

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ
АРХИТЕКТУРНО-СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ»

**Методические указания для выполнения
самостоятельных практических работ по дисциплинам
«Статистика» и «Эконометрика»**

ИНГАСУ
2010

ББК 65.в 6 + 65.051я73

УДК 330.115+311(075)

М 54

Методические указания для выполнения самостоятельных практических работ по дисциплинам «Статистика» и «Эконометрика» / сост. А. В. Елесин, Н. Ю. Прокопенко ; Нижегород. гос. архитектур.-строит. ун-т.- Н. Новгород : ННГАСУ, 2010. – 64 с.

В методических указаниях рассмотрены некоторые основные сведения из теории, приведены примеры решения типовых задач, ряд формул, необходимых для понимания изучаемого материала, задания для контрольной работы, список рекомендуемой литературы по учебным дисциплинам «Статистика» и «Эконометрика». Данное пособие может быть также полезно для студентов, изучающих дисциплины «Количественные методы анализа данных», «Методы прогнозирования», «Математическое и имитационное моделирование» и др.

Работа подготовлена на кафедре информационных систем в экономике.

©Елесин А.В., Прокопенко Н.Ю., 2010

©ННГАСУ, 2010

Содержание

Методические указания для выполнения самостоятельной практической работы по дисциплине «Статистика».....	4
Тема 1. Выборочный метод.....	5
Тема 2. Статистические оценки параметров распределения.....	10
Тема 3. Элементы теории корреляции.....	15
Тема 4. Статистическая проверка гипотез.....	23
Контрольные задания.....	30
Литература по дисциплине «Статистика».....	48
Методические указания для выполнения самостоятельной практической работы по дисциплине «Эконометрика».....	49
Литература по дисциплине «Эконометрика».....	64

Методические указания для выполнения самостоятельной практической работы по дисциплине «Статистика»

Перед тем как приступить к разбору решений типовых задач, изложенных в данной методической разработке, студентам рекомендуется изучить следующий теоретический материал:

Генеральная и выборочная совокупности. Виды выборок. Способы отбора. Статистическое распределение выборки. Полигон и гистограмма. Эмпирическая функция распределения. Статистические оценки параметров распределения. Несмещенные, эффективные и состоятельные оценки. Оценка генеральной средней по выборочной средней. Устойчивость выборочных средних. Выборочная дисперсия. Оценка генеральной дисперсии по исправленной выборочной. Точность оценки, доверительная вероятность (надежность). Доверительный интервал.

Функциональная, статистическая и корреляционная зависимость. Две основные задачи теории корреляции. Отыскание параметров выборочного уравнения прямой линии регрессии. Свойства выборочного коэффициента корреляции. Групповая и общая средняя. Групповая, внутригрупповая, межгрупповая дисперсии.

Статистические гипотезы. Нулевая и конкурирующая гипотезы. Простая и сложная гипотезы. Ошибки первого и второго рода, уровень значимости. Статистический критерий проверки нулевой гипотезы. Наблюдаемое значение критерия. Критерий согласия. Примеры критериев согласия. Проверка гипотез о распределении Пуассона и о нормальном распределении совокупности с помощью критерия согласия Пирсона. Методики вычисления теоретических частот распределения Пуассона и нормального распределения.

Тема 1. Выборочный метод.

Проведение экономических исследований связано с изучением свойств различных совокупностей однотипных объектов (людей, предприятий, товаров и т.п.). При этом каждый объект, входящий в состав совокупности, характеризуется некоторым числом – величиной изучаемого признака X . Для обозначения таких совокупностей вводится понятие *генеральной совокупности*.

Под *генеральной совокупностью* понимается вся совокупность однотипных объектов, которые изучаются в данном исследовании.

Пример генеральной совокупности – данные о доходах всех жителей какой-либо страны; о результатах голосования населения по какому-либо вопросу и т.д.

Однако на практике в большинстве случаев мы имеем дело только с частью возможных наблюдений, взятых из генеральной совокупности.

Выборка (выборочная совокупность) – это совокупность случайно отобранных объектов, составляющих лишь часть генеральной совокупности.

Объемом совокупности (выборочной или генеральной) называют число объектов этой совокупности.

В зависимости от способов отбора объектов из генеральной совокупности различают несколько типов выборок. Их типы, определения, свойства, примеры использования рекомендуется изучить самостоятельно.

