

Владимирский государственный университет

**Методические указания
к практическим занятиям
по курсу «Статистика»**

Владимир 2003

Министерство образования Российской Федерации
Владимирский государственный университет
Кафедра экономики и управления на предприятиях

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ
ПО КУРСУ «СТАТИСТИКА»**

Составители:
Е.М. МАРЧЕНКО
П.Н. ЗАХАРОВ

Владимир 2003

УДК 657.62.622

Рецензент
Кандидат экономических наук, доцент
Владимирского государственного университета
Н.В. Андреева

Печатается по решению редакционно-издательского совета
Владимирского государственного университета

Методические указания к практическим занятиям по курсу «Статистика» / Владим. гос. ун-т; Сост.: Е.М. Марченко, П.Н. Захаров. Владимир, 2002. 36 с.

Изложена методика решения, приведены примеры расчета типовых задач, даны задания для самостоятельной работы студентов по основным разделам общей теории статистики: сводка и группировка статистических данных, средние величины, показатели вариации, выборочное наблюдение, корреляционно-регрессионный анализ.

Предназначены для студентов 2-го курса специальностей 060811 экономика и управление строительством, 060815 экономика и управление на предприятиях городского хозяйства, 060817 экономика и управление туризмом и гостиничным хозяйством.

Табл. 7. Библиогр.: 5 назв.

УДК 657.62.622

ВВЕДЕНИЕ

Качество управления неразрывно связано с обоснованностью и достоверностью статистических показателей, на основе анализа которых и принимаются управленческие решения на уровне предприятия, региона, отрасли или национальной экономики в целом.

Общая теория статистики разрабатывает принципы и методы сбора, обработки и анализа общественно-экономических явлений. Социально-экономическая статистика формирует систему показателей, с помощью которых оценивается уровень развития производства, общества.

Статистика дает знания и навыки проведения статистических исследований, необходимые для будущей работы как специалистов, так и руководителей предприятий любой отрасли.

Методические указания к практическим занятиям по курсу «Статистика» переработаны в соответствии с учебным планом и предназначены для студентов экономических специальностей 060811, 060815, 060817.

Методические указания включают пять глав, каждая из которых состоит из краткой теоретической части, разбора типовых задач, вопросов для контроля и задач для самостоятельного решения.

Материал каждой главы разбит на 4 раздела, которые обозначены символами:

Общие положения	–	
Разбор типовых задач	–	
Вопросы для контроля	–	
Задачи для самостоятельной работы	–	

СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ



Аналитическими называют такие группировки, которые позволяют установить и изучить связь между результирующими и факторными признаками единиц однотипной совокупности при параллельном сопоставлении обобщенных значений признаков по группам.

Признаки, оказывающие влияние на другие, называются факторными, а признаки, значение которых изменяется под влиянием других признаков, – результирующими. Основанием группировки, как правило, служит признак-фактор, а по результирующим признакам рассчитываются групповые средние. По изменению величины групповых средних делают вывод о наличии связи между признаками.

Допустима разбивка совокупности на любое число групп. При этом следует руководствоваться следующими правилами:

- единицы в образованных группах должны быть однородны по группировочному признаку;
- группы по величине группировочного признака должны существенно отличаться друг от друга;
- количество единиц в группах должно быть достаточным для расчета надежных характеристик;
- групповые средние должны последовательно уменьшаться или последовательно увеличиваться.

Группировка может осуществляться с равными и неравными интервалами. Технически удобнее иметь дело с равными интервалами. В этом случае величина интервала определяется по формуле:

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

где X_{\max} – наибольшее значение признака, X_{\min} – наименьшее значение признака, n – число образуемых групп.

Интервалы могут быть замкнутыми, если указаны верхняя и нижняя границы, и открытыми, если указана одна из границ. При группировке с неравными интервалами желательно образование групп замкнутого типа с целью повышения точности расчетов. Величина открытого интервала определяется исходя из величин закрытых интервалов.

Если количественные группировочные признаки представлены целочисленными значениями (численность работников, количество членов семьи, количество станков), то верхнюю границу предыдущей группы и нижнюю последующей возможно обозначать различно с расхождением на единицу. Например, 1 – 10, 11 – 20. Если группировочный признак непрерывный, дробный, то различие границ может составлять часть единицы либо ноль, т.е. границы совпадают, например, 1 – 10, 10 – 20. При совпадении границ интервалов граничное значение (в нашем примере равно 10) включается в тот интервал, где оно совпадает с величиной верхней границы интервала, т.е. 1 – 10. При наличии открытых интервалов ориентируются по последнему интервалу. Если указано «20 и более», то 20 входит в данный интервал, если «более 20» – значение 20 входит в предыдущий интервал.

Задача № 1



По приведенным данным о рынке недвижимости (табл. 1) осуществите группировку квартир по районам города.

