномиальное распределение, определяемое вероятностями

$$P\{x = m\} = C_n^m p^m (1-p)^{n-m}, \ \ m = 0,1,...,n, \tag{3.1}$$

где $n > 0$ – целое, $0 < p < 1$;

– распределение Пуассона, определяемое вероятностями

$$P\{x = m\} = e^{-\lambda} \frac{\lambda^m}{m!}, \ \ m = 0,1,...,n, \tag{3.2}$$

где $\lambda > 0$.

Если существует такая функция $p(x)$, что вероятность попадания x в любой интервал (a, b) на прямой

$$P\{a < x < b\} = \int_a^b p(x)dx, \tag{3.3}$$

причем

$$p(x) \geq 0 = \int_{-\infty}^{\infty} p(x)dx = 1, \tag{3.4}$$

то распределение вероятностей величины x называют абсолютно непрерывным, а функция $p(x)$ носит название плотности вероятности.

В случае равномерного распределения на отрезке $[a, b]$

$$p(x) = \begin{cases} 0 & \text{при } x < a, \\ \dfrac{1}{b-a} & \text{при } x \in [a, b], \\ 0 & \text{при } x > b. \end{cases} \tag{3.5}$$

Важнейшее распределение вероятностей непрерывного типа – нормальное распределение с плотностью

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-a)^2}{2\sigma^2}}, \tag{3.6}$$

где $-\infty < a < \infty$ и $\sigma > 0$;

σ –дисперсия.

Вопросы для обсуждения

1 Понятие вероятности.

2 Виды распределения вероятности.

Задачи для решения

Задача 1. В партии, состоящей из 1 000 изделий, четыре изделия имеют дефекты. Для контроля отбираются 100 изделий. Найти вероятность того, что среди отобранных изделий не окажется бракованных.

Задача 2. Известно, что в среднем из 1 000 выданных кредитов примерно 12 не возвращаются в срок. В текущем году банк выдал 3 000 кредитов. Найти количество кредитов, которые не будут возвращены в срок.

Задача 3. Известно, что курс евро к рублю может возрасти с вероятностью 0,55, а курс доллара к рублю может возрасти с вероятностью 0,35. Вероятность того, что возрастут оба курса, составляет 0,3. Найти вероятность того, что курс евро или доллара по отношению к рублю возрастёт.

Задача 4. Известны вероятности дополнительной потребности фирмы в инженерах на предстоящие два года (таблица 3.1).

Таблица 3.1 – Исходные данные

Число инженеров, чел.	< 100	100…199	200…299	300…399	400…499	>=500
Вероятность	0,1	0,15	0,3	0,3	0,1	0,05

4 Нормальные, биномиальные и пуассоновские распределения

Теоретические сведения

Каждое распределение вероятности определяется некоторыми параметрами, которые служат обобщающими величинами (например, среднее, дисперсия), характеризующими данное распределение (т. е. их знание позволит подробно описать распределение).

С помощью соответствующей статистики можно произвести оценку этих параметров в выборке. В зависимости от того, является ли случайная переменная дискретной или непрерывной, распределение вероятности может быть либо дискретным, либо непрерывным.

Основные дискретные распределения и их параметры представлены в таблице 4.1.

Таблица 4.1 – Параметры дискретных распределений

Распределение	Вероятность	Параметры
Равномерное дискретное	$\frac{1}{N}$, $k = 1, 2, ..., N$	$N = 1, 2, ...$
Бернулли	$p_1 = p$, $p_0 = q$	$0 \le p \le 1$, $q = 1 - p$
Биномиальное	$C_n^k p^k q^{n-k}$, $k = 0,1,...,n$	$0 \le p \le 1$, $q = 1 - p$, $n = 1, 2, ...$
Пуассона	$\frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0,1,...$	$\lambda > 0$
Геометрическое	$q^{k-1}p$, $k = 1,2,...$	$0 \le p \le 1$, $q = 1 - p$
Отрицательно-биномиальное	$C_{k-1}^{r-1} p^r q^{k-r}$, $k = r, r+1,...$	$0 \le p \le 1$, $q = 1 - p$, $r = 1, 2, ...$

Нормально распределенные случайные величины широко распространены на практике, например, результаты измерения любой физической величины подчиняются нормальному закону распределения. Нормальным называется распределение вероятностей непрерывной случайной величины, которое описывается плотностью

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\overline{x})^2}{2\sigma^2}}, \qquad (4.1)$$

где σ – дисперсия;

$\overline{x}$ – среднее значение случайной величины x.

Вопросы для обсуждения

1 Биноминальное распределение.
2 Равномерное дискретное распределение.
3 Распределение Пуассона.

Задачи для решения

Задача 1. В лотерее на каждые 100 билетов приходится 15 выигрышей. Количество и размеры выигрышей представлены в таблице 4.1.

Таблица 4.1 – Исходные данные

Размер выигрыша, р.	2000	500	100
Количество билетов	1	4	10

Случайная величина X описывает размер выигрыша на один случайно выбранный билет. Составить ряд распределения случайной величины X. Построить кривую распределения вероятностей. Найти функцию распределения $p(x)$ и построить её график. Найти $P\{X < 500\}$, $P\{X < 2\ 100\}$, средний выигрыш на один билет и дисперсию выигрыша.