Пусть из генеральной совокупности извлечена выборка, причем x_1 наблюдалось n_1 раз, x_2 – n_2 раз, ..., x_k – n_k раз и $\sum_{i=1}^k n_i = n$ – объем выборки.

Наблюдаемые значения x_i называют вариантами, а последовательность вариантов, записанных в возрастающем порядке – вариационным рядом.

Числа наблюдений (n_i) называют частотами, а их отношения к объему выборки $\frac{n_i}{n} = w_i$ – относительными частотами.

Статистическим распределением выборки называют перечень вариант x_i и соответствующих им частот n_i или относительных частот w_i (причем сумма всех частот равна объему выборки, а сумма всех относительных частот равна 1).

x_i	x_1	x_2	x_k
n_i	n_1	n_2	n_k

или

x_i	x_1	x_2	x_k
w_i	w_1	w_2	w_k

Вариационный ряд, заданный в таком виде, называют *дискретным*. Геометрической характеристикой дискретного вариационного ряда является полигон частот.

Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)$, где x_i – варианты выборки, а n_i – соответствующие им частоты.

Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот. Для непрерывно распределенного признака весь интервал, в котором заключены все наблюдаемые значения признака, разбивают на ряд частичных интервалов длины h и находят n_i – сумму частот вариант попавших в i -й интервал. Такое распределение называют *интервальным вариационным рядом*.

Геометрической характеристикой интервального вариационного ряда является гистограмма частот.

Гистограммой частот называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длины h , а высоты равны отношению $\frac{n_i}{h}$.

Пример1. Из большой группы предприятий одной из отраслей промышленности случайным образом отобрано 30, по которым получены показатели основных фондов в млн. руб.:

3; 4; 2; 3; 3; 6; 5; 2; 4; 7; 5; 5; 3; 4; 3; 2; 6; 7; 5; 4; 3; 4; 5; 7; 6; 2; 3; 6; 6; 4.

Составить дискретное статистическое распределение выборки, записать распределение относительных частот, построить полигон частот.

Решение. Различные значения признака запишем в порядке возрастания и под каждым из них запишем соответствующие частоты. Получим дискретное статистическое распределение выборки:

x_i	2	3	4	5	6	7
n_i	4	7	6	5	5	3

Проверка: сумма всех частот должна быть равна объему выборки:

$$n=4+7+6+5+5+3=30.$$

Найдем относительные частоты:

$$w_1 = \frac{4}{30} = 0,13; \quad w_2 = \frac{7}{30} = 0,23;$$

$$w_3 = \frac{6}{30} = 0,2; \quad w_4 = \frac{5}{30} = 0,17;$$

$$w_5 = \frac{5}{30} = 0,17; \quad w_6 = \frac{3}{30} = 0,1.$$

Составим таблицу распределения относительных частот:

x_i	2	3	4	5	6	7
w_i	0,13	0,23	0,2	0,17	0,17	0,1

Контроль: $\sum w_i = 0,13 + 0,23 + 0,2 + 0,17 + 0,17 + 0,1 = 1$.

Строим полигон частот. Для этого строим точки с координатами $(x_i; n_i): (2;4), (3;7), (4;6), (5;5), (6;5), (7;3)$ и соединяем их последовательно отрезками.



Рис.1. Полигон частот для дискретного распределения

Пример 2. Выборочно обследовано 26 предприятий легкой промышленности по валовой продукции. Получены следующие результаты в млн. руб.:

15,0; 16,4; 17,8; 18,0; 18,4; 19,2; 19,8; 20,2; 20,6; 20,6; 20,6;
 21,3; 21,4; 21,7; 22,0; 22,2; 22,3; 22,7; 23,0; 24,2; 24,2; 25,1;
 25,3; 26,0; 26,5; 27,1.

Составить интервальное распределение выборки с началом $x_0=15$ и длиной частичного интервала $h=2,5$. Построить гистограмму частот.

Решение. Для составления интервального распределения составим таблицу. В первой строке расположим в порядке возрастания интервалы, длина каждого из которых $h=2,5$. Во второй строке запишем количество значений признака в выборке, попавших в этот интервал (т.е. сумму частот вариант, попавших в соответствующий интервал).

Частичный интервал	15-17,5	17,5-20	20-22,5	22,5-25	25-27,5
Частота интервала	2	5	10	4	5

Объем выборки $n=2+5+10+4+5=26$.