Таблица 1

Предложение двухкомнатных квартир по районам города

Тип квартиры	Район	Жил. площадь	Кухня	Сан. узел	Цена, тыс. руб.
См	Центральный	32,0	6,0	с	150,0
Из	Центральный	27,0	8,0	р	210,0
См	Железнодорожный	31,0	6,0	с	110,0
Из	Железнодорожный	32,0	8,0	р	240,0
См	Железнодорожный	29,0	6,0	с	130,0
Из	Железнодорожный	25,0	6,0	с	110,0
Из	Дзержинский	31,0	6,0	с	95,0
Из	Дзержинский	27,0	7,0	р	100,0
См	Дзержинский	25,0	6,0	р	80,0
Из	Дзержинский	28,0	7,0	р	95,0
Из	Кировский	32,0	6,0	с	70,0
Из	Кировский	24,0	8,0	с	95,0

Окончание табл. 1

Тип квартиры	Район	Жил. площадь	Кухня	Сан. узел	Цена, тыс. руб.
См	Кировский	28,0	7,0	р	90,0
Из	Кировский	31,0	6,0	с	80,0
Из	Кировский	32,0	6,0	с	80,0
Из	Калининский	36,0	13,0	р	300,0
См	Калининский	27,0	7,0	р	100,0
См	Калининский	30,0	6,0	р	95,0
Из	Калининский	31,0	9,0	р	120,0
См	Октябрьский	31,0	9,0	р	120,0
Из	Октябрьский	31,0	9,0	р	110,0
Из	Октябрьский	33,0	9,0	р	120,0
Из	Октябрьский	31,0	9,0	р	110,0

Примечание. См – квартиры со смежными комнатами, Из – с изолированными.

Решение

Для анализа структуры предложения квартир по районам города выполним группировку квартир.

При проведении любой группировки среди прочих решаются следующие задачи: выбор группировочного признака (основания группировки) и определение числа выделяемых групп.

В соответствии с условием задачи в основу группировки положим признак «район города». Так как группировочный признак атрибутивный, то количество групп совпадает с числом районов города.

После того как выбран группировочный признак, намечено число групп и образованы сами группы, необходимо отобрать показатели, которые характеризуют группы, и определить их величины по каждой группе. В качестве показателей, характеризующих структуру, рассчитаем количество и среднюю цену квартир в каждом районе и удельный вес района в общем предложении квартир.

Показатели, характеризующие предложение квартир по районам, разнесем по шести указанным группам и подсчитаем групповые итоги. Результаты группировки представим в таблице (табл. 2).

Таблица 2

Группировка квартир по районам города

Район	Количество квартир	Удельный вес района	Средняя цена квартиры, млн руб.
Дзержинский	4	17,39	92,5
Железнодорожный	4	17,39	122,5
Калининский	4	17,39	153,8
Кировский	5	21,74	83,0
Октябрьский	4	17,39	115,0
Центральный	2	8,70	180,0
Итого	23	100,00	122,2

Задача № 2

Распределите районы Н-ской области по величине розничного товарооборота за 1995 г. Число групп задайте равным 5.

Район	Розничный товарооборот тыс. руб.	Район	Розничный товарооборот тыс. руб.
1 Баганский	31331	16 Куйбышевский	36775
2 Барабинский	56440	17 Маслянинский	47248
3 Болотинский	99212	18 Мошковский	92955
4 Венгеровский	34088	19 Новосибирский	178291
5 Доволенский	43520	20 Ордынский	68865
6 Здвинский	38196	21 Сузунский	60674
7 Искимитский	208492	22 Северный	9767
8 Красноозерский	51387	23 Татарский	23944
9 Купинский	65680	24 Тогучинский	127725
10 Карганский	82972	25 Убинский	24559
11 Колыванский	45561	26 Усть-Тарский	21946
12 Коченевский	137445	27 Чановский	44876
13 Кочковский	28970	28 Черепановский	117021
14 Карасукский	104518	29 Чулымский	36637
15 Кыштовский	21253	30 Чистоозерный	33775

Решение

Определим величину равного интервала:

$$i = \frac{208492 - 9767}{5} = \frac{198725}{5} = 39745.$$

Результаты представим в табличном виде (табл. 3).

Таблица 3

Группировка районов по объему товарооборота

Группы районов	Число районов в группе	Районы	Средний размер товарооборота по группе, тыс. руб.
9767 – 49512	16	1, 4, 5, 6, 11, 13, 15, 16, 17, 22, 23, 25, 26, 27, 29, 30	32652,8
49512 – 89257	6	2, 8, 9, 10, 20, 21	64336,3
89257 – 129002	5	3, 14, 18, 24, 28	108286,2
129002 – 168747	1	12	137445,0
168747 – 208492	2	7, 19	193391,5
Итого	30		65804,1



1. Какая группировка называется аналитической?
2. Перечислите основные правила проведения группировки.
3. Какие виды интервалов используются при группировке дискретных и непрерывных величин?
4. Как осуществить перегруппировку факторного признака?

Задача № 1



По группе грузовых автотранспортных предприятий города имеется следующая информация за отчетный год:

Предприятие	Грузооборот, млн т·км	Сумма затрат на перевозки, тыс. руб.
1	62	29140
2	40	22040
3	38	21660
4	25	14625
5	15	9900

Пред- приятие	Грузооборот, млн т·км	Сумма затрат на перевозки, тыс. руб.
6	30	17100
7	52	25272
8	27	30800
9	47	26790
10	24	14160
11	18	11700
12	58	27750
13	44	22000
14	23	13317
15	32	17280
16	20	12000

Необходимо:

1) произвести группировку грузовых автотранспортных предприятий по размеру грузооборота, выделив следующие группы: до 20 млн т·км; 20 – 40 млн т·км; 40 млн т·км и более;

2) по каждой группе определить: число предприятий, общий объем грузооборота, общую сумму затрат на перевозки, среднюю величину затрат на 10 т·км;

3) представить решение в форме статистической таблицы;

4) сформулировать вывод.

