

**Часть I. Комплексное использование
статистических методов
при проведении анализа данных**

Раздел I

Методические рекомендации к выполнению расчётов

Практическая часть включает три *типовых* расчетных задания.

Задание 1. Исследование структуры совокупности.

Задание 2. Выявление наличия корреляционной связи между признаками, установление направления связи и оценка ее тесноты.

Задание 3. Применение метода выборочных наблюдений.

Методика комплексного применения статистических методов при выполнении расчетных заданий излагается в данном разделе с использованием демонстрационного примера.

Демонстрационный пример

При проведении статистического наблюдения за деятельностью коммерческих банков одного из регионов РФ за исследуемый период получены выборочные данные об объеме кредитных вложений и сумме прибыли по 30-ти банкам (выборка 20%-ная, механическая).

В проводимом исследовании эти банки выступают как *единицы выборочной совокупности*. *Генеральную совокупность* образуют *все коммерческие банки региона*. Анализируемыми признаками изучаемых единиц совокупности являются *Объем кредитных вложений* и *Сумма прибыли* банка.

Выборочные данные представлены в табл.1.

Исходные данные

Номер банка п/п	Объем кредитных вложений, млн руб.	Сумма прибыли, млн руб.	Номер банка п/п	Объем кредитных вложений, млн руб.	Сумма прибыли, млн руб.
1	150,0	45,1	16	167,1	58,0
2	40,0	6,2	17	130,0	47,0
3	180,0	67,0	18	171,0	64,7
4	88,3	27,3	19	148,3	46,2
5	170,0	62,5	20	150,0	53,7
6	169,0	60,0	21	180,0	67,0
7	70,0	16,9	22	198,1	68,0
8	112,0	20,9	23	200,0	70,0
9	170,0	65,0	24	211,0	80,1
10	93,3	16,0	25	190,0	67,7
11	136,4	69,0	26	205,0	72,0
12	120,0	35,0	27	225,0	84,0
13	135,4	53,4	28	230,0	87,0
14	173,0	66,2	29	240,0	90,2
15	160,0	56,0	30	230,0	85,0

Задание 1

По исходным данным (табл. 1) необходимо выполнить следующее:

1. Построить статистический ряд распределения банков по **Объему кредитных вложений**, образовав **четыре** группы с равными интервалами.
2. Графическим методом и путем расчётов определить значения *моды* и *медианы* полученного ряда распределения.
3. Рассчитать характеристики ряда распределения: *среднюю арифметическую*, *среднее квадратическое отклонение*, *коэффициент вариации*.
4. Вычислить *среднюю арифметическую* по исходным данным (табл. 1.1), сравнить её с аналогичным показателем, рассчитанным в п. 3 для интервального ряда распределения. Объяснить причину их расхождения.

Сделать выводы по результатам выполнения Задания 1.

Выполнение Задания 1

Целью выполнения данного Задания является изучение состава и структуры выборочной совокупности банков путем построения и анализа статистического ряда распределения банков по признаку *Объем кредитных вложений*.

1. Построение интервального ряда распределения банков по объему кредитных вложений

Для построения интервального вариационного ряда, характеризующего распределение банков по объему кредитных вложений, необходимо вычислить *величину и границы интервалов ряда*.

При построении ряда с равными интервалами величина интервала h определяется по формуле

$$h = \frac{x_{max} - x_{min}}{k}, \quad (1)$$

где x_{max} , x_{min} – наибольшее и наименьшее значения признака в исследуемой совокупности, k - число групп интервального ряда.

Число групп k задается в условии задания или рассчитывается по формуле Г.Стерджесса

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

где n - число единиц совокупности.

Определение величины интервала по формуле (1) при заданных $k = 4$, $x_{max} = 240$ млн руб., $x_{min} = 40$ млн руб.:

$$h = \frac{240 - 40}{4} = \frac{200}{4} = 50 \text{ млн руб.}$$

При $h = 50$ млн руб. границы интервалов ряда распределения имеют следующий вид (табл. 2):

Таблица 2

Номер группы	Нижняя граница, млн руб.	Верхняя граница, млн руб.
1	40	90
2	90	140
3	140	190
4	190	240

Для построения интервального ряда необходимо подсчитать число банков, входящих в каждую группу (*частоты групп*). При этом возникает вопрос, в какую группу включать единицы совокупности, у которых значения признака выступают одновременно и верхней, и нижней границами смежных интервалов (для демонстрационного примера – это 90, 140, 190 млн руб.). Отнесение таких единиц к

одной из двух смежных групп рекомендуется осуществлять *по принципу полуоткрытого интервала* [). Т.к. при этом верхние границы интервалов не принадлежат данным интервалам, то соответствующие им единицы совокупности включаются не в данную группу, а в следующую. В последний интервал включаются и *нижняя*, и *верхняя границы*.

Процесс группировки единиц совокупности по признаку *Объем кредитных вложений* представлен во вспомогательной (разработочной) таблице 3 (графа 4 этой таблицы необходима для построения аналитической группировки в Задании 2).

Таблица 3

Разработочная таблица для построения интервального ряда распределения и аналитической группировки

Группы банков по объему кредитных вложений, млн руб.	Номер банка	Объем кредитных вложений, млн руб.	Сумма прибыли, млн руб.
1	2	3	4
40 – 90	2	40,0	6,2
	7	70,0	16,9
	4	88,3	27,3
Всего	3	198,3	50,4
90 – 140	10	93,3	16,0
	8	112,0	20,9
	12	120,0	35,0
	17	130,0	47,0
	13	135,4	53,4
	11	136,4	69,0
Всего	6	727,1	241,3
140 – 190	19	148,3	46,2
	1	150,0	45,1
	20	150,0	53,7
	15	160,0	56,0
	16	167,1	58,0
	6	169,0	60,0
	5	170,0	62,5
	9	170,0	65,0
	18	171,0	64,7
	14	173,0	66,2
	3	180,0	67,0
	21	180,0	67,0
Всего	12	1988,4	711,4
191 – 240	25	190,0	67,7
	22	198,1	68,0
	23	200,0	70,0

	26	205,0	72,0
	24	211,0	80,1
	27	225,0	84,0
	28	230,0	87,0
	30	230,0	85,0
	29	240,0	90,2
Всего	9	1929,1	704,0
ИТОГО	30	4842,9	1707,1

На основе групповых итоговых строк «Всего» табл. 3 формируется итоговая таблица 4, представляющая *интервальный ряд распределения банков по объему кредитных вложений*.

Таблица 4
Распределение банков по объему кредитных вложений

Номер группы	Группы банков по объему кредитных вложений, млн руб., x	Число банков, f
1	40 – 90	3
2	90 – 140	6
3	140 – 190	12
4	190 – 240	9
	Итого	30

Помимо частот групп в абсолютном выражении в анализе интервальных рядов используются ещё три характеристики ряда, приведенные в графах 4 - 6 табл. 1.4. Это *частоты групп в относительном выражении, накопленные (кумулятивные) частоты S_j* , получаемые путем последовательного суммирования частот всех предшествующих (j-1) интервалов, и *накопленные частоты*, рассчитываемые по

формуле $\frac{S_j}{\sum f_j} 100$.

Таблица 5
Структура банков по объему кредитных вложений

№ группы	Группы банков по объему кредитных вложений, млн руб.	Число банков, f_j		Накопленная частота, S_j	Накопленная частота, %
		в абсолютном выражении	в % к итогу		
1	2	3	4	5	6
1	40 – 90	3	10,0	3	10,0
2	90 – 140	6	20,0	9	30,0
3	140 – 190	12	40,0	21	70,0
4	190 – 240	9	30,0	30	100,0
	Итого	30	100,0		

Вывод. Анализ интервального ряда распределения изучаемой совокупности банков показывает, что распределение банков по объему кредитных вложений не является равномерным: преобладают банки с кредитными вложениями от 140 млн руб. до 190 млн руб. (это 12 банков, доля которых составляет 40%); 30% банков имеют кредитные вложения менее 140 млн руб., а 70% – менее 190 млн руб.

1.2. Нахождение моды и медианы полученного интервального ряда распределения графическим методом и путем расчетов

Мода и медиана являются *структурными средними величинами*, характеризующими (наряду со средней арифметической) центр распределения единиц совокупности по изучаемому признаку.

Мода M_o для дискретного ряда – это значение признака, наиболее часто встречающееся у единиц исследуемой совокупности¹. В интервальном вариационном ряду модой приближенно считается *центральное значение модального интервала* (имеющего наибольшую частоту). Более точно моду можно определить графическим методом по гистограмме ряда (рис.1).

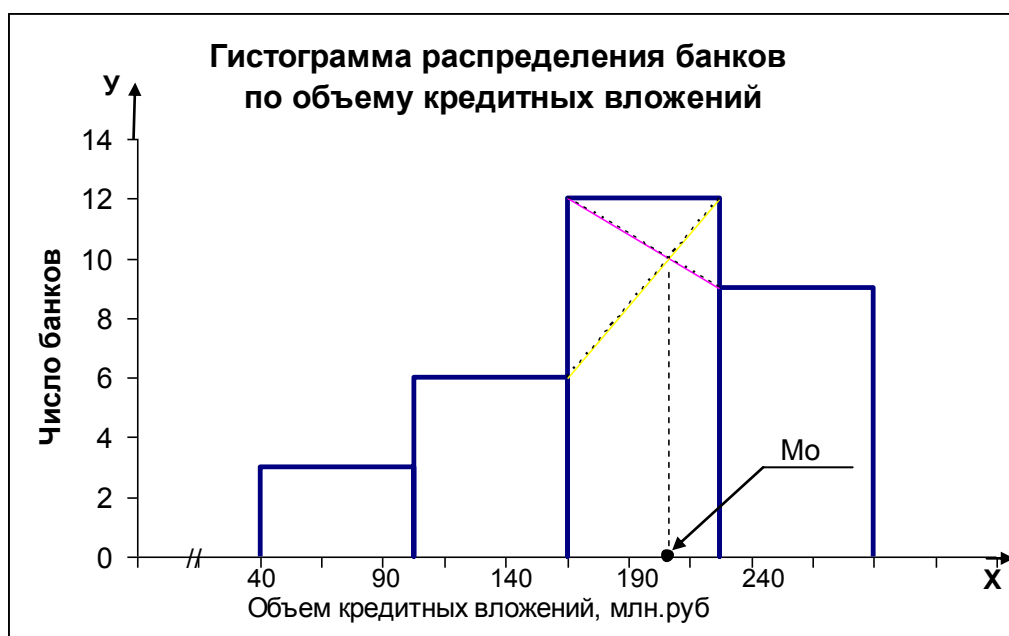


Рис. 1 Определение моды графическим методом

¹ Если в дискретном ряду все варианты встречаются одинаково часто, то в этом случае мода отсутствует. Могут быть распределения, где не один, а два (или более) варианта имеют наибольшие частоты. Тогда ряд имеет две (или более) моды, распределение является бимодальным (или многомодальным), что указывает на качественную неоднородность совокупности по изучаемому признаку.