Для построения гистограммы частот на оси абсцисс откладываем частичные интервалы; на каждом из них строим прямоугольники высотой $\frac{n_i}{h}$.

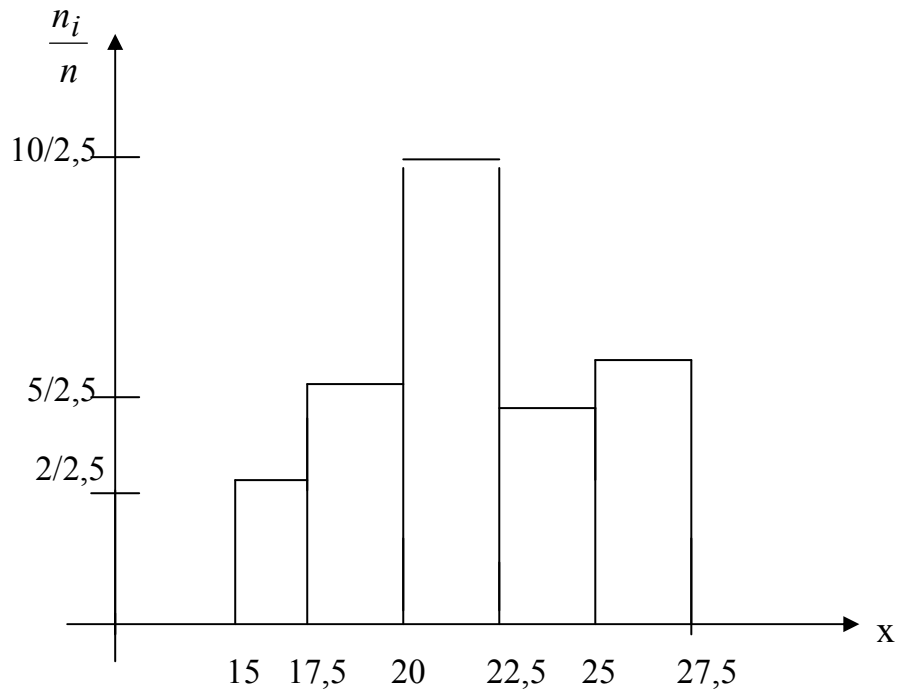


Рис.2. Гистограмма непрерывного распределения

Площадь каждого прямоугольника равна частоте интервала, на котором он построен. Сумма площадей этих прямоугольников равна объему выборки.

Тема 2. Статистические оценки параметров распределения

Пусть требуется изучить количественный признак генеральной совокупности. Располагая лишь выборочными значениями признака, можно оценить, а не определить точно, значения параметров; эти оценки будут случайными и меняться от выборки к выборке. Поэтому важно не только знать оценки параметров, определенные на основе выборочных данных, но и понимать меры их надежности.

Цель любого оценивания – получить как можно более точное значение неизвестной характеристики генеральной совокупности по данным выборочного наблюдения.

Статистической оценкой неизвестного параметра теоретического распределения называют функцию от наблюдаемых случайных величин.

В зависимости от способа выражения оценки делятся на точечные оценки, выражаемые одним числом, и интервальные оценки, определяющие числовой интервал, внутри которого может находиться оцениваемый параметр генеральной совокупности.

Генеральная совокупность характеризуется двумя сторонами:

- 1) видом распределения (например, равномерное, нормальное, Пуассоновское и т.д.); 2) параметрами распределения (например, математическое ожидание, среднее квадратическое отклонение и т.п.).

В связи с этим существует два класса оценок: оценки вида распределения и оценки параметров распределения.

Оценка Θ^* должна быть несмещенной, эффективной, состоятельной. Определения несмещенной, эффективной, состоятельной оценок рекомендуется изучить самостоятельно.