Задача № 2

По отдельным бригадам строительной организации имеются следующие данные за август:

Показатель	Бригада							
	1	2	3	4	5	6	7	8
Объем работ, тыс. руб.	819	1296	1340	1008	1468	1772	720	1904
Численность рабочих, чел.	16	24	25	21	27	32	15	34

Необходимо:

1) для выявления зависимости объема работ от числа рабочих, занятых в строительных бригадах, произвести группировку бригад по численности рабочих, выделив три группы с равными интервалами;

2) на основе выполненной группировки построить статистическую таблицу;

3) сформулировать вывод.

Задача № 3

По 12 партиям деталей, обрабатываемых рабочими производственного участка, имеются следующие данные:

Партия	Число операций при обработке одной детали	Число деталей в партии	Время на обработку всей партии, ч
1	2	12	3,86
2	3	16	1,97
3	3	4	1,83
4	4	12	8,10
5	5	20	4,40
6	5	8	4,70
7	6	12	5,90
8	8	4	5,38
9	11	4	3,80
10	12	4	4,40
11	11	3	3,75
12	9	1	1,45

Необходимо:

- 1) выполнить аналитическую группировку;
- 2) результаты группировки представить в форме групповой таблицы;
- 3) сформулировать вывод.

РАСЧЕТ СРЕДНИХ ВЕЛИЧИН

 В экономических исследованиях применяются две категории средних: степенные средние и структурные средние. К степенным средним относятся средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая и средняя кубическая. Все перечисленные виды степенных средних могут быть рассчитаны на основе формулы:

$$\bar{x} = \sqrt[m]{\frac{\sum x^m f}{\sum f}},$$

где x – варианты, f – частота признака, m – показатель степени.

В табл. 4 приведены формулы средних для различных значений m .

Виды степенных средних величин

Значение m	Вид средней	Формула средней	
		простой	взвешенной
-1	Гармоническая	$\bar{x} = \frac{n}{\sum \frac{1}{x}}$	$\bar{x} = \frac{\sum xf}{\sum \frac{1}{x} fx}$
0	Геометрическая	$\bar{x} = \sqrt[n]{\prod x}$	$\bar{x} = \sqrt[n]{x_1^{f_1} x_2^{f_2}}$
1	Арифметическая	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum xf}{\sum f}$
2	Квадратическая	$\bar{x} = \sqrt{\frac{\sum x^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum x^2 f}{\sum f}}$
3	Кубическая	$\bar{x} = \sqrt[3]{\frac{\sum x^3}{\sum n}}$	$\bar{x} = \sqrt[3]{\frac{\sum x^3 f}{\sum f}}$

Представителями структурных средних являются мода и медиана. Мода в вариационном ряду – варианта, соответствующая максимальному значению частоты.

Определение моды интервального ряда начинается с нахождения модального интервала. Модальным интервалом считается интервал, соответствующий максимальной частоте. Значение моды определяется по формуле:

$$M_0 = x_{m_0} + \frac{(f_{m_0} - f_{m_0-1})}{(f_{m_0} - f_{m_0-1}) + (f_{m_0} - f_{m_0+1})},$$

где x – нижняя граница интервала модального; f_{m_0} – частота модального интервала; f_{m_0-1} – частота интервала, предшествующего модальному; f_{m_0+1} – частота интервала, следующего за модальным.

Медианой называется варианта, которая делит пополам ряд, расположенный в порядке убывания или возрастания признака. Для нахождения медианы определяют ее место в ряду, используя формулу

$$N_{me} = \frac{n+1}{2}.$$

Для нахождения медианного интервала необходимо определить полусумму частот ряда. С помощью кумуляты (накопленной частоты) можно

отыскать интервал, в который попадает полусумма частот. Значение медианы интервального ряда рассчитывается по формуле

$$Me = x_{me} + \frac{\frac{\sum f}{2} - S_{me-1}}{f_{me}},$$

где x – нижняя граница медианного интервала; S – кумулята интервала, предшествующего медианному; $\frac{\sum f}{2}$ – полусумма частот ряда; f_{me} – частота медианного интервала.

Задача



По следующим данным определить среднее модальное и медианное значения стажа работы рабочих участка.

Стаж работы	До 5 лет	5–10	10–15	15 и более
Количество рабочих	2	6	15	7
Кумулята	2	8	23	30

Решение:

Определяем общий стаж всех рабочих:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{2,5 \cdot 2 + 7,5 \cdot 6 + 12,5 \cdot 15 + 17,5 \cdot 7}{2 + 6 + 15 + 7} = 12,0.$$

Для каждого интервала предварительно вычислялось среднее значение признака как полусумма нижнего и верхнего значений интервала. При этом величина открытых интервалов приравнивается к величине примыкающих к ним соседних интервалов, например,

$$\bar{x}_1 = \frac{0+5}{2} = 2,5.$$

Для расчета моды используются значения модального интервала, соответствующего максимальной частоте 15:

$$M_0 = 10 + 5 \cdot \frac{15-6}{(15-6)+(15-7)} = 12,6.$$

Для расчета медианы сначала определяется медианный интервал по полусумме частот и кумуляте. Так как кумулята третьего интервала – 23, что превышает полусумму частот (15), то медианным принимается интервал 10 – 15. Определяем величину медианы:

$$M_e = 10 + 5 \cdot \frac{\frac{30}{2} - 8}{15} = 12,3.$$



1. Как определить медианный интервал?
2. О чем говорит показатель моды?
3. Что показывает значение медианы?
4. Что такое кумулята?
5. Как определить модальный интервал?