Конкретное значение моды для интервального ряда рассчитывается по формуле:

$$M_o = x_{M_o} + h \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}, \quad (3)$$

где x_{M_o} – нижняя граница модального интервала,

h – величина модального интервала,

f_{M_o} – частота модального интервала,

f_{M_o-1} – частота интервала, предшествующего модальному,

f_{M_o+1} – частота интервала, следующего за модальным.

Согласно табл.1.3 модальным интервалом построенного ряда является интервал 140 – 190 млн. руб., так как его частота максимальна ($f_3 = 12$).

Расчет моды по формуле (3):

$$M_o = 140 + 50 \cdot \frac{12 - 6}{(12 - 6) + (12 - 9)} = 173,33 \text{ млн руб.}$$

Вывод. Для рассматриваемой совокупности банков наиболее распространенный объем кредитных вложений характеризуется средней величиной 173,33 млн руб.

Медиана Me – это значение признака, приходящееся на середину ранжированного ряда. По обе стороны от медианы находится одинаковое количество единиц совокупности.

Медиану можно определить графическим методом по кумулятивной кривой (рис. 2). Кумулята строится по накопленным частотам (табл. 5, графа 5).



Рис. 2. Определение медианы графическим методом

Конкретное значение медианы для интервального ряда рассчитывается по формуле:

$$Me = x_{Me} + h \frac{\frac{\sum_{j=1}^k f_j}{2} - S_{Me-1}}{f_{Me}}, \quad (4)$$

где x_{Me} – нижняя граница медианного интервала,

h – величина медианного интервала,

$\sum f_j$ – сумма всех частот,

f_{Me} – частота медианного интервала,

S_{Me-1} – кумулятивная (накопленная) частота интервала, предшествующего медианному.

Для расчета медианы необходимо, прежде всего, определить медианный интервал, для чего используются накопленные частоты (или частоты) из табл. 5 (графа 5). Так как медиана делит численность ряда пополам, она будет располагаться в том интервале, где накопленная частота **впервые равна** полусумме всех частот $\frac{\sum f_j}{2}$ или превышает ее (т.е. все предшествующие накопленные частоты меньше этой величины).

В демонстрационном примере медианным интервалом является интервал 140 – 190 млн. руб., так как именно в этом интервале накопленная частота $S_j = 21$ впервые превышает величину, равную половине численности единиц совокупности

$$\left(\frac{\sum f_j}{2} = \frac{n}{2} = \frac{30}{2} = 15 \right).$$

Расчет значения медианы по формуле (4):

$$Me = 140 + 50 \cdot \frac{\frac{30}{2} - 9}{12} = 165 \text{ млн руб.}$$

Вывод. В рассматриваемой совокупности банков половина банков имеют в среднем объем кредитных вложений не более 165 млн руб., а другая половина – не менее 165 млн руб.

3. Расчет характеристик ряда распределения

Для расчета характеристик ряда распределения \bar{x} , σ , σ^2 , V_σ на основе табл. 5 строится вспомогательная таблица 6 (x'_j – середина j-го интервала).

Таблица 6

Расчетная таблица для нахождения характеристик ряда распределения

Группы банков по объему кредитных вложений, млн руб.	Середина интервала, x'_j	Число банков, f_j	$x'_j \cdot f_j$	$x'_j - \bar{x}$	$(x'_j - \bar{x})^2$	$(x'_j - \bar{x})^2 f_j$
1	2	3	4	5	6	7
40 – 90	65	3	195	-95	9025	27075
90 – 140	115	6	690	-45	2025	12150
140 – 190	165	12	1980	5	25	300
190 – 240	215	9	1935	55	3025	27225
Итого		30	4800			66750

Расчет средней арифметической взвешенной:

$$\bar{x} = \frac{\sum_{j=1}^k x'_j \cdot f_j}{\sum_{j=1}^k f_j} = \frac{4800}{30} = 160 \text{ млн руб.} \quad (5)$$

Расчет среднего квадратического отклонения:

$$\sigma = \sqrt{\frac{\sum_{j=1}^k (x'_j - \bar{x})^2 f_j}{\sum_{j=1}^k f_j}} = \sqrt{\frac{66720}{30}} = 47,1699 \text{ млн руб.} \quad (6)$$

Расчет дисперсии:

$$\sigma^2 = 47,1699^2 = 2225,00$$

Расчет коэффициента вариации:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100 = \frac{47,1699}{160} 100 = 29,5\% \quad (7)$$

Вывод. Анализ полученных значений показателей \bar{x} и σ говорит о том, что средний объем кредитных вложений банков составляет 165 млн руб., отклонение от

среднего объема в ту или иную сторону составляет в среднем 47,17 млн руб. (или 29,5%), наиболее характерные значения объема кредитных вложений находятся в пределах от 112,83 млн руб. до 207,17 млн руб. (диапазон $\bar{x} \pm \sigma$).

Значение $V_\sigma = 29,5\%$ не превышает 33%, следовательно, вариация кредитных вложений в исследуемой совокупности банков незначительна и совокупность по данному признаку качественно однородна. Расхождение между значениями \bar{x} , Mo и Me незначительно ($\bar{x}=160$ млн руб., $Mo=173,33$ млн руб., $Me=165$ млн руб.), что подтверждает вывод об однородности совокупности банков. Таким образом, найденное среднее значение объема кредитных вложений банков (160 млн руб.) является типичной, надежной характеристикой исследуемой совокупности банков.

4. Вычисление средней арифметической по исходным данным

Для расчета применяется формула средней арифметической простой:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{4842,9}{30} = 161,43 \text{ млн руб.}, \quad (8)$$

Причина расхождения средних величин, рассчитанных по формулам (8) и (5), заключается в том, что по формуле (8) средняя определяется по фактическим значениям исследуемого признака для всех 30-ти банков, а по формуле (5) средняя вычисляется для интервального ряда, когда в качестве значений признака берутся середины интервалов x'_j и, следовательно, значение средней будет менее точным (за исключением случая равномерного распределения значений признака внутри каждой группы).

Задание 2

По исходным данным табл. 1 с использованием результатов выполнения Задания 1 необходимо выполнить следующее:

1. Установить наличие и характер корреляционной связи между признаками *Объем кредитных вложений* и *Сумма прибыли*, образовав по каждому признаку четыре группы с равными интервалами, используя методы:

а) аналитической группировки;

б) корреляционной таблицы.

2. Оценить тесноту корреляционной связи, используя коэффициент детерминации и эмпирическое корреляционное отношение.

Сделать выводы по результатам выполнения Задания 2.

Выполнение Задания 2

Целью выполнения данного Задания является выявление наличия корреляционной связи между факторным и результативным признаками, установление направления связи и оценка ее тесноты.

Факторный и результативный признаки либо задаются в условии задания, либо определяются путем проведения предварительного теоретического анализа. Лишь после того, как выяснена экономическая сущность явления и определены факторный и результативный признаки, приступают к проведению корреляционного анализа данных.

По условию Задания 2 факторным является признак *Объем кредитных вложений (X)*, результативным – признак *Сумма прибыли (Y)*.

1. Установление наличия и характера связи между признаками *Объем кредитных вложений* и *Сумма прибыли* методами аналитической группировки и корреляционной таблицы

1а. Применение метода аналитической группировки

При использовании метода аналитической группировки строится интервальный ряд распределения единиц совокупности по факторному признаку X и для каждой j -ой группы ряда определяется среднегрупповое значение \bar{y}_j результативного признака Y . Если с ростом значений фактора X от группы к группе *средние* значения \bar{y}_j *систематически* возрастают (или убывают), между признаками X и Y имеет место корреляционная связь.

Используя разработочную таблицу 3, строим аналитическую группировку, характеризующую зависимость между факторным признаком X – *Объем кредитных вложений* и результативным признаком Y – *Сумма прибыли*. Макет аналитической таблицы имеет следующий вид (табл. 7):

Таблица 7

Зависимость суммы прибыли банков от объема кредитных вложений

Номер группы	Группы банков по объему кредитных вложений, млн руб.	Число банков	Сумма прибыли, млн руб.	
			всего	в среднем на один банк
1				
2				
3				
4				
Итого				

Групповые средние значения \bar{y}_j получаем из таблицы 3 (графа 4), основываясь на итоговых строках «Всего». Построенную аналитическую группировку представляет табл. 8.

Таблица 8

Зависимость суммы прибыли банков от объема кредитных вложений

Номер группы	Группы банков по объему кредитных вложений, млн руб., x	Число банков, f_j	Сумма прибыли, млн руб.	
			всего	в среднем на один банк, \bar{y}_j
1	2	3	4	5=4:3
1	40 – 90	3	50,4	16,800
2	90 – 140	6	241,3	40,217
3	140 – 190	12	711,4	59,283
4	190 – 240	9	704,0	78,222
	Итого	30	1707,1	56,90

Вывод. Анализ данных табл. 8 показывает, что с увеличением объема кредитных вложений от группы к группе систематически возрастает и средняя прибыль по каждой группе банков, что свидетельствует о наличии прямой корреляционной связи между исследуемыми признаками.