Несмещенной, состоятельной и эффективной оценкой генеральной средней (математического ожидания признака X генеральной

совокупности) является выборочная средняя $\overline{x_g}$ – среднее арифметическое значений признака в выборке:

$$\overline{x_g} = \frac{x_1 + x_2 + \dots + x_k}{n}, \quad (1)$$

где n – объем выборки, x_i – значение признака в выборке. Если результаты выборки представлены в виде дискретного распределения:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

то

$$\overline{x_g} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n} = \frac{\sum_{i=1}^k x_i n_i}{n}. \quad (2)$$

Состоятельной, смещенной оценкой генеральной дисперсии (дисперсия признака X генеральной совокупности) является выборочная дисперсия:

$$D_g = \overline{x^2} - (\overline{x_g})^2, \quad (3)$$

где $\overline{x^2} = \frac{\sum_{i=1}^k x_i^2 n_i}{n}$, $\overline{x_g} = \frac{\sum_{i=1}^k x_i n_i}{n}$.

Несмещенной, состоятельной оценкой генеральной дисперсии является исправленная выборочная дисперсия

$$S^2 = \frac{n}{n-1} D_g. \quad (4)$$

Пример 1. При изучении производительности труда X тыс. руб. на одного работника было обследовано 10 предприятий и получены следующие значения:

4,2; 4,8; 4,7; 5,0; 4,9; 4,3; 3,9; 4,1; 4,3; 4,8.

Определить выборочное среднее $\overline{x_g}$, выборочную дисперсию, исправленное среднее квадратическое отклонение.

Решение. По формуле (1) находим выборочную среднюю при $n=10$:

$$\bar{x}_e = \frac{4,2 + 4,8 + 4,7 + 5,0 + 4,9 + 4,3 + 3,9 + 4,1 + 4,3 + 4,8}{10} = \frac{45}{10} = 4,5 \text{ (тыс.руб)}$$

По формуле (3) найдем выборочную дисперсию. Для этого вычислим $\overline{x^2}$ и $(\bar{x}_e)^2$.

$$\overline{x^2} = \frac{4,2^2 + 4,8^2 + 4,7^2 + 5,0^2 + 4,9^2 + 4,3^2 + 3,9^2 + 4,1^2 + 4,3^2 + 4,8^2}{10} = 20,382$$

$$(\bar{x}_e)^2 = (4,5)^2 = 20,25.$$

$$D_e = \overline{x^2} - (\bar{x}_e)^2 = 20,382 - 20,25 = 0,132.$$

Исправленное среднее квадратическое отклонение:

$$S = \sqrt{S^2} = \sqrt{\frac{10}{10-1} \times 0,132} = \sqrt{0,147} \approx 0,383.$$

Смысл полученных результатов заключается в следующем. Величина \bar{x}_e характеризует среднее значение признака X в пределах рассматриваемой выборки. Средняя производительность труда для изученных предприятий составила $\bar{x}_e = 4,5$ тыс. руб. на одного работника. Исправленное среднее квадратическое отклонение S описывает абсолютный разброс значений показателя X и в данном случае составляет $S=0,383$ тыс. руб.

Если дано интервальное распределение выборки, то надо перейти к дискретному, взяв за значения вариант середины частичных интервалов.

Выборочные оценки являются приближенными. Чтобы с помощью статистических данных можно было сделать правильные выводы, нужно знать точность и надежность этих оценок.

Пусть Θ^* – статистическая оценка неизвестного параметра Θ . Надежностью (доверительной вероятностью) оценки Θ по Θ^* называют вероятность γ , с которой осуществляется неравенство $|\Theta - \Theta^*| < \Delta$.

Обычно надежность оценки задается наперед, причем в качестве γ берут число, близкое к единице. По надежности γ ищут такое число Δ , чтобы

$$P(|\Theta - \Theta^*| < \Delta) = \gamma. \quad (5)$$

Число Δ называют точностью оценки, или предельной ошибкой. Из равенства (5) следует, что

$$P(\Theta^* - \Delta < \Theta < \Theta^* + \Delta) = \gamma. \quad (6)$$

Интервал $(\Theta^* - \Delta, \Theta^* + \Delta)$ называется доверительным интервалом; он называется интервальной оценкой неизвестного параметра Θ .

Интервальной оценкой с надежностью γ математического ожидания $M(X) = a$ нормально распределенного признака X генеральной совокупности при известном среднем квадратическом отклонении $\sigma = \sqrt{D(X)}$ этого признака служит доверительный интервал

$$\bar{x}_g - \frac{\sigma \cdot t}{\sqrt{n}} \leq a \leq \bar{x}_g + \frac{\sigma \cdot t}{\sqrt{n}}, \quad (7)$$

где n – объем выборки, \bar{x}_g – выборочная средняя, t – значение аргумента функции Лапласа $\Phi(t)$, при котором $\Phi(t) = \frac{\gamma}{2}$, $\frac{\sigma \cdot t}{\sqrt{n}} = \Delta$ – точность оценки.