Задача № 1



По приведенным данным определите среднюю цену помидоров на рынках города за месяц:

Рынок	Продано помидоров на сумму, руб.	Цена за кг, руб.
1	3100	50
2	4600	48
3	5200	60

Задача № 2

Определите среднюю себестоимость 1 кВт·ч электроэнергии по следующим данным:

ТЭЦ	Количество выработанной энергии, млн кВт·ч	Себестоимость 1 кВт·ч, руб.
1	6000	490
2	1200	520
3	3800	420

Задача № 3

Портфель состоит из акций трех компаний. Их доходность равна соответственно 15, 18 и 20 %, а доля в портфеле – 20, 50 и 30. Какова средняя доходность портфеля? Как следует увеличить доходность портфеля по сравнению с существующей?

Задача № 4

Вычислите моду и медиану количественного состава семей города на основании следующего их распределения по числу совместно проживающих членов семьи:

Число членов семьи	2	3	4	5	6	7	Всего
Удельный вес, %	15	34	25	16	8	2	100

Задача № 5

По приведенным данным, характеризующим распределение студентов одного из факультетов по возрасту, определите моду и медиану.

Возраст студентов, лет	17	18	19	20	21	22	23	24	Всего
Число студентов	20	80	90	110	130	170	90	60	750

РАСЧЕТ ПОКАЗАТЕЛЕЙ ВАРИАЦИИ И ОШИБОК ВЫБОРКИ

 К показателям вариации относят размах вариации R , среднее линейное отклонение d , дисперсию σ^2 и среднее квадратическое отклонение σ , коэффициент вариации v . Наиболее часто используют дисперсию и среднеквадратическое отклонение, расчет которых ведется по формулам:

$$R = x_{\max} - x_{\min}, \quad \sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}, \quad \sigma = \sqrt{\sigma^2}.$$

Относительная величина вариации определяется коэффициентом вариации v и рассчитывается по формуле

$$v = \frac{\sigma}{\bar{x}} 100\%,$$

Коэффициент показывает, какую долю средней составляет вариация признака в совокупности. При v более 33,3 % средняя величина считается ненадежной, нетипичной для данной выборки.

Формула расчета средней (стандартной) ошибки выборки имеет вид:

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

где n – объем выборки или число наблюдений.

Предельная ошибка выборки равна

$$\Delta = t \cdot \mu,$$

где t – коэффициент доверия.

Значения доверительного интервала находят с определенной долей вероятности, соответствующей значениям t . Величины t и вероятности $P(t)$ определяют по специальным статистическим таблицам нормального распределения.

$$\tilde{x} - \Delta \leq \tilde{x} \leq \tilde{x} + \Delta.$$

По результатам расчета доверительного интервала можно судить с определенной вероятностью о том, в каких пределах может изменяться средняя генеральной совокупности, если известна средняя выборки.

Задача № 1



Удельный вес основных рабочих в трех цехах предприятия составил 80, 75 и 90 % от общей численности рабочих.

Определить дисперсию и среднеквадратическое отклонение доли основных рабочих по предприятию в целом, если численность всех рабочих трех цехов составила соответственно 100, 200 и 150 человек.

Решение

Рабочие предприятия подразделяются на две группы: основные и ремонтно-вспомогательные рабочие.

Общая численность основных рабочих по предприятию в целом составит:

$$\sum n_0 = 0,8 \cdot 100 + 0,75 \cdot 200 + 0,9 \cdot 150 = 365 \text{ человек.}$$

Доля основных рабочих по предприятию

$$p = \frac{\sum n_0}{\sum n} = \frac{365}{100 + 200 + 150} = \frac{365}{450} = 0,811.$$

Дисперсия альтернативного признака

$$\sigma^2 = p(1-p) = 0,811(1-0,811) = 0,1533.$$

Среднеквадратическое отклонение доли основных рабочих по предприятию в целом

$$\sigma = \sqrt{pq} = \sqrt{0,1533} = 0,3915.$$

Задача № 2

Дисперсия признака равна 600. Объем совокупности равен 10. Сумма квадратов индивидуальных значений признака равна 6250.

Найти среднюю величину.

Решение

Средняя величина признака:

$$\bar{x} = \sqrt{\bar{x}^2 - \sigma^2} = \sqrt{625 - 600} = 5.$$



1. Какие показатели характеризуют вариацию признака?
2. С какой целью рассчитываются показатели вариации?
3. Какой показатель отражает абсолютную меру отклонения от средней величины, а какой – относительную величину?
4. С какой целью осуществляется выборочное наблюдение?
5. О чем говорит величина доверительного интервала?

Задача № 1



По приведенным данным, характеризующим распределение студентов одного из факультетов по возрасту, определите дисперсию, среднее линейное отклонение, относительные показатели вариации возраста студентов.