Задача 2. Построить ожидаемое распределение результатов испытаний, которое было бы получено для 256 абсолютно невежественных экзаменующихся, случайно угадывающих ответы на четыре вопроса с четырьмя возможными вариантами ответа на каждый вопрос (из которых один и только один верен).

Задача 3. На автовокзале время прибытия автобусов различных рейсов объявляет дежурный. Появление информации о различных рейсах происходит случайно и независимо друг от друга. В среднем на автовокзал прибывает пять рейсов каждые полчаса.

1 Составить ряд распределения числа сообщений о прибытии автобусов в течение получаса.

2 Найти числовые характеристики этого распределения.

3 Записать функцию распределения вероятностей и построить ее график.

4 Чему равна вероятность того, что в течение получаса прибудут не менее трех автобусов?

5 Чему равна вероятность того, что в течение четверти часа не прибудет ни один автобус?

Задача 4. В среднем в магазин заходят три человека в минуту. Найти вероятность того, что за 2 мин в магазин зайдет не более одного человека.

Задача 5. Автоматический токарный станок настроен на выпуск деталей со средним диаметром 2,00 см и со средним квадратическим отклонением 0,005 см. Действует нормальный закон распределения. Компания технического сервиса рекомендует остановить станок для технического обслуживания и корректировки в случае, если образцы деталей, которые он производит, имеют средний диаметр более 2,01 см либо менее 1,99 см.

1 Найти вероятность остановки станка, если он настроен по инструкции на 2,00 см.

2 Если станок начнет производить детали, которые в среднем имеют слишком большой диаметр, а именно 2,02 см, какова вероятность того, что станок будет продолжать работать?

5 Многошаговые процедуры принятия решений в условиях неопределенности

Теоретические сведения

В многошаговых процедурах принятия решений в условиях неопределенности проводится оценка стратегий.

Хотя ситуации принятия решений существенно отличаются друг от друга, тем не менее они имеют три общие составляющие:

1) множество вариантов решений (стратегий) для лица, принимающего решение;

2) множество возможных исходов и вероятностей этих исходов;

3) оценка результатов (обычно в денежной форме) при различных вариантах принятых решений и исходов.

В том случае, если эти составляющие описаны корректно, лицо, принимающее решение, может рассчитывать найти оптимальную стратегию с точки зрения некоторого критерия.

Процесс принятия решений характеризуется четырьмя основными свойствами.

1 Альтернативные планы действий. Прежде чем сделать окончательный выбор, лицо, принимающее решение, должно оценить несколько вариантов.

2 События или экономическое положение. Лицо, принимающее решение, должно перечислить возможные события и вычислить их вероятность.

3 Таблица выигрышей. Для того чтобы оценить план действий, лицо, принимающее решение, должно сопоставить каждому событию возможный выигрыш или проигрыш. В деловых приложениях выигрыш и проигрыш обычно выражаются в терминах прибыли или убытка, хотя существуют и другие формы, например, моральное удовлетворение или польза.

4 Критерий принятия решения. Лицо, принимающее решение, должно установить критерий выбора наилучшего плана действий. Для этого сначала требуется вычислить вероятность каждого события, используя статистические данные, экспертные оценки или знания о функции распределения. Выбор оптимального плана действий основывается на вероятностях событий и размерах упущенной выгоды при каждом варианте.

Ожидаемая прибыль (expected monetary value – *EMV*) для каждого варианта действий представляет собой сумму доходов xij, полученных при событии i, умноженных на вероятности *Pi*:

$$EMV(j) = \sum_{i=1}^{N} x_{ij} P_i, \tag{5.1}$$

где $EMV(j)$ – ожидаемая прибыль при варианте действий j;

x_{ij} – выигрыш, полученный при варианте действий j и событии i;

P_i – вероятность события i.

Дерево решений представляет собой графический инструмент, который помогает производить действия в многошаговых процедурах принятия решений в условиях неопределенности, а именно:

1) описание возможных стратегий игрока, принимающего решение;

2) описание неопределенных исходов (неизвестные стратегии второй стороны) и их вероятностей;

3) вычисление *EMV* по стратегиям первого игрока;

4) выбор стратегии с максимальным значением *EMV*.

Вопросы для обсуждения

1 Понятие неопределенности в принятии решений.

2 Критерии выбора оптимального решения.

Задачи для решения

Задача 1. Некая компания, специализирующаяся на программных разработках, планирует принять участие в тендере на получение некоторого госзаказа на разработку информационной системы. Тендер проводится закрытым способом, то есть предложения подаются участниками в запечатанных конвертах и неизвестны другим участникам. По оценкам компании, участие в тендере обойдется в 5000 р., а выполнение заказа в 95 000 р. Из опыта предыдущих тендеров известно, что с вероятностью 30 % конкуренции вообще не будет. Кроме того, известно, что цена подобного тендера имеет следующие условные вероятности (таблица 5.1).

Таблица 5.1 – Условные вероятности цены тендера

Цена тендера, р.	Вероятность
Менее 115000	0,2
От 115000 до 120000	0,4
От 120000 до 125000	0,3
Более 125000	0,1

Необходимо принять решение, участвовать ли в тендере, и если да, то с какой ценой. Необходимо выбрать решение, которое максимизирует ожидаемую прибыль.