1б. Применение метода корреляционной таблицы.

Корреляционная таблица представляет собой комбинацию двух рядов распределения. Строки таблицы соответствуют группировке единиц совокупности по факторному признаку X , а графы – группировке единиц по результативному признаку Y . На пересечении j -ой строки и k -ой графы указывается число единиц совокупности, входящих в j -ый интервал по факторному признаку и в k -ый интервал

по результативному признаку. *Концентрация частот около диагонали* построенной таблицы свидетельствует о наличии корреляционной связи между признаками. Связь прямая, если частоты располагаются по диагонали, идущей от левого верхнего угла к правому нижнему. Расположение частот по диагонали от правого верхнего угла к левому нижнему говорит об обратной связи.

Для построения корреляционной таблицы необходимо знать величины и границы интервалов по двум признакам **X** и **Y**. Величина интервала и границы интервалов для факторного признака **X** – *Объем кредитных вложений* известны из табл. 8. Для результативного признака **Y** – *Сумма прибыли* величина интервала определяется по формуле (1) при $k = 4$, $y_{max} = 90,2$ млн руб., $y_{min} = 6,2$ млн руб.:

$$h = \frac{90,2 - 6,2}{4} = \frac{84}{4} = 21 \text{ млн руб.}$$

Границы интервалов ряда распределения результативного признака **Y** имеют следующий вид (табл. 9):

Таблица 9

Номер группы	Нижняя граница, млн руб.	Верхняя граница, млн руб.
1	6,2	27,2
2	27,2	48,2
3	48,2	69,2
4	69,2	90,2

Подсчитывая с использованием принципа *полуоткрытого интервала* [) число банков, входящих в каждую группу (частоты групп), получаем *интервальный ряд распределения результативного признака* (табл. 10).

Таблица 10

Распределение банков по сумме прибыли

Группы банков по сумме прибыли, млн. руб., x	Число банков, f_j
6,2 – 27,2	4
27,2 – 48,2	5
48,2 – 69,2	14
69,2 – 90,2	7
Итого	30

Используя группировки по факторному и результативному признакам, строим корреляционную таблицу (табл. 11).

Таблица 11

Корреляционная таблица зависимости суммы прибыли банков от объема кредитных вложений

Группы банков по размеру кредитных вложений, млн руб.	Группы банков по сумме прибыли, млн руб.				
	6,2 – 27,2	27,2 – 48,2	48,2 – 69,2	69,2 – 90,2	Итого
40 – 90	2	1			3
90 – 140	2	2	2		6
140 – 190		2	10	0	12
190 – 240			2	7	9
Итого	5	5	5	15	30

Вывод. Анализ данных табл. 11 показывает, что распределение частот групп произошло вдоль диагонали, идущей из левого верхнего угла в правый нижний угол таблицы. Это свидетельствует о наличии прямой корреляционной связи между объемом кредитных вложений и суммой прибыли банков.

2. Измерение тесноты корреляционной связи с использованием коэффициента детерминации и эмпирического корреляционного отношения

Для измерения тесноты связи между факторным и результативным признаками рассчитывают специальные показатели – эмпирический коэффициент детерминации η^2 и эмпирическое корреляционное отношение η .

Эмпирический коэффициент детерминации η^2 оценивает, насколько вариация результативного признака Y объясняется вариацией фактора X (остальная часть вариации Y объясняется вариацией прочих факторов). Показатель η^2 рассчитывается как доля межгрупповой дисперсии в общей дисперсии по формуле

$$\eta^2 = \frac{\delta_x^2}{\sigma_o^2}, \quad (9)$$

где σ_o^2 – общая дисперсия признака Y ,

δ_x^2 – межгрупповая (факторная) дисперсия признака Y .

Значения показателя η^2 изменяются в пределах $0 \leq \eta^2 \leq 1$. При отсутствии

корреляционной связи между признаками **X** и **Y** имеет место равенство $\eta^2 = 0$, а при наличии функциональной связи между ними - равенство $\eta^2 = 1$.

Общая дисперсия σ_o^2 характеризует вариацию результативного признака, сложившуюся под влиянием всех действующих на **Y** факторов (систематических и случайных). Этот показатель вычисляется по формуле

$$\sigma_o^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_o)^2}{n}, \quad (10)$$

где y_i – индивидуальные значения результативного признака;

\bar{y}_o – общая средняя значений результативного признака;

n – число единиц совокупности.

Общая средняя \bar{y}_o вычисляется как средняя арифметическая простая по всем единицам совокупности:

$$\bar{y}_o = \frac{\sum_{i=1}^n \bar{y}_i}{n}, \quad (11)$$

или как средняя взвешенная по частоте групп интервального ряда:

$$\bar{y}_o = \frac{\sum_{j=1}^k \bar{y}_j f_j}{\sum_{j=1}^k f_j} \quad (12)$$

Для вычисления \bar{y}_o удобно использовать формулу (11), т.к. в табл. 8 (графы 3 и 4 итоговой строки) имеются значения числителя и знаменателя формулы.

Расчет \bar{y}_o по формуле (11):

$$\bar{y}_o = \frac{1707,1}{30} = 56,90 \text{ млн руб.}$$

Для расчета общей дисперсии σ_o^2 применяется вспомогательная таблица 12.

Таблица 12

Вспомогательная таблица для расчета общей дисперсии

Номер банка п/п	Прибыль, млн руб.	$y_i - \bar{y}_0$	$(y_i - \bar{y}_0)^2$	y_i^2
1	2	3	4	5
1	45,1	-11,803	139,3187	2034,01
2	6,2	-50,703	2570,8280	38,44
3	67,0	10,097	101,9427	4489,00
4	27,3	-29,603	876,3573	745,29
5	62,5	5,597	31,3227	3906,25
6	60,0	3,097	9,5893	3600,00
7	16,9	-40,003	1600,2667	285,61
8	20,9	-36,003	1296,2400	436,81
9	65,0	8,097	65,5560	4225,00
10	16,0	-40,903	1673,0827	256,00
11	69,0	12,097	146,3293	4761,00
12	35,0	-21,903	479,7560	1225,00
13	53,4	-3,503	12,2733	2851,56
14	66,2	9,297	86,4280	4382,44
15	56,0	-0,903	0,8160	3136,00
16	58,0	1,097	1,2027	3364,00
17	47,0	-9,903	98,0760	2209,00
18	64,7	7,797	60,7880	4186,09
19	46,2	-10,703	114,5613	2134,44
20	53,7	-3,203	10,2613	2883,69
21	67,0	10,097	101,9427	4489,00
22	68,0	11,097	123,1360	4624,00
23	70,0	13,097	171,5227	4900,00
24	80,1	23,197	538,0853	6416,01
25	67,7	10,797	116,5680	4583,29
26	72,0	15,097	227,9093	5184,00
27	84,0	27,097	734,2293	7056,00
28	87,0	30,097	905,8093	7569,00
29	90,2	33,297	1108,6680	8136,04
30	85,0	28,097	789,4227	7225,00
Итого	1707,1	1650,197	14192,2897	111331,97

Расчет общей дисперсии по формуле (10):

$$\sigma_o^2 = \frac{14192,2897}{30} = 473,0763$$

Общая дисперсия может быть также рассчитана по формуле

$$\sigma_o^2 = \overline{y_i^2} - (\bar{y})^2,$$

где $\overline{y_i^2}$ – средняя из квадратов значений результативного признака,

$(\bar{y})^2$ – квадрат средней величины значений результативного признака.

Для демонстрационного примера

$$\overline{y_i^2} = \frac{111331,97}{30} = 3711,0657$$
$$(\bar{y})^2 = 56,903^2 = 3237,9893$$

Тогда

$$\sigma_o^2 = \overline{y_i^2} - (\bar{y})^2 = 3711,0657 - 3237,9893 = 473,0763$$

Межгрупповая дисперсия δ_x^2 измеряет **систематическую вариацию** результативного признака, обусловленную влиянием признака-фактора **X** (по которому произведена группировка). Воздействие фактора **X** на результативный признак **Y** проявляется в отклонении групповых средних \bar{y}_j от общей средней \bar{y}_0 . Показатель δ_x^2 вычисляется по формуле

$$\delta_x^2 = \frac{\sum_{j=1}^k (\bar{y}_j - \bar{y}_0)^2 f_j}{\sum_{j=1}^k f_j}, \quad (13)$$

где \bar{y}_j – групповые средние,

\bar{y}_0 – общая средняя,

f_j – число единиц в j-ой группе,

k – число групп.

Для расчета межгрупповой дисперсии δ_x^2 строится вспомогательная таблица 13. При этом используются групповые средние значения \bar{y}_j из табл. 8 (графа 5).

Вспомогательная таблица для расчета межгрупповой дисперсии

Группы банков по размеру кредитных вложений, млн руб.	Число банков, f_j	Среднее значение \bar{y}_j в группе	$\bar{y}_j - \bar{y}_0$	$(\bar{y}_j - \bar{y}_0)^2 f_j$
1	2	3	4	5
40 – 90	3	16,800	-40,103	4824,8320
90 – 140	6	40,216	-16,687	1670,6690
140 – 190	12	59,283	2,380	67,9728
190 – 240	9	78,222	21,319	4090,4552
Итого	30			10653,9291

Расчет межгрупповой дисперсии δ_x^2 по формуле (11):

$$\delta_x^2 = \frac{10653,9291}{30} = 355,1310$$

Расчет эмпирического коэффициента детерминации η^2 по формуле (9):

$$\eta^2 = \frac{\delta_x^2}{\sigma_o^2} = \frac{355,1310}{473,0763} = 0,7507 \text{ или } 75,1\%$$

Вывод. 75,1% вариации суммы прибыли банков обусловлено вариацией объема кредитных вложений, а 24,9% – влиянием прочих неучтенных факторов.