Возраст студентов, лет	17	18	19	20	21	22	23	24	Всего
Число студентов	20	80	90	110	130	170	90	60	750

Задача № 2

По переписи населения 1926 г. в России доля грамотных среди женщин составляла 46 %, а среди мужчин – 77 %. Определите общий (средний) процент грамотности всего населения и дисперсию этого показателя, если женщины составляли 53 % в общей численности населения.

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ



Расчет средней (стандартной) ошибки выборки зависит от способа отбора и процедуры выборки (табл. 5).

Таблица 5

Определение средней (стандартной) ошибки выборки

Тип выборки		Величина средней ошибки выборочной средней	Величина средней ошибки выборочной доли
по процедуре выборки	по способу отбора		
Собственно-случайная	Повторная	$\sqrt{\frac{\sigma^2}{n}}$	$\sqrt{\frac{\omega(1-\omega)}{n}}$
	Бесповторная	$\sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}$

Тип выборки		Величина средней ошибки выборочной средней	Величина средней ошибки выборочной доли
По процедуре выборки	По способу отбора		
Механическая	—	$\sqrt{\frac{\sigma^2(N-n)}{n(N-n)}}$	$\sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}$
Типичская пропорционально объему типичских групп	Повторная	$\sqrt{\frac{\bar{\sigma}^2}{n}}$	$\sqrt{\frac{\omega(1-\omega)}{n}}$
	Бесповторная	$\sqrt{\frac{\bar{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}$
Серийная	Повторная	$\sqrt{\frac{\delta^2}{m}}$	$\sqrt{\frac{\delta_{\omega}^2}{m} \left(1 - \frac{m}{M}\right)}$
	Бесповторная	$\sqrt{\frac{\delta^2}{m} \left(1 - \frac{m}{M}\right)}$	$\sqrt{\frac{\delta_{\omega}^2}{m} \left(1 - \frac{m}{M}\right)}$

Примечание. Символ n – объем выборки или число наблюдений; m – число отобранных серий; δ^2 – межгрупповая (межсерийная) дисперсия; $\bar{\sigma}^2$ – средняя из внутригрупповых дисперсий; ω – доля единиц, обладающих признаком; δ_{ω}^2 – дисперсия (межгрупповая) альтернативного признака.

Формулы необходимого объема выборки для различных способов формирования выборочной совокупности могут быть выведены из соответствующих формул расчета средней ошибки выборки.

Задача № 1



В 50 туристических агентствах города проводится обследование среднемесячного количества реализованных путевок методом механического отбора. Определить необходимую численность выборки, на основе обследования которой с вероятностью 0,683 ошибка не будет превышать 4 путевок, если по данным пробного обследования дисперсия составляет 225.

Решение

Рассчитаем необходимый объем выборки:

$$n = \frac{1^2 \cdot 225 \cdot 50}{4^2 \cdot 50 + 1^2 \cdot 225} = \frac{11250}{1025} = 11.$$

Задача № 2

Для оценки потерь рабочего времени рабочих строительной организации с численностью 480 человек была проведена 25 %-я выборка методом собственно-случайного бесповторного отбора. Результаты показали, что 10 % обследуемых потеряли более 45 минут рабочего времени в день. Установить доверительный интервал доли рабочих, имеющих потери времени свыше 45 минут с вероятностью 0,683.

Решение

Объем выборочной совокупности составит: $n = 480 \cdot 0,25 = 120$ человек.

Рассчитаем предельную ошибку выборочной доли, имея в виду, что при $p = 0,683$ $t = 1$.

$$\Delta = t \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)} = 1 \cdot \sqrt{\frac{0,1(1-0,1)}{120} \left(1 - \frac{120}{480}\right)} = 0,024;$$
$$10 - 2,4 \leq p \leq 10 + 2,4;$$
$$7,6 \leq p \leq 12,4.$$



1. С какой целью осуществляется выборочное наблюдение?
2. Как рассчитать среднюю ошибку выборки для собственно-случайного бесповторного отбора данных?
3. По какой формуле рассчитывается предельная ошибка выборки?
4. О чем говорит величина доверительного интервала?

Задача № 1



На промышленном предприятии работают 200 бригад рабочих. Определить необходимое количество бригад для проведения выборочного обследования на предмет профессиональных заболеваний, если ошибка выборки не должна превышать 5 %, межсерийная дисперсия доли равна 225, а требуемая вероятность составляет 95,4 %.

Задача № 2

Что произойдет с предельной ошибкой выборки, если вероятность гарантирования ошибки: а) увеличить с 0,954 до 0,997; б) уменьшить 95,4 % до 68,3 %?

Задача № 3

Отделом технического контроля методом случайного бесповторного отбора было проверено 80 изделий, из которых 7 оказались бракованными. Можно ли с вероятностью 0,997 утверждать, что доля бракованных изделий во всей партии не превышает 9 %, если процент отбора равен 10 %?

Задача № 4

Обследование семейных бюджетов населения города на основе 10 %-й типической пропорциональной выборки показало, что в 35 бюджетах одиноких граждан доля расходов на оплату жилья составила 9 %, а в 115 семейных бюджетах – 6 % соответственно.

С вероятностью 68,3 % установить границы доли расходов на оплату жилья населением города.

КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ



Корреляционный анализ определяет тесноту связи между двумя признаками (при парной связи) или между результативным и множеством факторных признаков (при многофакторной связи). Регрессионный анализ заключается в определении аналитического выражения связи, характеризующей изменение результирующего признака под влиянием одного или нескольких независимых факторов.