Задача 2. Некоторая компания решает вопрос о представлении некоторого нового продукта на общенациональный рынок. Неопределенность заключается в том, как отреагирует рынок на этот новый продукт. Рассматривается вопрос об апробации нового продукта первоначально на некотором региональном рынке. Таким образом, первоначальное решение, которое необходимо принять компании, – это проводить ли первоначальный маркетинг продукта на регио-

нальном уровне. Компания предполагает, что выход на региональный уровень потребует затрат на 3 тыс. р., а выход на общенациональный рынок потребует вложения 90 тыс. р. Если не проводить первоначальных пробных продаж на региональном уровне, то решение о выходе на общенациональный рынок можно принять незамедлительно.

Компания рассматривает результаты продаж как успешные, средние или отрицательные в зависимости от объемов продаж. Для регионального уровня этим градациям соответствуют объемы в 200, 100 и 30 тыс. экземпляров, а для общенационального – 6000, 3000 и 900 тыс. экземпляров соответственно.

Исходя из данных по результатам региональных тестирований аналогичных видов продукции компания оценивает вероятности вышеуказанных трех исходов как 0,3, 0,6 и 0,1.

Кроме того, исследуя данные о соотношении результатов региональных продаж с последующими продажами на общенациональном рынке, компания сумела оценить следующие условные вероятности (таблица 5.2).

Таблица 5.2 – Условные вероятности продаж на общенациональном рынке

		Результаты на общенациональном рынке			
			Успешные	Средние	Отрицательные
0,3	Результаты продаж на региональном рынке	Успешные	0,8	0,15	0,05
0,6		Средние	0,3	0,5	0,2
0,1		Отрицательные	0,05	0,25	0,7

Кроме этого, известно, что каждая продажа приносит прибыль в 18 р. как на региональном рынке, так и на общенациональном.

Задача состоит в принятии обоснованной стратегии выхода (или не выхода) на рынок с новой товарной позицией.

6 Выборки и распределения выборок

Теоретические сведения

Под выборочным наблюдением понимается такое несплошное наблюдение, при котором статистическому обследованию подвергаются единицы изучаемой совокупности, отобранные случайным способом.

Цель выборочного наблюдения состоит в том, чтобы по характеристикам отобранной части единиц (выборка) судить о характеристиках всей совокупности (генеральная).

В статистике принято строго различать параметры и свойства генеральной совокупности и их оценки по выборочным данным. С этой целью принята следующая система обозначений: генеральные параметры обозначаются греческими буквами, выборочные показатели, которые рассматриваются как оценки генеральных параметров, – латинскими буквами (таблица 6.1).

Таблица 6.1 – Показатели выборочного наблюдения и генеральной совокупности

Показатель	Генеральный параметр	Выборочный показатель
Объем (число единиц) совокупности	N	n
Среднее значение признака	$\bar{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
Доля единиц, обладающих изучаемым признаком	$p = M / N$ M – число единиц, обладающих определенным признаком в ГС	$W = m / n$ m – число единиц, обладающих определенным признаком в ВС
Доля единиц, не обладающих изучаемым признаком	q	$1 - w$
Дисперсия количественного признака	$\sigma^2_{\bar{x}} = \frac{\sum (x_i - \bar{x})^2}{N}$	$\sigma^2_{\tilde{x}} = \frac{\sum (x_i - \tilde{x})^2}{n}$
Дисперсия доли	$\sigma^2_p = p(1-p)$	$\sigma^2_w = w(1-w)$

В связи с тем, что существуют различные методы, виды и способы отбора единиц из генеральной совокупности, формулы для расчета средней ошибки выборки также будут различаться (таблица 6.2).

Формулы для нахождения необходимой численности выборки при разных способах отбора представлены в таблице 6.3.

В статистических исследованиях с помощью формулы предельной ошибки можно решать ряд задач.

1 Определять возможные пределы нахождения характеристики генеральной совокупности на основе данных выборки.

Доверительные интервалы для генеральной средней можно установить на основе соотношений

$$\tilde{x} - \Delta_x \le \bar{x} \le \tilde{x} + \Delta_x, \qquad (6.1)$$

где $\tilde{x}$ и $\bar{x}$ – генеральная и выборочная средние соответственно;

Δ_x – предельная ошибка выборочной средней.

Таблица 6.2 – Формулы для расчета характеристик

Способ отбора	Оцениваемый параметр	Повторный отбор	Бесповторный отбор
Собственно случайный и механический	Средняя	$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2_{\tilde{x}}}{n}}$	$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2_{\tilde{x}}}{n}(1-\frac{n}{N})}$
	Доля	$\mu_w = \sqrt{\frac{w(1-w)}{n}}$	$\mu_w = \sqrt{\frac{w(1-w)}{n}(1-\frac{n}{N})}$
Типический	Средняя	$\mu_{\tilde{x}} = \sqrt{\frac{\overline{\sigma^2_i}}{n}}$	$\mu_{\tilde{x}} = \sqrt{\frac{\overline{\sigma^2_i}}{n}(1-\frac{n}{N})}$
	Доля	$\mu_w = \sqrt{\frac{\overline{w_i(1-w_i)}}{n}}$	$\mu_w = \sqrt{\frac{\overline{w_i(1-w_i)}}{n}(1-\frac{n}{N})}$
Серийный	Средняя	$\mu_{\tilde{x}} = \sqrt{\frac{\delta^2_{\tilde{x}}}{s}}$	$\mu_{\tilde{x}} = \sqrt{\frac{\delta^2_{\tilde{x}}}{s}(\frac{S-s}{S-1})}$
	Доля	$\mu_w = \sqrt{\frac{\delta^2_w}{s}}$	$\mu_w = \sqrt{\frac{\delta^2_w}{s}(\frac{S-s}{S-1})}$