Эмпирическое корреляционное отношение η оценивает тесноту связи между факторным и результативным признаками и вычисляется по формуле

$$\eta = \sqrt{\eta^2} = \sqrt{\frac{\delta_x^2}{\sigma_o^2}} \quad (14)$$

Значение показателя изменяются в пределах $0 \leq \eta \leq 1$. Чем ближе значение η к 1, тем теснее связь между признаками. Для качественной оценки тесноты связи на основе η служит шкала Чэддока (табл. 14):

Таблица 14

Шкала Чэддока

η	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характеристика силы связи	Слабая	Умеренная	Заметная	Тесная	Весьма тесная

Расчет эмпирического корреляционного отношения η по формуле (14):

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma_o^2}} = \sqrt{0,7507} = 0,866 \text{ или } 86,6\%$$

Вывод. Согласно шкале Чэддока связь между объемом кредитных вложений и суммой прибыли банков является тесной.

Задание 3

По результатам выполнения Задания 1 с вероятностью 0,954 необходимо определить:

1) ошибку выборки средней величины объема кредитных вложений банков и границы, в которых будет находиться генеральная средняя.

2) ошибку выборки доли банков с объемом кредитных вложений 175 млн руб. и выше, а также границы, в которых будет находиться генеральная доля.

3) необходимый объем выборки при заданной предельной ошибке выборки, равной 10 млн руб.

Выполнение Задания 3

Целью выполнения данного Задания является определение для генеральной совокупности коммерческих банков региона границ, в которых будут находиться величина среднего объема кредитных вложений банков и доля банков с объемом кредитных вложений не менее 175 млн руб.

1. Определение ошибки выборки для среднего объема кредитных вложений банков и границ, в которых будет находиться генеральная средняя

Применение выборочного метода наблюдения всегда связано с *установлением степени достоверности оценок показателей генеральной совокупности*, полученных на основе значений показателей выборочной совокупности. Достоверность этих оценок зависит от репрезентативности выборки, т.е. от того, насколько полно и адекватно представлены в выборке статистические свойства генеральной совокупности. Как правило, генеральные и выборочные характеристики не совпадают, а отклоняются на некоторую величину ϵ , которую называют *ошибкой выборки (ошибкой репрезентативности)*.

Значения признаков единиц, отобранных из генеральной совокупности в выборочную, всегда случайны, поэтому и статистические характеристики выборки случайны, следовательно, и ошибки выборки также случайны. Ввиду этого принято вычислять два вида ошибок - среднюю $\mu_{\tilde{x}}$ и предельную $\Delta_{\tilde{x}}$.

Средняя ошибка выборки $\mu_{\tilde{x}}$ - это среднее квадратическое отклонение всех возможных значений выборочной средней от генеральной средней, т.е. от своего математического ожидания $M[\bar{x}]$.

Величина средней ошибки выборки рассчитывается **дифференцированно** (по различным формулам) в зависимости от **вида и способа отбора единиц** из генеральной совокупности в выборочную.

Для собственно-случайной и механической выборки с бесповторным способом отбора средняя ошибка $\mu_{\tilde{x}}$ выборочной средней \tilde{x} определяется по формуле

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}, \quad (15)$$

где σ^2 – общая дисперсия выборочных значений признаков,

N – число единиц в генеральной совокупности,

n – число единиц в выборочной совокупности.

Предельная ошибка выборки $\Delta_{\tilde{x}}$ определяет границы, в пределах которых будет находиться генеральная средняя:

$$\begin{aligned} \bar{x} &= \tilde{x} \pm \Delta_{\tilde{x}}, \\ \tilde{x} - \Delta_{\tilde{x}} &\leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}}, \end{aligned} \quad (16)$$

где \tilde{x} – выборочная средняя,

\bar{x} – генеральная средняя.

Границы задают **доверительный интервал генеральной средней**, т.е. случайную область значений, которая с вероятностью P гарантированно содержит значение генеральной средней. Эту вероятность P называют **доверительной вероятностью** или **уровнем надёжности**.

В экономических исследованиях чаще всего используются доверительные вероятности $P=0.954$, $P=0.997$, реже $P=0.683$.

В математической статистике доказано, что предельная ошибка выборки Δ кратна средней ошибке μ с **коэффициентом кратности t** (называемым также **коэффициентом доверия**), который зависит от значения доверительной

вероятности P . Для предельной ошибки выборочной средней $\Delta_{\tilde{x}}$ это теоретическое положение выражается формулой

$$\Delta_{\tilde{x}} = t(P)\mu_{\tilde{x}} \quad (17)$$

Значения t вычислены заранее для различных доверительных вероятностей P и *протабулированы* (таблицы функции Лапласа Φ). Для наиболее часто используемых уровней надежности P значения t задаются следующим образом (табл. 15):

Таблица 15

Доверительная вероятность P	0,683	0,866	0,954	0,988	0,997	0,999
Значение t	1,0	1,5	2,0	2,5	3,0	3,5

По условию демонстрационного примера выборочная совокупность насчитывает 30 банков, выборка 20% механическая, следовательно, *генеральная совокупность включает 150 банков*. Выборочная средняя \tilde{x} , дисперсия σ^2 определены в Задании 1 (п. 3). Значения параметров, необходимых для решения задачи, представлены в табл. 16:

Таблица 16

P	t	n	N	\tilde{x}	σ^2
0,954	2	30	150	160	2225

Расчет средней ошибки выборки по формуле (15):

$$\mu_{\tilde{x}} = \sqrt{\frac{2225}{30} \left(1 - \frac{30}{150}\right)} = 7,703 \text{ млн руб.},$$

Расчет предельной ошибки выборки по формуле (17):

$$\Delta_{\tilde{x}} = 2 \cdot 7,703 = 15,406 \text{ млн руб.}$$

Определение по формуле (16) доверительного интервала для генеральной средней:

$$160 - 15,406 \leq \bar{x} \leq 160 + 15,406, \\ 144,594 \text{ млн руб.} \leq \bar{x} \leq 175,406 \text{ млн руб.}$$

Вывод. На основании проведенного выборочного обследования коммерческих банков региона с вероятностью 0,954 можно утверждать, что для

генеральной совокупности банков средний объем кредитных вложений банка находится в пределах от 144,59 млн руб. до 175,41 млн руб.

2. Определение ошибки выборки для доли банков с объемом кредитных вложений 175млн руб. и выше, а также границ, в которых будет находиться генеральная доля

Доля единиц выборочной совокупности, обладающих тем или иным заданным свойством, выражается формулой

$$w = \frac{m}{n}, \quad (18)$$

где m – число единиц совокупности, обладающих заданным свойством;

n – общее число единиц в совокупности.

Для собственно-случайной и механической выборки с бесповторным способом отбора предельная ошибка выборки Δ_w доли единиц, обладающих заданным свойством, рассчитывается по формуле

$$\Delta_w = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}, \quad (19)$$

где w – доля единиц совокупности, обладающих заданным свойством;

$(1-w)$ – доля единиц совокупности, не обладающих заданным свойством,

N – число единиц в генеральной совокупности,

n – число единиц в выборочной совокупности.

Предельная ошибка выборки Δ_w определяет границы, в пределах которых будет находиться генеральная доля p единиц, обладающих заданным свойством:

$$w - \Delta_w \leq p \leq w + \Delta_w \quad (20)$$

По условию Задания 3 исследуемым свойством является *равенство или превышение объема кредитных вложений банка величины 175 млн руб.*

Число банков с заданным свойством определяется из табл. 3 (графа 3):

$$m=11$$

Расчет выборочной доли по формуле (18):

$$w = \frac{11}{30} = 0,367$$

Расчет по формуле (19) предельной ошибки выборки для доли:

$$\Delta_w = 2\sqrt{\frac{0,367(1 - 0,367)}{30} \left(1 - \frac{30}{150}\right)} = 0,157$$

Определение по формуле (20) доверительного интервала генеральной доли:

$$0,367 - 0,157 \leq p \leq 0,367 + 0,157$$
$$0,21 \leq p \leq 0,524$$

или

$$21\% \leq p \leq 52,4\%$$

Вывод. С вероятностью 0,954 можно утверждать, что в генеральной совокупности банков доля банков с объемом кредитных вложений 175 млн руб. и выше будет находиться в пределах от 21% до 52,4%.

3. Определение необходимого объема выборки с заданным значением допустимой предельной ошибки выборки, равной 10 млн руб.

Для собственно-случайной и механической выборки с бесповторным способом отбора необходимый объем выборки для средней количественного признака вычисляется по формуле

$$n = \frac{t^2 \sigma^2 N}{N \Delta_{\tilde{x}}^2 + t^2 \sigma^2} \quad (21)$$

По условию демонстрационного примера ошибка выборки $\Delta_{\tilde{x}}$ не должна превышать 10 млн руб. Параметры t , N и σ^2 известны из решений предыдущих задач.

Расчет необходимой численности выборки по формуле (21):

$$n = \frac{4 \cdot 2225 \cdot 150}{150 \cdot 100 + 4 \cdot 2225} = 56 \text{ банков}$$

Вывод. Для того, чтобы обеспечить для среднего объема кредитных вложений банков предельную ошибку выборки, равную 10 млн руб., необходимо из 150 банков, составляющих генеральную совокупность, отобрать в выборочную совокупность 56 банков.