Оценка параметров уравнения регрессии осуществляется методом наименьших квадратов, в соответствии с которым параметры модели находятся на основе минимальной суммы квадратов отклонений эмпирических значений результирующего признака и теоретических значений, рассчитанных по уравнению регрессии.

Система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов приведена в табл. 6.

Таблица 6

Системы нормальных уравнений

Вид зависимости	Уравнение связи	Система нормальных уравнений
Линейная	$y = a_0 + a_1 x$	$na_0 + a_1 \sum x = \sum y$ $a_0 \sum x + a_1 \sum x^2 = \sum xy$
Параболическая второго порядка	$y = a_0 + a_1 x + a_2 x^2$	$na_0 + a_1 \sum x + a_2 \sum x^2 = \sum y$ $a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum yx^2$ $a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum xy$
Гиперболическая	$y = a_0 + a_1 \frac{1}{x}$	$a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2} = \sum \frac{y}{x}$ $na_0 + a_1 \sum \frac{1}{x} = \sum y$

Примечание. Символ n – объем совокупности.

Связь между тремя и более связанными между собой признаками называется множественной или многофакторной регрессией:

$$y_{1,2,\dots,k} = f(x_1, x_2, \dots, x_k).$$

Построение моделей множественной регрессии включает несколько этапов:

1. Выбор формы связи (уравнения регрессии).
2. Отбор факторных признаков.
3. Устранение мультиколлинеарности, то есть тесной связи между факторными признаками, включенными в модель (при $r_{xy} > 0,8$).
4. Обеспечение достаточного объема выборки.

В силу простоты и логичности экономической интерпретации в качестве теоретической кривой в основном используют линейные модели. Отбор факторов в модель обычно осуществляется способом шаговой регрессии, который заключается в последовательном включении в модель факторов с последующей проверкой их значимости. При этом проверяется, увеличился ли множественный коэффициент корреляции R . В случае увеличения коэффициента корреляции включение проверяемого фактора в модель обязательно. Большой объем наблюдений необходим для обеспечения адекватности модели.

Тесноту связи между признаками можно оценить с помощью ряда показателей (табл. 7).

Таблица 7

Показатели, характеризующие тесноту связи между признаками

Показатель	Размах вариации	Назначение	Формула расчета
Линейный коэффициент корреляции	$-1 \leq r \leq 1$	При наличии линейной зависимости между двумя признаками	$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$
Эмпирическое корреляционное отношение	$0 \leq \eta \leq 1$	При наличии линейной и нелинейной зависимостей	$N = \sqrt{\frac{s^2}{b^2}}$
Теоретическое корреляционное отношение	$0 \leq \eta \leq 1$	При наличии линейной и нелинейной зависимостей	$N = \sqrt{\frac{s^2}{b^2}}$

Показатель	Размах вариации	Назначение	Формула расчета
Множественный коэффициент корреляции	$0 \leq R \leq 1$	При наличии линейной связи между резульативным и несколькими факторными признаками и между каждой парой факторных признаков	$R_{y/x_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$
Частный коэффициент корреляции	$0 \leq r_{yx_1}(x_2)$	Определяет тесноту связи между двумя признаками при фиксированном значении других признаков	$r_{yx_1}(x_2) = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}$

Частный коэффициент эластичности определяется по формуле

$$\varepsilon_{x_i} = a_i \frac{\bar{x}_i}{\bar{y}_i}$$

и показывает, на сколько процентов в среднем изменится значение резульативного признака при изменении факторного признака на 1 %.

Значимость линейного коэффициента корреляции проверяется на основе t – критерия Стьюдента:

$$t_p = \sqrt{\frac{r^2}{1 - r^2}}(n - 2).$$

При $t_p > t_{kp}$ можно говорить о значимости линейного коэффициента корреляции и, следовательно, о статистической существенной зависимости между x и y .

Задача



По группе предприятий за отчетный год имеются следующие данные:

Пред- приятия	Годовая производитель- ность труда работников, тыс. руб.	Вооружен- ность труда основным капиталом, тыс. руб./чел.	Удельный вес оборудования в стоимости основного ка- питала	Теку- честь кадров, %	Интегральный показатель использования рабочего времени
1	360	15,2	0,39	9,1	0,96
2	298	12,8	0,29	10,1	0,80
3	328	13,8	0,34	5,0	0,84
4	330	14,0	0,36	7,0	0,86
5	366	16,3	0,47	9,0	0,98
6	316	12,6	0,28	4,0	0,83
7	334	13,2	0,32	12,0	0,87
8	300	12,9	0,29	6,5	0,84
9	314	13,1	0,33	8,0	0,81
10	320	12,5	0,28	7,0	0,85
11	362	15,7	0,40	8,05	0,97
12	332	13,5	0,34	5,0	0,83

На основании приведенных данных требуется:

- 1) составить уравнение множественной зависимости производительности труда, обосновав систему факторов, включенных в модель;
- 2) определить совокупный коэффициент корреляции и частные коэффициенты корреляции;
- 3) сопоставить роль различных факторов в формировании моделируемого показателя.

Решение

Результативный показатель – годовая производительность труда работника y .

Признаки-факторы:

- вооруженность труда основным капиталом – x_1 ;
- удельный вес оборудования в стоимости основного капитала – x_2 ;
- текучесть кадров – x_3 ;
- интегральный показатель использования рабочего времени – x_4 .