Таблица 6.3 – Определение численности выборки

Способ отбора	Оцениваемый параметр	Повторный отбор	Бесповторный отбор
Собственно случайный и механический	Средняя	$n = \frac{t^2\sigma^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}$	$n = \frac{t^2\sigma^2_{\tilde{x}}N}{N\Delta^2_{\tilde{x}} + t^2\sigma^2_{\tilde{x}}}$
	Доля	$n = \frac{t^2pq}{\Delta^2_w}$	$n = \frac{Nt^2pq}{N\Delta^2_w + t^2pq}$
Типический	Средняя	$n = \frac{t^2\overline{\sigma^2_i}}{\Delta^2_{\tilde{x}}}$	$n = \frac{Nt^2\overline{\sigma^2_i}}{N\Delta^2_{\tilde{x}} + t^2\overline{\sigma^2_i}}$
	Доля	$n = \frac{t^2\overline{pq}}{\Delta^2_w}$	$n = \frac{Nt^2\overline{pq}}{N\Delta^2_w + t^2\overline{pq}}$
Серийный	Средняя	$s = \frac{t^2\delta^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}$	$s = \frac{St^2\delta^2_{\tilde{x}}}{(S-1)\Delta^2_{\tilde{x}} + t^2\delta^2_{\tilde{x}}}$
	Доля	$s = \frac{t^2\delta^2_w}{\Delta^2_w}$	$s = \frac{St^2\delta^2_w}{(S-1)\Delta^2_w + t^2\delta^2_w}$

Доверительные интервалы для генеральной доли устанавливаются на основе соотношений

$$p = w \pm \Delta_p, \;\; w - \Delta_p \le p \le w + \Delta_p. \qquad (6.2)$$

2 Определять доверительную вероятность, которая означает, что характеристика генеральной совокупности отличается от выборочной на заданную величину. Доверительная вероятность является функцией от t, где

$$t = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}}. \qquad (6.3)$$

Доверительная вероятность по величине t определяется по специальной таблице.

3 Определять необходимый объем выборки с помощью допустимой величины ошибки:

$$\Delta_p = t\mu_p. \qquad (6.4)$$

Вопросы для обсуждения

1 Определение размера выборки.
2 Показатели выборочного наблюдения и генеральной совокупности.
3 Методы, виды и способы отбора единиц из генеральной совокупности.

Задачи для решения

Задача 1. По данным пробного обследования среднее квадратическое отклонение веса нарезных батонов составило 15,7 г. Установить оптимальный объем выборки из партии нарезных батонов (2 000 шт.), чтобы с вероятностью 0,997 предельная ошибка выборки не превысила 3 % веса 500-граммового батона.

Задача 2. Для определения срока службы станков было проведено 20-процентное выборочное обследование по методу случайного бесповторного отбора. Полученные данные занесены в таблицу 6.4.

Таблица 6.4 − Результаты выборки

Срок службы станков, лет	До 3	3…5	5…7	7…9	9…11	Св. 11
Число станков, шт.	98	101	118	105	99	90

На основе этих данных определить:
− средний срок службы станка в выборочной совокупности;
− дисперсию;
− коэффициент вариации;

− с вероятностью 0,997 (t = 3) предельную ошибку выборки и возможные пределы среднего срока службы станков по генеральной совокупности;

− с вероятностью 0,954 (t = 2) ошибку выборочной доли и границы удельного срока службы станков в генеральной совокупности.

Задача 3. В целях контроля качества комплектующих из партии изделий, упакованных в 50 ящиков по 20 изделий в каждом, была произведена 10-процентная серийная выборка. По попавшим в выборку ящикам среднее отклонение параметров изделия от нормы соответственно составило 9; 11; 12; 8 и 14 мм. С вероятностью 0,954 определить среднее отклонение параметров по всей партии в целом.

Задача 4. В таблице 6.5 имеются данные о результате выборочного обследования незанятого, ищущего работу, населения, осуществленного на основе собственно случайной повторной выборки.

Таблица 6.5 − Результаты выборочного обследования незанятого населения

Возраст, лет	До 25	25 … 35	35 … 45	45 … 55	55 и более
Численность лиц данного возраста	18	38	70	45	26

С вероятностью 0,954 определить границы: среднего возраста незанятого населения; доли (удельного веса) лиц моложе 25 лет в общей численности незанятого населения.

Задача 5. При изучении покупательского спроса произведено 5-процентное выборочное бесповторное обследование розничной продажи ноутбуков при собственно случайном способе отбора. Полученное распределение реализованных ноутбуков по уровню цен представлено в таблице 6.6.