Раздел II

Образец выполнения и оформления работ

Имеются следующие выборочные данные по 30-ти однотипным фирмам одного из регионов РФ, характеризующие деятельность фирм за исследуемый период (выборка 10%-ная механическая):

Таблица 1

Исходные данные

Номер фирмы п/п	Среднесписочная численность менеджеров, чел.	Объём продаж, млн руб.	Номер фирмы п/п	Среднесписочная численность менеджеров, чел.	Объём продаж, млн руб.
1	35	3,30	16	50	4,00
2	26	2,80	17	31	3,00
3	20	2,50	18	33	3,30
4	24	2,60	19	41	3,50
5	30	3,00	20	31	3,00
6	33	3,10	21	36	3,35
7	28	2,90	22	36	3,45
8	35	3,40	23	37	3,47
9	42	3,60	24	37	3,50
10	27	2,90	25	38	3,60
11	36	3,30	26	39	3,70
12	22	2,60	27	43	3,60
13	25	2,80	28	44	4,00
14	35	3,35	29	46	3,90
15	32	3,10	30	48	3,80

Цель статистического исследования - анализ совокупности фирм по признакам

Среднесписочная численность менеджеров и Объём продаж, включая:

- изучение структуры совокупности по признаку *Среднесписочная численность менеджеров*;
- выявление наличия корреляционной связи между признаками *Среднесписочная численность менеджеров* и *Объём продаж* фирм, установление направления связи и оценка её тесноты;
- применение выборочного метода для определения статистических характеристик генеральной совокупности фирм.

Задание 1

По исходным данным (табл. 1) необходимо выполнить следующее:

1. Построить статистический ряд распределения фирм по **среднесписочной численности менеджеров**, образовав **шесть** групп с равными интервалами.
2. Графическим методом и путем расчетов определить значения *моды* и *медианы* полученного ряда распределения.
3. Рассчитать характеристики ряда распределения: *среднюю арифметическую*, *среднее квадратическое отклонение*, *коэффициент вариации*.
4. Вычислить *среднюю арифметическую* по исходным данным (табл. 1), сравнить её с аналогичным показателем, рассчитанным в п. 3 для интервального ряда распределения. Объяснить причину их расхождения.

Сделать выводы по результатам выполнения Задания 1.

Выполнение Задания 1

Целью выполнения данного Задания является изучение состава и структуры выборочной совокупности фирм путем построения и анализа статистического ряда распределения фирм по признаку *Среднесписочная численность менеджеров*.

1. Построение интервального ряда распределения фирм по среднесписочной численности менеджеров

Для построения интервального ряда распределения определяем величину интервала h по формуле:

$$h = \frac{x_{max} - x_{min}}{k},$$

где x_{max} , x_{min} – наибольшее и наименьшее значения признака в исследуемой совокупности, k - число групп интервального ряда.

При заданных $k = 6$, $x_{max} = 50$ чел. и $x_{min} = 20$ чел.

$$h = \frac{50 - 20}{6} = 5 \text{ чел.}$$

При $h = 5$ чел. границы интервалов ряда распределения имеют следующий вид (табл. 2):

Таблица 2

Номер группы	Нижняя граница, чел.	Верхняя граница, чел.
1	20	25
2	25	30
3	30	35
4	35	40
5	40	45
6	45	50

Определяем число фирм, входящих в каждую группу, используя *принцип полуоткрытого интервала* [), согласно которому фирмы со значениями признаков, которые служат одновременно верхними и нижними границами смежных интервалов (25, 30, 35, 40 и 45 чел.), будем относить ко второму из смежных интервалов.

Для определения числа фирм в каждой группе строим разработочную таблицу 3 (данные графы 4 потребуются при выполнении Задания 2).

Таблица 3

Разработочная таблица для построения интервального ряда распределения и аналитической группировки

Группы фирм по среднесписочной численности менеджеров, чел.	Номер фирмы	Среднесписочная численность менеджеров, чел.	Объем продаж, млн руб.
1	2	3	4
20 – 25	3	20	2,50
	12	22	2,60
	4	24	2,60
Всего	3	66	7,70
25 – 30	13	25	2,80
	2	26	2,80
	10	27	2,90
	7	28	2,90
Всего	4	106	11,40
30 – 35	5	30	3,00
	17	31	3,00
	20	31	3,00
	15	32	3,10
	6	33	3,10
	18	33	3,30
Всего	6	190	18,50

35-40	1	35	3,30
	8	35	3,40
	14	35	3,35
	11	36	3,30
	21	36	3,35
	22	36	3,45
	23	37	3,47
	24	37	3,50
	25	38	3,60
	26	39	3,70
Всего	10	364	34,42
40-45	19	41	3,50
	9	42	3,60
	27	43	3,60
	28	44	4,00
Всего	4	170	14,70
45-50	29	46	3,90
	30	48	3,80
	16	50	4,00
Всего	3	144	11,70
Итого	30	1040	98,42

На основе групповых итоговых строк «Всего» табл. 3 формируем итоговую таблицу 4, представляющую *интервальный ряд распределения фирм по среднесписочной численности менеджеров.*

Таблица 4

Распределение фирм по среднесписочной численности менеджеров

Номер группы	Группы фирм по среднесписочной численности менеджеров, чел., x	Число фирм, f_j
1	20-25	3
2	25-30	4
3	30-35	6
4	35-40	10
5	40-45	4
6	45-50	3
	ИТОГО	30

Приведем еще три характеристики полученного ряда распределения - частоты групп в относительном выражении, накопленные (кумулятивные) частоты S_j , получаемые путем последовательного суммирования частот всех

предшествующих (**j-1**) интервалов, и *накопленные частоты*, рассчитываемые по

формуле $\frac{S_j}{\sum f_j} 100$.

Таблица 5

Структура фирм по среднесписочной численности менеджеров

Номер группы	Группы фирм по среднесписочной численности менеджеров, чел., <i>x</i>	Число фирм, <i>f</i>		Накопленная частота <i>S_j</i>	Накопленная частотность, %
		в абсолютном выражении	в % к итогу		
1	2	3	4	5	6
1	20-25	3	10	3	10
2	25-30	4	13	7	23
3	30-35	6	20	13	43
4	35-40	10	34	23	77
5	40-45	4	13	27	90
6	45-50	3	10	30	100
	ИТОГО	30	100		

Вывод. Анализ интервального ряда распределения изучаемой совокупности фирм показывает, что распределение фирм по среднесписочной численности менеджеров не является равномерным: преобладают фирмы со среднесписочной численностью менеджеров от 35 чел. до 40 чел. (это 10 фирм, доля которых составляет 34%); самые малочисленные группы фирм имеют 25-30 чел. и 45-50 чел., каждая из которых включает 3 фирмы, что составляет по 10% от общего числа фирм.

2. Нахождение моды и медианы полученного интервального ряда распределения графическим методом и путем расчетов

Для определения моды графическим методом строим по данным табл. 4 (графы 2 и 3) гистограмму распределения фирм по изучаемому признаку.

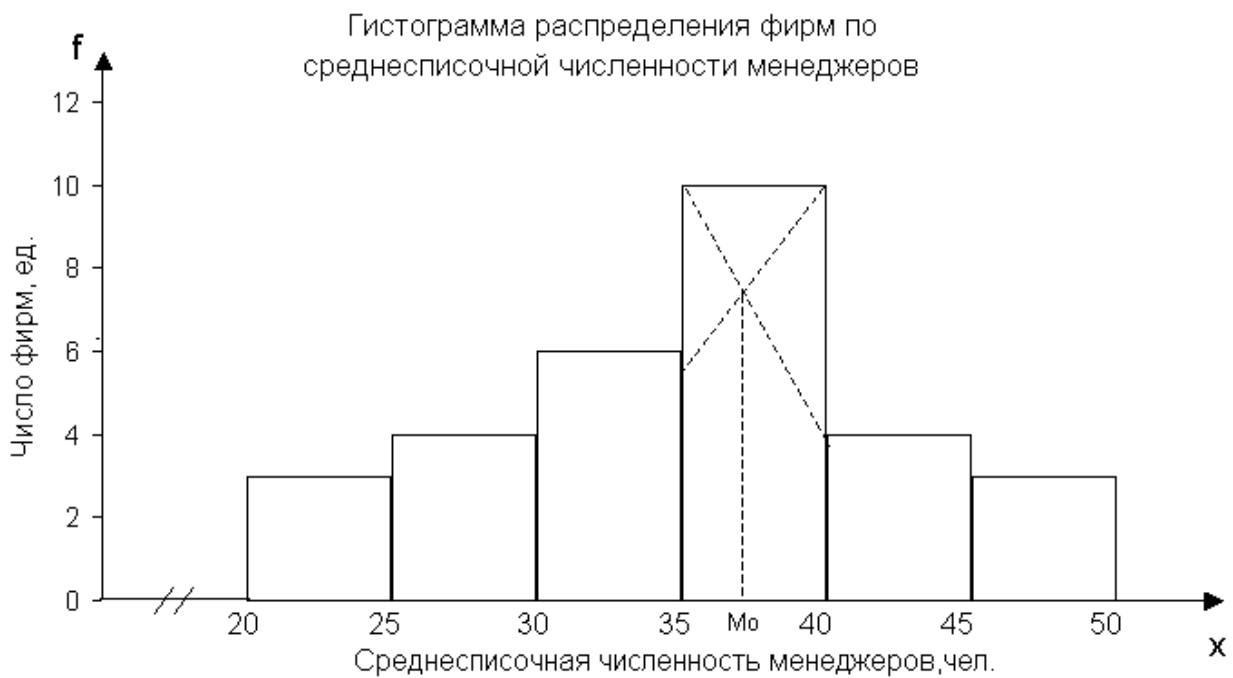


Рис. 1. Определение моды графическим методом

Расчет конкретного значения моды для интервального ряда распределения производится по формуле:

$$M_o = x_{M_o} + h \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})},$$

где x_{M_o} – нижняя граница модального интервала,

h – величина модального интервала,

f_{M_o} – частота модального интервала,

f_{M_o-1} – частота интервала, предшествующего модальному,

f_{M_o+1} – частота интервала, следующего за модальным.