Пример 2. В ходе обследования банковских счетов была проведена случайная выборка записей по вкладам. Из выборки $n=100$ оказалось, что средний размер вклада составляет 1 837 д.е.; среднее квадратическое отклонение размера вклада равно 280 д.е. Найти с надежностью $\gamma=0,95$ доверительный интервал для среднего размера a вкладов по всем счетам, если известно, что размер вкладов распределен по нормальному закону.

Решение. По условию $\bar{x}_g = 1837$; $n=100$; $\sigma=280$; $\gamma=0,95$. По таблице

значений функции $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{z^2}{2}} dz$ находим t из условия

$\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475$, получаем $t=1,96$. По формуле (7) находим

доверительный интервал:

$$1837 - \frac{1,96 \times 280}{\sqrt{100}} \leq a \leq 1837 + \frac{1,96 \times 280}{\sqrt{100}},$$

$$1837 - 54,88 \leq a \leq 1837 + 54,88,$$

$$1782,12 \leq a \leq 1891,88.$$

Это означает, что с вероятностью, равной 0,95, можно утверждать, что средний размер вклада генеральной совокупности находится в пределах от 1 782,12 д.е. до 1 891,88 д.е. Интервал $\pm 54,88$ составляет примерно $\pm 3\%$ среднего размера вклада в выборке (1 837). Это не очень большое отклонение, поэтому среднее значение выборки можно считать надежной оценкой среднего значения генеральной совокупности. Однако существует вероятность, равная 0,05 того, что можно получить значение вне доверительного интервала.

Тема 3. Элементы теории корреляции

Различные экономические показатели не являются независимыми, а связаны между собой; например, цена какого-либо товара и величина спроса на этот товар, объем производства и прибыль фирмы, располагаемый доход и объем личного потребления, инфляция и безработица. Взаимосвязи показателей в экономике редко имеют простой функциональный вид, поскольку на интересующий нас показатель, кроме явно учитываемых факторов, влияет еще множество других, которые являются случайными. Поэтому одной из основных задач в экономических исследованиях является анализ зависимостей между переменными.

Пусть требуется оценить связь между переменными X и Y . Возникает два вопроса: 1) связаны ли между собой эти переменные; 2) какова теснота этой связи?

В качестве характеристики тесноты линейной связи между количественными признаками в выборке используется выборочный

коэффициент корреляции r_B :

$$r_B = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

Свойства выборочного коэффициента корреляции:

- 1) значения r_B заключены в промежутке от -1 до +1.
- 2) если $r_B = 0$, то между X и Y отсутствует линейная корреляционная связь, но возможно наличие между ними другого типа связи.
- 3) если $r_B > 0$, то увеличение признака X в среднем приводит к увеличению признака Y . Если $r_B < 0$, то с увеличением X в среднем признак Y уменьшается.
- 4) если $|r_B| = 1$, то между X и Y существует линейная функциональная зависимость, не искажаемая действием случайных факторов.

Для качественной оценки тесноты корреляционной связи между X и Y можно воспользоваться таблицей Чеддока (табл.1).

Таблица 1

Диапазон изменения $ r_B $	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Характер тесноты связи	слабая	умеренная	заметная	высокая	весьма высокая

Пример. Выборочно обследовано 100 снабженческо-сбытовых предприятий некоторого региона по количеству работников X и объемам складской реализации Y (д.е.). Результаты представлены в корреляционной таблице (табл.2).

Таблица 2

X \ Y	5	15	25	35	45	n_y
130	7	1				8
132	2	7	1			10
134	1	5	4	1		11
136		1	15	10	8	34
138			3	12	15	30
140				1	6	7
n_x	10	14	23	24	29	$n=100$

По данным исследования требуется:

- 1) в прямоугольной системе координат построить эмпирические ломаные регрессии Y на X и X на Y, сделать предположение о виде корреляционной связи;
- 2) оценить тесноту линейной корреляционной связи;
- 3) составить линейные уравнения регрессии Y на X и X на Y, построить их графики в одной системе координат;
- 4) используя полученные уравнения регрессии, оценить ожидаемое среднее значение признака Y при $x=40$ чел. Дать экономическую интерпретацию полученных результатов.