Таблица 6.6 – Результаты выборки

Распределение по уровню цен, р.	До 1000	1000…1500	1500…2000	2000…2500	Св. 2500
Количество ноутбуков, шт.	142	232	197	147	117

На основе приведенных данных вычислить: среднюю цену ноутбука (способом момента); среднее квадратическое отклонение; коэффициент вариации.

По результатам выборочного наблюдения определить:

– с вероятностью 0,954 возможные значения доли продажи ноутбуков по цене свыше 1500 р. в общем объеме продажи ноутбуков;

– с вероятностью 0,997 (t = 3) возможные границы средней цены ноутбука.

7 Оценка доверительных интервалов

Теоретические сведения

Доверительный интервал – это интервал, построенный с помощью случайной выборки из распределения с неизвестным параметром, такой, что он содержит данный параметр с заданной вероятностью.

Модель адаптивного прогнозирования будет включать следующие этапы.

Этап 1. Для исследования динамики развития воспользуемся моделью Брауна. Расчетное значение в момент времени t получается по формуле

$$Y_p(t) = a_0(t-1) + a_1(t-1)k(t = 1,\ 2, ..., N), \tag{7.1}$$

где k – количество шагов прогнозирования (обычно $k = 1$).

Это значение сравнивается с фактическим уровнем и полученная ошибка прогноза $E(t) = Y(t) - Y_p(t)$ используется для корректировки модели.

Для расчета первоначальных значений a0 и a1 используются данные по первым пяти наблюдениям. Расчет ведется по формулам

$$a_1 = \Sigma((t - t_{cp})(Y_t - Y_{cp})) / \Sigma(t - t_{cp})^2; \tag{7.2}$$

$$a_0 = Y_{cp} - a_1 t_{cp}. \tag{7.3}$$

Корректировка параметров на каждом шаге осуществляется по формулам

$$a_0(t) = a_0(t-1) + a_1(t-1) + E(t)(1-\beta^2); \tag{7.4}$$

$$a_1(t) = a_1(t-1) + E(t)(1-\beta)^2, \tag{7.5}$$

где β – коэффициент дисконтирования данных, отражающий большую степень доверия к более поздним данным. Его значение должно быть в интервале от 0 до 1.

Такой процесс модификации модели в зависимости от ее текущих прогнозных качеств обеспечивает адаптацию к новым закономерностям развития. Для прогнозирования используется модель, полученная на последнем шаге (при $t = N$).

Этап 2. Качество модели определяется ее адекватностью исследуемому процессу и точностью. Адекватность характеризуется наличием определенных статистических свойств, а точность – степенью близости к фактическим данным. Модель считается адекватной, если ряд остатков обладает свойствами: независимости уровней; их случайности; соответствия нормальному закону распределения.

При проверке *независимости* (отсутствии автокорреляции) определяется отсутствие в ряду остатков систематической составляющей. Это проверяется с помощью *d*-критерия Дарбина–Уотсона, в соответствии с которым вычисляется *коэффициент d*:

$$d = \frac{\sum_{t=2}^{N} (e_{(t)} - e_{(t-1)})^2}{\sum_{t=1}^{N} e_{(t)}^{\ 2}}. \qquad (7.6)$$

Если $0 \le d$ (или d') $\le d1$ – *модель неадекватна* (уровни ряда остатков сильно автокоррелированы);

$d2 \le d$ (или d') ≤ 2 – *модель адекватна;*

$d1 \le d$ (или d') $\le d2$ – однозначного вывода сделать нельзя и необходимо применять другие критерии (на основе коэффициента автокорреляции $r(1)$).

Для проверки *случайности* уровней ряда использован критерий поворотных точек. В соответствии с этим критерием каждый уровень ряда сравнивается с двумя соединенными с ними. Если он больше или меньше их, то эта точка считается поворотной. Далее подсчитывается сумма поворотных точек p. В случайном ряду чисел должно выполняться строгое неравенство

$$p > [(2N-1)/3 - 2\sqrt{(16N-29)/90}]. \qquad (7.7)$$

Квадратные скобки означают целую часть числа.

Соответствие ряда остатков нормальному закону распределения важно с точки зрения правомерности построения доверительных интервалов прогноза. В этой связи определяется близость к соответствующим параметрам нормального закона распределения коэффициентов асимметрии – A_c и эксцесса $Э_к$ наблюдений около модели:

$$A_c = \sum_{t=1}^{N} e_{(t)}^{\ 3} / N / \sqrt{(1/N\sum_{t=1}^{N} e_{(t)}^{\ 2})^3}\ ; \qquad (7.8)$$

$$Э_к = \sum_{t=1}^{N} e_{(t)}^{\ 4} / N / (1/N\sum_{t=1}^{N} e_{(t)}^{\ 2})^2 - 3\,. \qquad (7.9)$$

Если эти коэффициенты приблизительно равны нулю, то ряд остатков распределен в соответствии с нормальным законом. Для оценки степени их близости к нулю вычисляются дисперсии:

$$S_a = 6(N-2)/(N+1)/(N+3); \qquad (7.10)$$

$$S_э = 24N(N-2)(N-3)/(N+1)/(N+3)/(N+5). \qquad (7.11)$$

Если $|Э_к| \le 1{,}5\ S_э$, $|A_c| \le 1{,}5\ S_a$, то распределение ряда остатков не противоречит нормальному закону распределения. Если $|Э_к| \ge 2S_э$, $|A_c| \ge 2\ S_a$, то ряд остатков не соответствует нормальному закону распределения. В случае попадания в зону неопределенности (между полутора и двумя СКО) используются другие критерии, в частности RS-критерий:

$$RS = \left(E_{\max} - E_{\min}\right) / S, \tag{7.12}$$

где $E_{\max}$ – максимальный уровень ряда остатков;

$E_{\min}$ – минимальный уровень ряда остатков;

S – среднее квадратическое отклонение.