Согласно табл. 4 модальным интервалом построенного ряда является интервал 35 - 40 чел., т.к. он имеет наибольшую частоту ($f_4=10$). Расчет моды:

$$M_o = 35 + 5 \cdot \frac{10 - 6}{(10 - 6) + (10 - 4)} = 35 + 5 \cdot \frac{4}{10} = 37 \text{ чел.}$$

Вывод. Для рассматриваемой совокупности фирм наиболее распространенная среднесписочная численность менеджеров характеризуется средней величиной 37 человек.

Для определения медианы графическим методом строим по данным табл. 5 (графы 2 и 5) кумуляту распределения фирм по изучаемому признаку.

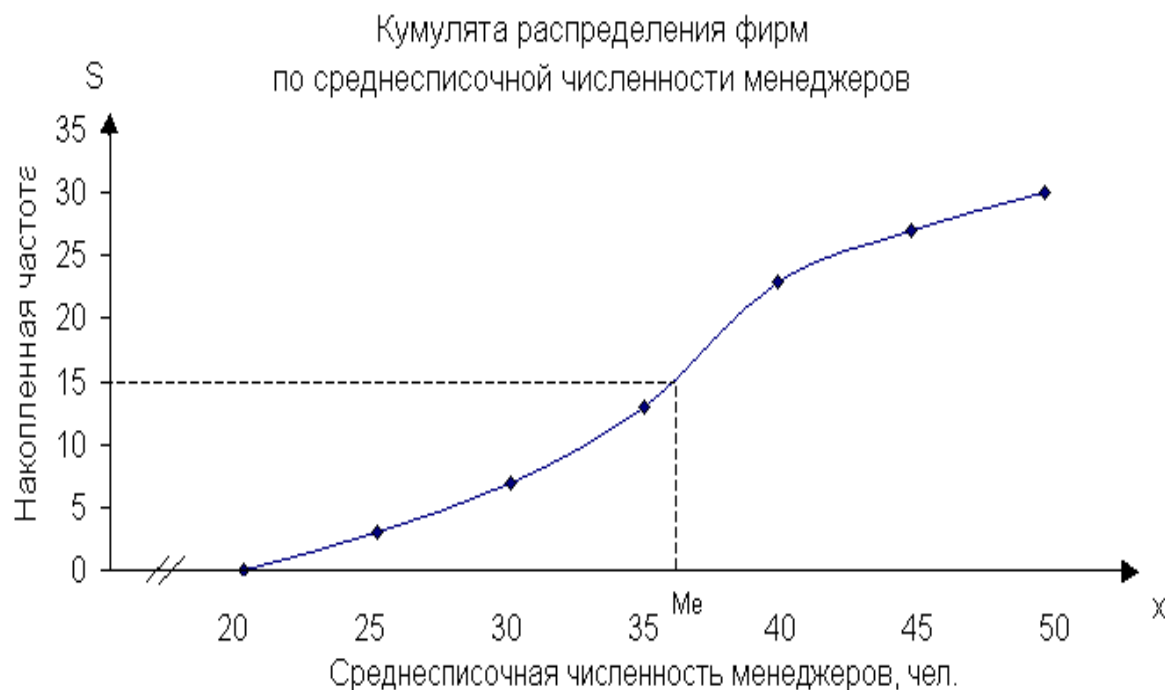


Рис. 2. Определение медианы графическим методом

Расчет конкретного значения медианы для интервального ряда распределения производится по формуле

$$Me = x_{Me} + h \frac{\frac{\sum_{j=1}^k f_j}{2} - S_{Me-1}}{f_{Me}},$$

где x_{Me} — нижняя граница медианного интервала,

h — величина медианного интервала,

$\sum f_j$ — сумма всех частот,

f_{Me} — частота медианного интервала,

S_{Me-1} — кумулятивная (накопленная) частота интервала, предшествующего медианному.

Определяем медианный интервал, используя графу 5 табл. 5. Медианным интервалом является интервал 35-40 чел., т.к. именно в этом интервале накопленная частота $S_j=23$ впервые превышает полусумму всех частот ($\frac{\sum f_i}{2} = \frac{30}{2} = 15$).

Расчет медианы:

$$Me = 35 + 5 \cdot \frac{15 - 13}{10} = 36 \text{ чел.}$$

Вывод. В рассматриваемой совокупности фирм половина фирм имеют среднесписочную численность менеджеров не более 36 человек, а другая половина – не менее 36 человек.

3. Расчет характеристик ряда распределения

Для расчета характеристик ряда распределения \bar{x} , σ , σ^2 , V_σ на основе табл. 5 строим вспомогательную таблицу 6 (x'_j – середина интервала).

Таблица 6

Расчетная таблица для нахождения характеристик ряда распределения

Группы фирм по среднесписочной численности менеджеров, чел.	Середина интервала, x'_j	Число фирм, f_j	$x'_j f_j$	$x'_j - \bar{x}$	$(x'_j - \bar{x})^2$	$(x'_j - \bar{x})^2 f_j$
1	2	3	4	5	6	7
20-25	22,5	3	67,5	-12,833	164,6859	494,0577
25-30	27,5	4	110,0	-7,833	61,3559	245,4236
30-35	32,5	6	195,0	-2,833	8,0259	48,1553
35-40	37,5	10	375,0	2,167	4,6959	46,9589
40-45	42,5	4	170,0	7,167	51,3659	205,4636
45-50	47,5	3	142,5	12,167	148,0359	444,1077
ИТОГО		30	1060,0			1484,16668

Рассчитаем среднюю арифметическую взвешенную:

$$\bar{x} = \frac{\sum_{j=1}^k x'_j f_j}{\sum_{j=1}^k f_j} = \frac{1060}{30} = 35,33 \approx 35 \text{ чел.}$$

Рассчитаем среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum_{j=1}^k (x'_j - \bar{x})^2 f_j}{\sum_{j=1}^k f_j}} = \sqrt{\frac{1484,16668}{30}} = 7,0336 \approx 7 \text{ чел.}$$

Рассчитаем дисперсию:

$$\sigma^2 = 7,0336^2 = 49,47222$$

Рассчитаем коэффициент вариации:

$$V_{\sigma} = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{7,0336}{35,33} \cdot 100 = 19,9\%$$

Вывод. Анализ полученных значений показателей \bar{x} и σ говорит о том, что средняя величина среднесписочной численности менеджеров составляет 35 чел., отклонение от этой величины в ту или иную сторону составляет в среднем 7 чел. (или 19,9%), наиболее характерная среднесписочная численность менеджеров находится в пределах от 28 до 42 чел. (диапазон $\bar{x} \pm \sigma$).

Значение $V_{\sigma} = 19,9\%$ не превышает 33%, следовательно, вариация среднесписочной численности менеджеров в исследуемой совокупности фирм незначительна и совокупность по данному признаку однородна. Расхождение между значениями \bar{x} , Mo и Me незначительно ($\bar{x}=35$ чел., $Mo=37$ чел., $Me=36$ чел.), что подтверждает вывод об однородности совокупности фирм. Таким образом, найденное среднее значение среднесписочной численности менеджеров (35 чел.) является типичной, надежной характеристикой исследуемой совокупности фирм.

4. Вычисление средней арифметической по исходным данным о среднесписочной численности менеджеров фирм

Для расчета применяется формула средней арифметической простой:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1040}{30} = 34,67 \approx 35 \text{ чел.},$$

Причина расхождения средних величин, рассчитанных по исходным данным (34,67 чел.) и по интервальному ряду распределения (35,33 чел.), заключается в том, что в первом случае средняя определяется по *фактическим значениям* исследуемого признака для всех 30-ти фирм, а во втором случае в качестве значений признака берутся *середины интервалов* x'_j и, следовательно, значение средней будет менее точным. Вместе с тем, при округлении обеих рассматриваемых величин их значения совпадают (35 чел.), что говорит о достаточно равномерном распределении численности менеджеров внутри каждой группы интервального ряда.

Задание 2

По исходным данным (табл. 1) с использованием результатов выполнения Задания 1 необходимо выполнить следующее:

1. Установить наличие и характер корреляционной связи между признаками *Среднесписочная численность менеджеров* и *Объём продаж*, образовав шесть групп с равными интервалами по каждому из признаков, используя методы:
 - а) аналитической группировки;
 - б) корреляционной таблицы.
 2. Измерить тесноту корреляционной связи, используя *коэффициент детерминации* и *эмпирическое корреляционное отношение*.
- Сделать **выводы** по результатам выполнения задания 2.

Выполнение задания 2

Целью выполнения данного задания является выявление наличия корреляционной связи между факторным и результативным признаками, а также установление направления связи и оценка ее тесноты.

По условию Задания 2 факторным является признак *Среднесписочная численность менеджеров*, результативным – признак *Объём продаж*.

1. Установление наличия и характера корреляционной связи между признаками *Среднесписочная численность менеджеров* и *Объём продаж* методами аналитической группировки и корреляционных таблиц

1а. Применение метода аналитической группировки

Аналитическая группировка строится по факторному признаку X и для каждой j -ой группы ряда определяется среднегрупповое значение \bar{y}_j результативного признака Y . Если с ростом значений фактора X от группы к группе средние значения \bar{y}_j *систематически* возрастают (или убывают), между признаками X и Y имеет место корреляционная связь.