Решение.

1. Для построения эмпирических ломаных регрессии вычислим условные средние $\overline{Y_x}$ и $\overline{X_y}$. Вычисляем $\overline{Y_x}$. Так как при $x=5$ признак Y имеет распределение:

Y	130	132	134
n_i	7	2	1

то условное среднее $\overline{Y_{x=5}} = \frac{130 \cdot 7 + 132 \cdot 2 + 134 \cdot 1}{7 + 2 + 1} = 130,8.$

При $x=15$ признак Y имеет распределение:

Y	130	132	134	136
n_i	1	7	5	1

тогда $\overline{Y_{x=15}} = \frac{130 \cdot 1 + 132 \cdot 7 + 134 \cdot 5 + 136 \cdot 1}{14} = 132,86.$

Аналогично вычисляются все $\overline{Y_x}$ и $\overline{X_y}$. Получим таблицы, выражающие корреляционную зависимость Y от X , (табл.3) и X от Y (табл.4).

Таблица 3

X	5	15	25	35	45
$\overline{Y_x}$	130,8	132,86	135,74	137,08	137,86

Таблица 4

Y	130	132	134	136	138	140
$\overline{X_y}$	6,25	14	19,54	32,35	39	43,57

В прямоугольной системе координат построим точки $A_i(x_i, \overline{Y_{x_i}})$, соединим их отрезками прямых, получим эмпирическую линию регрессии Y на X . Аналогично строятся точки $B_i(\overline{X_{y_i}}, y_i)$ и эмпирическая линия регрессии X на Y (см. рис. 1).

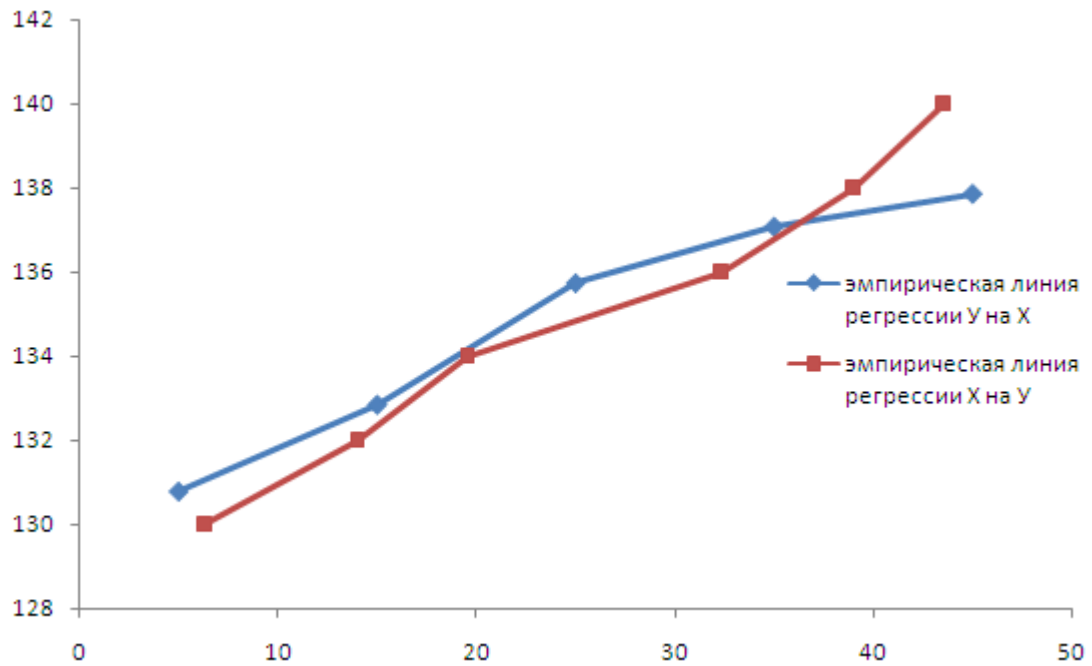


Рис.3. Эмпирические ломаные регрессии

Построенные эмпирические ломаные регрессии Y на X и X на Y свидетельствуют о том, что между количеством работающих (X) и объемом складских реализаций (Y) существует линейная зависимость. Из графика видно, что с увеличением X , \bar{Y}_x также увеличивается, поэтому можно выдвинуть гипотезу о прямой линейной корреляционной зависимости между количеством работающих и объемом складских реализаций.