Если значение этого критерия попадает между табулированными границами с заданным уровнем вероятности, то гипотеза о нормальном распределении ряда остатков принимается.

В статистическом анализе известно большое число характеристик точности. Наиболее часто используются:

– максимальная по абсолютной величине ошибка

$$E_{\max} = \max |e(t)|; \tag{7.13}$$

– относительная максимальная ошибка

$$E_{отн} = E_{\max} / Y_{ср} 100\ \%. \tag{7.14}$$

Этап 3. В моделях скользящего среднего (Брауна, Хольта) прогнозные точечные оценки $Y_p(t,k)$ уровня ряда $Y(t+k)$, вычисляются в момент времени t на k шагов вперед путем подстановки в нее значения k:

$$Y_p(t,k) = A_0(t) + A_1(t) \cdot k, \tag{7.15}$$

где $A_0(t)$ – оценка текущего (t-го) уровня;

$A_1(t)$ – оценка текущего прироста.

Границы доверительного интервала определяются на основе точечной оценки путем вычитания из нее и сложения с ней величины $U(k)$:

$$U(k) = S_e \cdot \sqrt{1 + C(k)}. \tag{7.16}$$

Величина $C(k)$ рассчитывается по формуле

$$C(k) = \alpha\left(1,25 + \alpha\, k\right). \tag{7.17}$$

Вопросы для обсуждения

1 Понятие доверительного интервала.

2 Методы определения доверительного интервала.

Задание

1 Построить адаптивную модель Брауна (используя данные таблицы 7.1). Провести прогноз на один и два периода вперед.

2 Оценить качество построенной модели.

3 Определить доверительный интервал для прогнозных значений ($\beta = 0{,}7$).

Таблица 7.1 – Исходные данные

t	1	2	3	4	5	6	7	8	9	10	11
$Y(t)$	16	19	20	25	30	29	35	42	50	65	75

8 Проверка гипотез

Теоретические сведения

Проверка гипотезы о нормальном законе распределения проводится с помощью критерия согласия Пирсона. Критерий основан на сравнении эмпирических и теоретических частот. Эмпирические частоты получают в результате эксперимента, а теоретические частоты рассчитывают по формулам (таблица 8.1). Предварительно необходимо найти точечные оценки параметров предполагаемого распределения по выборке, а затем использовать их для вычисления теоретических частот. Критерий имеет χ^2-распределение $ck = m - r - 1$ степенями свободы, где m – число интервалов вариационного ряда; r – число параметров теоретического распределения, вычисленных по экспериментальным данным. Малочисленные эмпирические частоты ($ni_{эмп} < 5$) следует объединить с соседним интервалом, в этом случае и соответствующие им теоретические частоты также складываются. Если производилось объединение частот, то при определении числа степеней свободы следует в качестве m принять число интервалов, оставшихся после объединения частот.

Таблица 8.1 – Определение теоретических частот основных распределений

Распределение	Параметры распределения	Теоретические частоты	Критерий
Нормальное	$\mu = \overline{x_в}$ $\sigma = \sigma_в$	$n_{iтеор} = \frac{nh}{\sigma} \cdot \varphi(u_i)$, $u_i = \frac{x_i - \mu}{\sigma}$, $\varphi(u_i)$	$\chi^2 = \sum_{i=1}^{m} \frac{(n_{i\,эксп} - n_{i\,теор})^2}{n_{i\,теор}}$ Критическая область – правосторонняя $\chi^2_{кр}(\alpha; k)$, $k = m - r - 1$
Показательное	$\lambda = \frac{1}{\overline{x_в}}$	$n_{i\,теор} = n\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right)$	
Равномерное	$a = \overline{x_в} - \sigma_в\sqrt{3}$ $b = \overline{x_в} + \sigma_в\sqrt{3}$	$n_{i\,теор} = \frac{nh}{b - a}$	

Проверка гипотезы о наличии тренда.

Метод проверки разностей средних уровней. Реализация этого метода состоит из четырех этапов.

На первом этапе исходный временной ряд $y_1, y_2, y_3, \ldots, y_n$ разбивается на две примерно равные по числу уровней части: в первой части n_1 первых уровней исходного ряда, во второй – из остальных уровней ($n_1 + n_2 = n$).