Для определения возможности включения факторов в модель строится матрица парных коэффициентов корреляции. Результаты расчета дали следующую матрицу:

	y	x_1	x_2	x_3	x_4
y	1	0,911	0,903	0,249	0,979
x_1	0,911	1	0,977	0,264	0,812
x_2	0,903	0,977	1	-0,009	0,710
x_3	0,249	0,264	-0,009	1	0,930
x_4	0,979	0,812	0,710	0,930	1

Цифры первой строки матрицы парных коэффициентов корреляции показывают, что фактор x_3 (текучесть кадров) не следует включать в модель, так как связь результативного показателя с ним слабая ($r_{x_3 y} = 0,249$). С остальными факторами связь тесная, и, если нет мультиколлинеарности, они могут быть включены в модель.

Сначала проверяется возможность включения в модель факторов x_1 и x_2 . В качестве критерия принимается соблюдение следующих неравенств:

$$r_{x_1 y} > r_{x_1 x_2};$$

$$r_{x_2 y} > r_{x_1 x_2}.$$

Фактически эти неравенства не соблюдаются, так как

$$0,911 < 0,977;$$

$$0,903 < 0,977.$$

Следовательно, в модель должен быть включен фактор x_1 , так как связь результативного показателя с ним более тесная ($r_{x_1 y} = 0,911$).

Далее проверяется возможность включения в модель факторов x_1 и x_4 на основе следующих неравенств:

$$r_{x_1 y} > r_{x_1 x_4};$$

$$r_{x_4 y} > r_{x_1 x_4}.$$

Фактически эти неравенства соблюдаются, так как

$$0,911 > 0,812;$$

$$0,979 > 0,812.$$

Таким образом, в модель множественной зависимости могут быть включены два фактора: x_1 и x_4 . Линейное уравнение имеет следующий вид:

$$\bar{y} = a + b_1 x_1 + b_4 x_4.$$

Система нормальных уравнений для нахождения параметров a , b_1 и b_4 следующая:

$$\begin{cases} \sum y = na + b_1 \sum x_1 + b_4 \sum x_4; \\ \sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_4 \sum x_4 x_1; \\ \sum yx_4 = a \sum x_4 + b_1 \sum x_1 x_4 + b_4 \sum x_4^2. \end{cases}$$

$$\begin{cases} 3960 = 12a + b_1 \cdot 165,6 + b_4 \cdot 10,44; \\ 54937,8 = a_1 \cdot 65,6 + b_1 \cdot 2303 + b_4 \cdot 144,85; \\ 3459,9 = a_1 \cdot 0,44 + b_1 \cdot 144,85 + b_4 \cdot 9,127. \end{cases}$$

Решение системы уравнений дает следующие значения параметров:
 $a=48,27$; $b_1=4,43$; $b_4=253,63$.

Модель зависимости производительности труда от факторов имеет следующий вид:

$$\bar{y} = 48,27 + 4,43x_1 + 253,63x_4.$$

Вспомогательная таблица для расчета ошибки модели

Предприятие	Годовая производительность труда работников, тыс. руб.	Теоретический уровень годовой производительности труда работника по уравнению связи \bar{y} , тыс. руб.	$y - \bar{y}$	$(y - \bar{y})^2$
1	360	359,1	+ 0,9	0,81
2	298	307,8	- 9,8	96,04
3	328	332,4	- 4,4	19,36
4	330	328,4	+ 1,6	2,56
5	366	369,0	- 3,0	9,00
6	316	314,6	+ 1,4	1,96
7	334	324,4	+ 9,6	92,16
8	300	318,9	- 18,5	342,25
9	314	311,7	+ 2,3	5,29
10	320	319,2	+ 0,8	0,64
11	362	363,8	- 1,8	3,24
12	332	328,3	+ 3,7	13,69
Итого				587,00

Средняя квадратическая ошибка уравнения:

$$\sigma = \sqrt{\frac{\sum (y - \bar{y})^2}{n-1}} = \sqrt{\frac{587}{12-3}} = 8,08;$$

$$\frac{\sigma}{\bar{y}} \cdot 100 = \frac{8,08}{330} \cdot 100 = 2,45 \%$$

Уравнение достаточно хорошо отражает взаимосвязь производительности труда и двух ее факторов.

Определим совокупный коэффициент корреляции:

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_4}^2 - 2r_{yx_1} r_{yx_4} r_{x_1 x_4}}{1 - r_{x_1 x_4}^2}};$$

$$R = \sqrt{\frac{0,911^2 + 0,979^2 - 2 \cdot 0,911 \cdot 0,979 \cdot 0,812}{1 - 0,812^2}} = \sqrt{0,9979} = 0,99$$

(парные коэффициенты корреляции взяты из матрицы).

Близость совокупного коэффициента корреляции к единице означает: роль не учтенных в модели факторов ничтожна, и есть основания считать, что параметры регрессионной модели отражают степень эффективности включенных в нее факторов.