Используя разработочную таблицу 3, строим аналитическую группировку, характеризующую зависимость между факторным признаком X - *Среднесписочная численность менеджеров* и результативным признаком Y - *Объём продаж*. Макет аналитической таблицы имеет следующий вид (табл. 7):

Таблица 7

Зависимость объема продаж от среднесписочной численности менеджеров

Номер группы	Группы фирм по среднесписочной численности менеджеров, чел., x	Число фирм, f_j	Объем продаж, млн руб.	
			всего	в среднем на одну фирму, \bar{y}_j
1	2	3	4	5=4:3
1				
2				
3				
4				
5				
6				
	ИТОГО			

Групповые средние значения \bar{y}_j получаем из таблицы 3 (графа 4), основываясь на итоговых строках «Всего». Построенную аналитическую группировку представляет табл. 8:

Таблица 8

Зависимость объема продаж от среднесписочной численности менеджеров

Номер группы	Группы фирм по среднесписочной численности менеджеров, чел., x	Число фирм, f_j	Объем продаж, млн руб.	
			всего	в среднем на одну фирму, \bar{y}_j
1	2	3	4	5=4:3
1	20-25	3	7,70	2,5667
2	25-30	4	11,40	2,8500
3	30-35	6	18,50	3,0833
4	35-40	10	34,42	3,4420
5	40-45	4	14,70	3,6750
6	45-50	3	11,70	3,9000
	ИТОГО	30	98,42	

Вывод. Анализ данных табл. 8 показывает, что с увеличением среднесписочной численности менеджеров от группы к группе систематически возрастает и средний объем продаж по каждой группе фирм, что свидетельствует о наличии прямой корреляционной связи между исследуемыми признаками.

16. Применение метода корреляционных таблиц

Корреляционная таблица строится как комбинация двух рядов распределения по факторному признаку X и результативному признаку Y . На пересечении j -ой строки и k -ой графы таблицы указывается число единиц совокупности, входящих в j -ый интервал по признаку X и в k -ый интервал по признаку Y . Концентрация частот около диагонали построенной таблицы свидетельствует о наличии корреляционной связи между признаками - прямой или обратной. Связь прямая, если частоты располагаются по диагонали, идущей от левого верхнего угла к правому нижнему, обратная - по диагонали от правого верхнего угла к левому нижнему.

Для построения корреляционной таблицы необходимо знать величины и границы интервалов по двум признакам X и Y . Для факторного признака X – *Среднесписочная численность менеджеров* эти величины известны из табл. 4. Определяем величину интервала для результативного признака Y – *Объем продаж* при $k = 6$, $y_{max} = 4,0$ млн руб., $y_{min} = 2,5$ млн руб.:

$$h = \frac{4,0 - 2,5}{6} = 0,25 \text{ млн руб.}$$

Границы интервалов ряда распределения результативного признака Y имеют вид:

Таблица 9

Номер группы	Нижняя граница, млн руб.	Верхняя граница, млн руб.
1	2,50	2,75
2	2,75	3,00
3	3,00	3,25
4	3,25	3,50
5	3,50	3,75
6	3,75	4,00

Подсчитывая для каждой группы число входящих в нее фирм с использованием *принципа полуоткрытого интервала* [], получаем *интервальный ряд распределения результативного признака* (табл. 10).

Таблица 10

Интервальный ряд распределения фирм по объёму продаж

Группы фирм по объёму продаж, млн руб., y	Число фирм, f_i
2,50-2,75	3
2,75-3,00	4
3,00-3,25	5
3,25-3,50	8
3,50-3,75	6
3,75-4,00	4
ИТОГО	30

Используя группировки по факторному и результативному признакам, строим корреляционную таблицу (табл. 11).

Таблица 11

Корреляционная таблица зависимости объема продаж от среднесписочной численности менеджеров

Группы фирм по среднесписочной численности менеджеров, чел.	Группы фирм по объёму продаж, млн руб.						ИТОГО
	2,50-2,75	2,75-3,00	3,00-3,25	3,25-3,50	3,50-3,75	3,75-4,00	
20-25	3						3
25-30		4					4
30-35			5	1			6
35-40				7	2	1	10
40-45					4		4
45-50						3	3
ИТОГО	3	4	5	8	6	4	30

Вывод. Анализ данных табл. 11 показывает, что распределение частот групп произошло вдоль диагонали, идущей из левого верхнего угла в правый нижний угол таблицы. Это свидетельствует о наличии прямой корреляционной связи между среднесписочной численностью менеджеров и объемом продаж фирмами.

2. Измерение тесноты корреляционной связи с использованием коэффициента детерминации η^2 и эмпирического корреляционного отношения η

Коэффициент детерминации η^2 характеризует силу влияния факторного (группировочного) признака X на результативный признак Y и рассчитывается как доля межгрупповой дисперсии δ_x^2 признака Y в его общей дисперсии σ_0^2 :

$$\eta^2 = \frac{\delta_x^2}{\sigma_0^2}$$

где σ_o^2 – общая дисперсия признака Y ,

δ_x^2 – межгрупповая (факторная) дисперсия признака Y .

Общая дисперсия σ_o^2 характеризует вариацию результативного признака, сложившуюся под влиянием *всех действующих на Y факторов (систематических и случайных)* и вычисляется по формуле

$$\sigma_o^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_o)^2}{n}, \quad (10)$$

где y_i – индивидуальные значения результативного признака;

\bar{y}_o – общая средняя значений результативного признака;

n – число единиц совокупности.

Межгрупповая дисперсия δ_x^2 измеряет *систематическую вариацию* результативного признака, обусловленную *влиянием признака-фактора X* (по которому произведена группировка) и вычисляется по формуле

$$\delta_x^2 = \frac{\sum_{j=1}^k (\bar{y}_j - \bar{y}_o)^2 f_j}{\sum_{j=1}^k f_j}, \quad (13)$$

где \bar{y}_j – групповые средние,

\bar{y}_o – общая средняя,

f_j – число единиц в j -ой группе,

k – число групп.

Для расчета показателей σ_o^2 и δ_x^2 необходимо знать величину *общей средней \bar{y}_o* , которая вычисляется как *средняя арифметическая простая* по всем единицам совокупности:

$$\bar{y}_0 = \frac{\sum_{i=1}^n y_i}{n}$$

Значения числителя и знаменателя формулы имеются в табл. 8 (графы 3 и 4 итоговой строки). Используя эти данные, получаем общую среднюю \bar{y}_0 :

$$\bar{y}_0 = \frac{98,42}{30} = 3,281 \text{ млн руб.}$$

Для расчета общей дисперсии σ_0^2 применяется вспомогательная таблица 12.

Таблица 12

Вспомогательная таблица для расчета общей дисперсии

Номер фирмы	Объём продаж, млн. руб.	$y_i - \bar{y}_0$	$(y_i - \bar{y}_0)^2$
1	2	3	4
1	3,30	0,019	0,0004
2	2,80	0,481	0,2310
3	2,50	0,781	0,6094
4	2,60	0,681	0,4633
5	3,00	0,281	0,0788
6	3,10	0,181	0,0326
7	2,90	0,381	0,1449
8	3,40	0,119	0,0142
9	3,60	0,319	0,1020
10	2,90	0,381	0,1449
11	3,30	0,019	0,0004
12	2,60	0,681	0,4633
13	2,80	0,481	0,2310
14	3,35	0,069	0,0048
15	3,10	0,181	0,0326
16	4,00	0,719	0,5174
17	3,00	0,281	0,0788
18	3,30	0,019	0,0004
19	3,50	0,219	0,0481
20	3,00	0,281	0,0788
21	3,35	0,069	0,0048
22	3,45	0,169	0,0287
23	3,47	0,189	0,0358
24	3,50	0,219	0,0481
25	3,60	0,319	0,1020
26	3,70	0,419	0,1758
27	3,60	0,319	0,1020
28	4,00	0,719	0,5174
29	3,90	0,619	0,3836
30	3,80	0,519	0,2697
Итого	98,42		4,9452

Рассчитаем общую дисперсию:

$$\sigma_o^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_o)^2}{n} = \frac{4,9452}{30} = 0,1648$$

Для расчета межгрупповой дисперсии δ_x^2 строится вспомогательная таблица 13. При этом используются групповые средние значения \bar{y}_j из табл. 8 (графа 5).

Таблица 13

Вспомогательная таблица для расчета межгрупповой дисперсии

Группы фирм по среднесписочной численности менеджеров, чел., x	Число фирм, f_j	Среднее значение в группе, млн руб. \bar{y}_j	$\bar{y}_j - \bar{y}_o$	$(\bar{y}_j - \bar{y}_o)^2 f_j$
1	2	3	4	5
20-25	3	2,567	-0,714	1,5294
25-30	4	2,850	-0,431	0,7420
30-35	6	3,083	-0,197	0,2338
35-40	10	3,442	0,161	0,2602
40-45	4	3,675	0,394	0,6219
45-50	3	3,900	0,619	1,1506
ИТОГО	30			4,5379

Рассчитаем межгрупповую дисперсию:

$$\delta_x^2 = \frac{\sum_{j=1}^k (\bar{y}_j - \bar{y}_o)^2 f_j}{\sum_{j=1}^k f_j} = \frac{4,5379}{30} = 0,1513$$

Определяем коэффициент детерминации:

$$\eta^2 = \frac{\delta_x^2}{\sigma_o^2} = \frac{0,1513}{0,1648} = 0,9181 \text{ или } 91,8\%$$

Вывод. 91,8% вариации объёма продаж товаров фирмами обусловлено вариацией среднесписочной численности менеджеров по продажам, а 8,2% – влиянием прочих неучтенных факторов.

Эмпирическое корреляционное отношение η оценивает *тесноту связи* между факторным и результативным признаками и вычисляется по формуле

$$\eta = \sqrt{\eta^2} = \sqrt{\frac{\delta_x^2}{\sigma_o^2}}$$

Рассчитаем показатель η :

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma_o^2}} = \sqrt{0,9181} = 0,96$$

Вывод: согласно шкале Чэддока связь между среднесписочной численностью менеджеров и объёмом продаж фирмами является весьма тесной.

Задание 3

По результатам выполнения Задания 1 с вероятностью 0,954 необходимо определить:

1) ошибку выборки для средней величины среднесписочной численности менеджеров, а также границы, в которых будет находиться генеральная средняя.

2) ошибку выборки доли фирм со среднесписочной численностью менеджеров 40 человек и более, а также границы, в которых будет находиться генеральная доля фирм.