На втором этапе для каждой из этих частей вычисляются средние значения и дисперсии:

$$\overline{y_1} = \frac{\sum_{t=1}^{n_1} y_t}{n_1}; \qquad \sigma_1{}^2 = \frac{\sum_{t=1}^{n_1} (y_t - \overline{y_1})^2}{n_1 - 1}; \tag{8.1}$$

$$\overline{y_2} = \frac{\sum_{t=n_1+1}^{n} y_t}{n_2}; \qquad \sigma_2^2 = \frac{\sum_{t=n_1+1}^{n} (y_t - \overline{y_2})^2}{n_2 - 1}. \tag{8.2}$$

Третий этап заключается в проверке равенства (однородности) дисперсий обеих частей ряда с помощью F-критерия Фишера, которая основана на сравнении расчетного значения этого критерия:

$$F = \begin{cases} \sigma_1^2 / \sigma_2^2, & \text{если } \sigma_1^2 > \sigma_2^2; \\ \sigma_2^2 / \sigma_1^2, & \text{если } \sigma_1^2 < \sigma_2^2 \end{cases} \tag{8.3}$$

с табличным (критическим) значением критерия Фишера $F\alpha$ с заданным уровнем значимости (уровнем ошибки) α. В качестве α чаще всего берут значения 0,1 (10-процентная ошибка), 0,05 (5-процентная ошибка), 0,01 (1-процентная ошибка). Величина 1- α называется доверительной вероятностью.

Если расчетное значение F меньше табличного $F\alpha$, то гипотеза о равенстве дисперсий принимается и переходят к четвертому этапу. Если F больше или равно $F\alpha$, гипотеза о равенстве дисперсий отклоняется и делается вывод, что данный метод для определения наличия тренда ответа не дает.

На четвертом этапе проверяется гипотеза об отсутствии тренда с использованием критерия Стьюдента. Для этого определяется расчетное значение критерия Стьюдента по формуле (гипотеза о равенстве математических ожиданий двух выборок)

$$t = \frac{\left|\overline{y_1} - \overline{y_2}\right|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \tag{8.4}$$

$$\sigma = \sqrt{\frac{(n_1 - 1)\cdot\sigma_1^2 + (n_2 - 1)\cdot\sigma_2^2}{n_1 + n_2 - 2}}. \tag{8.5}$$

Если расчетное значение t меньше табличного значения статистики Стьюдента $t\alpha$ с заданным уровнем значимости α, гипотеза принимается, т. е. тренда нет, в противном случае тренд есть. Заметим, что в данном случае табличное значение $t\alpha$ берется для числа степеней свободы, равного $n_1 + n_2 - 2$, при этом

данный метод применим только для рядов с монотонной тенденцией.

Метод Фостера–Стюарта. Этот метод обладает большими возможностями и дает более надежные результаты по сравнению с предыдущим. Кроме тренда самого ряда (как говорят, тренда в среднем), он позволяет установить наличие тренда дисперсии временного ряда: если тренда дисперсии нет, то разброс уровней ряда постоянен, если дисперсия увеличивается, то ряд «раскачивается», и т. д.

Реализация метода также содержит четыре этапа.

На первом этапе производится сравнение каждого уровня исходного временного ряда, начиная со второго уровня, со всеми предыдущими, при этом определяются две числовые последовательности:

$$k_t = \begin{cases} 1, \text{ если } y_t \text{ больше всех предыдущих уровней;} \\ 0 - \text{ в противном случае;} \end{cases} \tag{8.6}$$

$$l_t = \begin{cases} 1, \text{ если } y_t \text{ меньше всех предыдущих уровней;} \\ 0 - \text{ в противном случае;} \end{cases} \tag{8.7}$$

$$t = 1, 2, 3, \ldots, n.$$

На втором этапе вычисляются величины s и d:

$$s = \sum_{t=2}^{n} (k_t + l_t); \tag{8.8}$$

$$d = \sum_{t=2}^{n} (k_t - l_t). \tag{8.9}$$

Нетрудно заметить, что величина s, характеризующая изменение временного ряда, принимает значения от 0 (все уровни ряда равны между собой) до $n - 1$ (ряд монотонный). Величина d характеризует изменение дисперсии уровней временного ряда и изменяется от минус $(n - 1)$ (ряд монотонно убывает) до $(n - 1)$ (ряд монотонно возрастает).

Третий этап заключается в проверке гипотез, можно ли считать случайными:

1) отклонение величины s от величины μ – математического ожидания величины s для ряда, в котором уровни расположены случайным образом;

2) отклонение величины d от нуля. Эта проверка проводится с использованием расчетных значений t-критерия Стьюдента для средней и для дисперсии:

$$t_s = \frac{|s - \mu|}{\sigma_1}; \tag{8.10}$$

$$t_d = \frac{|d - 0|}{\sigma_2}, \tag{8.11}$$

где μ – математическое ожидание величины *s,* определенной для ряда, в котором уровни расположены случайным образом;

σ_1 – среднеквадратическое отклонение для величины *s*;

σ_2 – среднеквадратическое отклонение для величины *d.*

На четвертом этапе расчетные значения *ts* и *td* сравниваются с табличным значением *t*-критерия Стьюдента с заданным уровнем значимости *tα.* Если расчетное значение меньше табличного, то гипотеза об отсутствии соответствующего тренда принимается; в противном случае тренд есть. Например, если *ts* больше табличного значения *tα,* а *td* меньше *tα,* то для данного временного ряда имеется тренд в среднем, а тренда дисперсии уровней ряда нет.

Критерий серий, основанный на медиане выборки, реализуется в виде следующей последовательности шагов.

1 Из исходного ряда y_t длиной *n* образуется ранжированный (вариационный) ряд y_t ($y|_n$, $y|_{n-1}$, ... , $y|_1$) , где $y|_1$ – наименьшее значение ряда y_t .