Частные коэффициенты корреляции следующие:

а) частный коэффициент корреляции между результативным признаком y и фактором x_1 при элиминировании фактора x_4 :

$$r_{yx_1}(x_4) = \frac{r_{yx_1} - r_{yx_4} \cdot r_{x_1 x_4}}{\sqrt{(1 - r_{yx_4}^2) \cdot (1 - r_{x_1 x_4}^2)}} = \frac{0,979 - 0,911 \cdot 0,812}{\sqrt{(1 - 0,911^2)(1 - 0,812^2)}} = 0,967;$$

б) частный коэффициент корреляции между результативным признаком y и фактором x_4 при элиминировании фактора x_1 :

$$r_{yx_4}(x_1) = \frac{r_{yx_4} - r_{yx_1} \cdot r_{x_1 x_4}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1 x_4}^2)}} = \frac{0,979 - 0,911 \cdot 0,812}{\sqrt{(1 - 0,911^2)(1 - 0,812^2)}} = 0,987.$$

Сопоставление полученных ранее частных коэффициентов корреляции с вычисленными ранее парными коэффициентами корреляции подтверждает наличие тесной связи между результативным и факторными признаками.

Для сравнения роли отдельных факторов в формировании показателя производительности труда определяются коэффициенты эластичности:

а) для фактора x_1 $\varepsilon_1 = b_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 4,43 \cdot \frac{13,8}{330} = 0,185$;

$$\bar{x}_1 = \frac{165,6}{12} = 13,8; \bar{y} = \frac{3960}{12} = 330;$$

б) для фактора x_4 $\varepsilon_4 = b_4 \cdot \frac{\bar{x}_4}{\bar{y}} = 253,63 \cdot \frac{0,87}{330} = 0,669$;

$$\bar{x}_4 = \frac{10,44}{12} = 0,87.$$

Следовательно, при увеличении вооруженности труда основным капиталом на 1 % производительность труда возрастает лишь на 0,185 %. Увеличение показателя использования рабочего времени на 1 % повлечет рост производительности труда на 0,669 %.



1. Какие характеристики используются для проверки надежности и типичности средних величин?
2. Как исключить коррелируемость независимых признаков?
3. С какой целью рассчитывают уравнение регрессии?
4. Для чего нужны стандартные коэффициенты регрессии?
5. О чем говорит коэффициент множественной детерминации?
6. При выполнении какого условия можно говорить об устойчивости коэффициентов корреляции?

Задача № 1



Зависимость между объемом произведенной продукции и балансовой прибылью по 10 предприятиям одной из отраслей промышленности характеризуется следующими данными:

Объем реализованной продукции, млн руб.	Балансовая прибыль, млн руб.
491,8	133,8
483,0	124,1
481,7	62,4
478,7	62,9
476,9	51,4
475,2	72,4
474,4	99,3
4589,5	49,0
452,9	104,0
446,5	116,1

Определить вид корреляционной зависимости, построить уравнение регрессии, рассчитать параметры уравнения, вычислить тесноту связи. Объяснить полученные характеристики.

Задача № 2

По данным об объеме услуг по месяцам и уровне механизации труда предприятий городского хозяйства составить уравнение регрессии

и рассчитать показатели тесноты связи между факторами. Дать экономическую интерпретацию полученных показателей.

Предприятие	Месячный объем услуг, млн руб.	Уровень механизации	Предприятие	Месячный объем услуг, млн.руб.	Уровень механизации
1	90	95	11	65	77
2	77	64	12	95	90
3	80	77	13	90	85
4	90	93	14	91	90
5	91	64	15	100	99
6	100	98	16	110	100
7	101	99	17	109	98
8	105	100	18	107	89
9	110	100	19	89	95
10	99	96	20	98	99

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Общая теория статистики: Практикум. – М.: ИНФРА-М, 1999. – 139 с.
2. Сборник задач по общей теории статистики: Учеб. пособие / Под ред. Л.К. Серга. – М: Филинь, 1999. – 360 с.
3. Практикум по теории статистики: Учеб. пособие / Под ред. Р.А. Шмойловой. – М: Финансы и статистика, 2003. – 416 с.
4. Ефимова М.Р., Ганченко О.И., Петрова Е.В. Практикум по общей теории статистики: Учеб. пособие. – М.: Финансы и статистика, 2003. – 336 с.
5. Минашкин В.Г., Гусынин А.Б., Садовникова Н.А., Шмойлова Р.А. Теория статистики: Учеб.-практ. пособие / Моск. гос. ун-т экономики, статистики и информатики. – М.: МЭСИ, 1999. – 260 с.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ	4
РАСЧЕТ СРЕДНИХ ВЕЛИЧИН	10
РАСЧЕТ ПОКАЗАТЕЛЕЙ ВАРИАЦИИ И ОШИБОК ВЫБОРКИ	14
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ	16
КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ	19
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	27

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРАКТИЧЕСКИМ
ЗАНЯТИЯМ ПО КУРСУ
«СТАТИСТИКА»

Составители:

МАРЧЕНКО Елизавета Маратовна
ЗАХАРОВ Павел Николаевич

Ответственный за выпуск – зав. кафедрой профессор В.Ф. Архипова

Редактор И.В. Усачева
Корректор Е.В. Афансьева
Компьютерная верстка Д.Н. Ях

ЛР № 020275. Подписано в печать 15.12.03.

Формат 60x84/16. Бумага для множит. техники. Гарнитура Таймс.
Печать на ризографе. Усл. печ. л. 1,86. Уч.-изд. л. 1,95. Тираж 200 экз.
Заказ

Редакционно-издательский комплекс
Владимирского государственного университета.
600000, Владимир, ул. Горького, 87.