Выполнение Задания 3

Целью выполнения данного Задания является определение для генеральной совокупности фирм региона границ, в которых будут находиться средняя величина среднесписочной численности менеджеров, и доля фирм со среднесписочной численностью менеджеров не менее 40 человек.

1. Определение ошибки выборки для величины среднесписочной численности менеджеров, а также границ, в которых будет находиться генеральная средняя

Применяя выборочный метод наблюдения, необходимо рассчитать ошибки выборки (ошибки репрезентативности), т.к. генеральные и выборочные характеристики, как правило, не совпадают, а отклоняются на некоторую величину ε .

Принято вычислять два вида ошибок выборки - *среднюю $\mu_{\bar{x}}$* и *предельную $\Delta \bar{x}$* .

Для расчета средней ошибки выборки $\mu_{\bar{x}}$ применяются *различные формулы в зависимости от вида и способа отбора единиц* из генеральной совокупности в выборочную.

Для *собственно-случайной* и *механической* выборки с *бесповторным способом отбора* средняя ошибка $\mu_{\tilde{x}}$ для выборочной средней \tilde{x} определяется по формуле

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

где σ^2 – общая дисперсия изучаемого признака,

N – число единиц в генеральной совокупности,

n – число единиц в выборочной совокупности.

Предельная ошибка выборки $\Delta_{\tilde{x}}$ определяет границы, в пределах которых будет находиться генеральная средняя:

$$\begin{aligned} \bar{x} &= \tilde{x} \pm \Delta_{\tilde{x}}, \\ \tilde{x} - \Delta_{\tilde{x}} &\leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}}, \end{aligned}$$

где \tilde{x} – выборочная средняя,

\bar{x} – генеральная средняя.

Предельная ошибка выборки $\Delta_{\tilde{x}}$ кратна средней ошибке $\mu_{\tilde{x}}$ с *коэффициентом кратности t* (называемым также коэффициентом доверия):

$$\Delta_{\tilde{x}} = t(P)\mu_{\tilde{x}}$$

Коэффициент кратности t зависит от значения *доверительной вероятности P*, гарантирующей вхождение генеральной средней в интервал $\tilde{x} \pm \Delta_{\tilde{x}}$, называемый *доверительным интервалом*.

Наиболее часто используемые доверительные вероятности P и соответствующие им значения t задаются следующим образом (табл. 14):

Таблица 14

Доверительная вероятность P	0,683	0,866	0,954	0,988	0,997	0,999
Значение t	1,0	1,5	2,0	2,5	3,0	3,5

По условию Задания 2 выборочная совокупность насчитывает 30 фирм, выборка 10% механическая, следовательно, *генеральная совокупность включает*

300 фирм. Выборочная средняя \bar{x} , дисперсия σ^2 определены в Задании 1 (п. 3). Значения параметров, необходимых для решения задачи, представлены в табл. 15:

Таблица 15

P	t	n	N	\bar{x}	σ^2
0,954	2	30	300	35,33	49,4722

Рассчитаем среднюю ошибку выборки:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{49,4722}{30} \cdot (1 - 0,10)} = 1,22 \approx 1 \text{ чел}$$

Рассчитаем предельную ошибку выборки:

$$\Delta_{\bar{x}} = 2 \cdot 1,22 = 2,44 \approx 2 \text{ чел.}$$

Определим доверительный интервал для генеральной средней:

$$35,33 - 2,44 \leq \bar{x} \leq 35,33 + 2,44$$

$$32,89 \leq \bar{x} \leq 37,77$$

или

$$33 \text{ чел.} \leq \bar{x} \leq 38 \text{ чел.}$$

Вывод. На основании проведенного выборочного обследования с вероятностью 0,954 можно утверждать, что для генеральной совокупности фирм средняя величина среднесписочной численности менеджеров находится в пределах от 33 до 38 человек.

2. Определение ошибки выборки для доли фирм со среднесписочной численностью менеджеров 40 человек и более, а также границ, в которых будет находиться генеральная доля

Доля единиц выборочной совокупности, обладающих тем или иным заданным свойством, выражается формулой

$$w = \frac{m}{n},$$

где m – число единиц совокупности, обладающих заданным свойством;

n – общее число единиц в совокупности.

Для *собственно-случайной* и *механической выборки с бесповторным способом отбора* предельная ошибка выборки Δ_w доли единиц, обладающих заданным свойством, рассчитывается по формуле

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

где w – доля единиц совокупности, обладающих заданным свойством;

$(1-w)$ – доля единиц совокупности, не обладающих заданным свойством,

N – число единиц в генеральной совокупности,

n – число единиц в выборочной совокупности.

Предельная ошибка выборки Δ_w определяет границы, в пределах которых будет находиться генеральная доля p единиц, обладающих исследуемым признаком:

$$w - \Delta_w \leq p \leq w + \Delta_w$$

По условию Задания 3 исследуемым свойством фирм является *равенство или превышение среднесписочной численности менеджеров величины 40 человек*.

Число фирм с данным свойством определяется из табл. 3 (графа 3):

$$m=7$$

Рассчитаем выборочную долю:

$$w = \frac{7}{30} = 0,233$$

Рассчитаем предельную ошибку выборки для доли:

$$\Delta_w = 2 \sqrt{\frac{0,233(1 - 0,233)}{30} \left(1 - \frac{30}{300}\right)} = 0,147$$

Определим доверительный интервал генеральной доли:

$$0,233 - 0,147 \leq p \leq 0,233 + 0,147$$

$$0,086 \leq p \leq 0,380$$

или

$$8,6\% \leq p \leq 38\%$$

Вывод. С вероятностью 0,954 можно утверждать, что в генеральной совокупности фирм региона доля фирм со среднесписочной численностью менеджеров 40 человек и более будет находиться в пределах от 8,6% до 38%.

Практическое задание по вариантам

(номер варианта получить у преподавателя)

t – номер вашего варианта

Исходные данные:

Имеются следующие выборочные данные по 30-ти однотипным предприятиям одного из регионов РФ, характеризующие деятельность фирм за исследуемый период (выборка 20%-ная механическая):

Таблица 1

Номер предприятия п/п	Вложения в рекламу, млн. руб	Сумма прибыли, млн руб.	Номер предприятия п/п	Вложения в рекламу, млн. руб	Сумма прибыли, млн руб.
1	$1,0*0,5t$	$25,1+t$	16	$3,4*0,5t$	$19,9+t$
2	$3,0*0,5t$	$3,0+t$	17	$5,0*0,5t$	$23,5+t$
3	$4,0*0,5t$	$25,4+t$	18	$5,5*0,5t$	$33,2+t$
4	$5,2*0,5t$	$15,1+t$	19	$9,0*0,5t$	$23,7+t$
5	$8,1*0,5t$	$27,2+t$	20	$9,2*0,5t$	$25,2+t$
6	$4,5*0,5t$	$28,3+t$	21	$4,7*0,5t$	$34,5+t$
7	$3,5*0,5t$	$9,0+t$	22	$6,7*0,5t$	$30,9+t$
8	$6,0*0,5t$	$11,0+t$	23	$10,2*0,5t$	$32,8+t$
9	$7,1*0,5t$	$33,0+t$	24	$11,0*0,5t$	$41,0+t$
10	$3,3*0,5t$	$10,0+t$	25	$8,0*0,5t$	$33,0+t$
11	$8,2*0,5t$	$38,2+t$	26	$9,0*0,5t$	$36,0+t$
12	$6,0*0,5t$	$24,2+t$	27	$8,5*0,5t$	$42,0+t$
13	$7,8*0,5t$	$25,7+t$	28	$11,1*0,5t$	$44,1+t$
14	$6,7*0,5t$	$32,1+t$	29	$9,2*0,5t$	$48,0+t$
15	$6,3*0,5t$	$26,0+t$	30	$8,4*0,5t$	$43,1+t$

Цель статистического исследования - анализ совокупности предприятий по признакам *Вложения в рекламу* и *Сумма прибыли*, включая:

- изучение структуры совокупности по признаку *Вложения в рекламу*;
- выявление наличия корреляционной связи между признаками *Вложения в рекламу* и *Сумма прибыли* предприятий, установление направления связи и оценка её тесноты;
- применение выборочного метода для определения статистических характеристик генеральной совокупности фирм.

Задания 1 (Занятие 1,2)

По исходным данным (табл. 1) необходимо выполнить следующее:

1. Построить статистический ряд распределения предприятий по объемам ***вложения в рекламу***, образовав ***пять*** групп с равными интервалами.
2. Графическим методом и путем расчетов определить значения *моды* и *медианы* полученного ряда распределения.

3. Рассчитать характеристики ряда распределения: *среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации.*
4. Вычислить *среднюю арифметическую* по исходным данным (табл. 1), сравнить её с аналогичным показателем, рассчитанным в п. 3 для интервального ряда распределения. Объяснить причину их расхождения.

Сделать выводы по результатам выполнения Задания 1.

Задание 2 (Занятие 3,4)

По исходным данным (табл. 1) с использованием результатов выполнения Задания 1 необходимо выполнить следующее:

2. Установить наличие и характер корреляционной связи между признаками *Вложения в рекламу* и *Сумма прибыли*, образовав пять групп с равными интервалами по каждому из признаков, используя методы:
 - а) аналитической группировки;
 - б) корреляционной таблицы.
2. Измерить тесноту корреляционной связи, используя *коэффициент детерминации и эмпирическое корреляционное отношение.*

Сделать выводы по результатам выполнения Задания 2.

Задание 3 (Занятие 5)

По результатам выполнения Задания 1 с вероятностью 0,954 необходимо определить:

- 3) ошибку выборки для средней величины *Вложения в рекламу*, а также границы, в которых будет находиться генеральная средняя.
- 4) ошибку выборки доли предприятий *Вложениями в рекламу* более 3 млн. руб, а также границы, в которых будет находиться генеральная доля предприятий.