2 Определяется медиана этого вариационного ряда *Me*. В случае нечетного значения *n* ($n = 2m + 1$) $Me = y|_{m+1}$, в противном случае $Me = (y|_m + y|_{m+1}) / 2$.

3 Образуется последовательность δ_i из плюсов и минусов по следующему правилу:

$$\delta_i = \begin{cases} +, \text{ если } y_t > Me, \ t = 1, 2, ..., n; \\ -, \text{ если } y_t < Me, \ t = 1, 2, ..., n. \end{cases} \tag{8.12}$$

Если значение y_t равно медиане, то это значение пропускается.

4 Подсчитывается $\vartheta(n)$ – число серий в совокупности δ_i, где под серией понимается последовательность подряд идущих плюсов или минусов. Один плюс или один минус тоже будет считаться серией.

Определяется $\tau\max(n)$ – протяженность самой длинной серии.

5 Проверка гипотезы основывается на том, что при условии случайности ряда (при отсутствии систематической составляющей) протяженность самой длинной серии не должна быть слишком большой, а общее число серий – слишком маленьким.

Поэтому для того, чтобы не была отвергнута гипотеза о случайности исходного ряда (об отсутствии систематической составляющей), должны выполняться следующие неравенства (для 5-процентного уровня значимости):

$$\begin{cases} \tau_{\max}(n) < [3,3(\lg n + 1)]; \\ \nu(n) > [\frac{1}{2}(n + 1 - 1,96\sqrt{n-1})]. \end{cases} \tag{8.13}$$

Если хотя бы одно из неравенств нарушается, то гипотеза об отсутствии тренда отвергается.

Квадратные скобки в правой части неравенства означают целую часть числа. Напомним, что целая часть числа А – это целое число, ближайшее к А и непревосходящее его.

Вопросы для обсуждения

1 Критерий согласия Пирсона.
2 Методы проверки гипотезы о наличии тренда.

Задачи для решения

Задача 1. В таблице 8.1 приведены годовые данные об объемах продаж некоторого товара.

Таблица 8.1 – Исходные данные

Номер года	1	2	3	4	5	6	7	8	9	10
Продажи	4400	5200	4950	5700	6000	7100	6900	7200	8200	7950

Продолжение таблицы 8.1

Номер года	11	12	13	14	15	16	17	18	19	20
Продажи	8100	8500	8400	7300	7750	8000	7950	8400	9050	9450

Необходимо:

1) определить абсолютные приросты, темпы роста, цепные и базисные, абсолютные значения одного процента прироста по годам.

Абсолютное значение 1 % прироста определяется по формуле

$$\alpha = \frac{\Delta y}{T_{пр}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_i} \cdot 100} = 0,01 \cdot y_{i-1}; \qquad (8.14)$$

2) определить средние величины уровня ряда, абсолютного прироста, темпов роста и прироста.

Средний абсолютный прирост определяется по формуле

$$\overline{\Delta}_y = \frac{\sum(\Delta y_i^ц)}{n} = \frac{y_n - y_0}{n}. \qquad (8.15)$$

Средний темп роста $\overline{T_p} = \sqrt{\frac{y_n}{y_0}}$;

3) определить среднеквадратическое отклонение и дисперсию;

4) дать прогноз на 21-й год на основании среднего темпа роста;

5) на основании данных задачи проверить гипотезу о наличии тренда во временном ряду:

а) методом проверки разности средних уровней;

б) методом Фостера–Стюарта;

в) используя критерий серий, основанный на медиане выборки.

Задача 2. В таблице 8.2 представлены изменения выработки на одного основного рабочего. Проверить гипотезу, что выработка рабочих цеха распределена по нормальному закону.

Таблица 8.2 – Исходные данные

Выработка, %	94…100	100…106	106…112	112…118	118…124	124…130	130…136	136…142
Число рабочих	3	7	11	20	28	19	10	2

Список литературы

1 **Бродецкий, Г. Л.** Экономико-математические методы и модели в логистике: процедуры оптимизации: учебник для студентов вузов / Г. Л. Бродецкий, Д. А. Гусев. – 2-е изд., стер. – Москва : Академия, 2014. – 288 с.

2 **Ефимова, М. Р.** Общая теория статистики : учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. − Москва : ИНФРА-М, 2013. – 416 с.

3 **Захаренков, С. Н.** Статистика: учебное пособие для вузов / С. Н. Захаренков, В. А. Тарловская. – Минск : Современная школа, 2010. – 272 с.

4 **Иода, Е. В.** Статистика : учебное пособие / Е. В. Иода. − Москва : Вузовский учебник; ИНФРА-М, 2012. – 303 с.

5 **Колесникова, И. В.** Статистика. Практикум : учебное пособие для вузов / И. В. Колесникова, Г. В. Круглякова. − Минск : Вышэйшая школа, 2011. – 285 с.

6 **Косоруков, О. А.** Методы количественного анализа в бизнесе: учебник / О. А. Косоруков. – Москва : ИНФРА-М, 2005. – 367 с.

7 **Мхитарян, В. С.** Анализ данных в MS Excel : учебное пособие / В. С. Мхитарян, В. Ф. Шишов, А. Ю. Козлов. – Москва : КУРС, 2019. – 368 с.