2. Оценим тесноту связи. Вычислим выборочный коэффициент корреляции.

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \quad \bar{x} = \frac{\sum x_i n_x}{n}, \quad \bar{y} = \frac{\sum y_j n_y}{n},$$

$$\overline{x^2} = \frac{\sum x_i^2 n_x}{n}, \quad \overline{y^2} = \frac{\sum y_j^2 n_y}{n}, \quad \overline{xy} = \frac{\sum x_i y_j n_{ij}}{n};$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2}, \quad \sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2};$$

$$\bar{x} = \frac{5 \cdot 10 + 14 \cdot 14 + 25 \cdot 23 + 35 \cdot 24 + 45 \cdot 29}{100} = 29,8;$$

$$\bar{y} = \frac{130 \cdot 8 + 132 \cdot 10 + 134 \cdot 11 + 136 \cdot 34 + 138 \cdot 30 + 140 \cdot 7}{100} = 135,78;$$

$$\overline{x^2} = \frac{5^2 \cdot 10 + 14^2 \cdot 14 + 25^2 \cdot 23 + 35^2 \cdot 24 + 45^2 \cdot 29}{100} = 1059;$$

$$\overline{y^2} = \frac{130^2 \cdot 8 + 132^2 \cdot 10 + 134^2 \cdot 11 + 136^2 \cdot 34 + 138^2 \cdot 30 + 140^2 \cdot 7}{100} = 18443,4$$

$$\begin{aligned} \overline{xy} = \frac{1}{100} & (130 \cdot 5 \cdot 7 + 130 \cdot 15 \cdot 1 + 132 \cdot 52 + 132 \cdot 15 \cdot 7 + 132 \cdot 25 \cdot 1 + 134 \cdot 5 \cdot 1 + \\ & + 134 \cdot 15 \cdot 5 + 134 \cdot 25 \cdot 4 + 134 \cdot 35 \cdot 1 + 136 \cdot 15 \cdot 1 + 136 \cdot 25 \cdot 15 + 136 \cdot 35 \cdot 10 + \\ & + 136 \cdot 45 \cdot 8 + 138 \cdot 25 \cdot 3 + 138 \cdot 35 \cdot 12 + 138 \cdot 45 \cdot 15 + 140 \cdot 35 \cdot 1 + \\ & + 140 \cdot 45 \cdot 6) = 4075,55 ; \end{aligned}$$

$$\sigma_x = \sqrt{1059 - (29,8)^2} = 13,08; \quad \sigma_y = \sqrt{18443,4 - (135,78)^2} = 2,68;$$

$$r_B = \frac{4075,55 - 29,8 \cdot 135,78}{13,08 \cdot 2,68} = 0,84.$$

Это значение r_B говорит о том, что линейная связь между количеством работников и объемом складских реализаций высокая. Этот вывод подтверждает первоначальное предположение, сделанное исходя из графика.

3. Запишем уравнения регрессии:

$$\overline{y_x} - \bar{y} = r_B \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad \overline{x_y} - \bar{x} = r_B \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y}).$$

Подставляя в эти уравнения найденные величины, получаем искомые уравнения регрессии:

1) уравнение регрессии Y на X: $\overline{y_x} - 135,78 = 0,84 \cdot \frac{2,68}{13,08} (x - 29,8)$, или

$$\overline{y_x} = 0,17x + 130,71;$$

2) уравнение регрессии X на Y:

$$\overline{x_y} - 29,8 = 0,84 \cdot \frac{13,08}{2,68} (y - 135,78), \text{ или } \overline{x_y} = 4,1y - 526,9.$$

Построим графики найденных уравнений регрессии.

Зададим координаты двух точек, удовлетворяющих уравнению

$$\bar{y}_x = 0,17x + 130,71.$$

Пусть $x = 10$, тогда $\bar{y}_x = 132,41$.

$A_1(10; 132,41)$,

Если $x = 40$, тогда $\bar{y}_x = 137,51$.

$A_2(40; 137,51)$

Аналогично находим точки, удовлетворяющие уравнению

$$\bar{x}_y = 4,1y - 526,9:$$

$B_1(10,2; 131)$, $B_2(43; 139)$

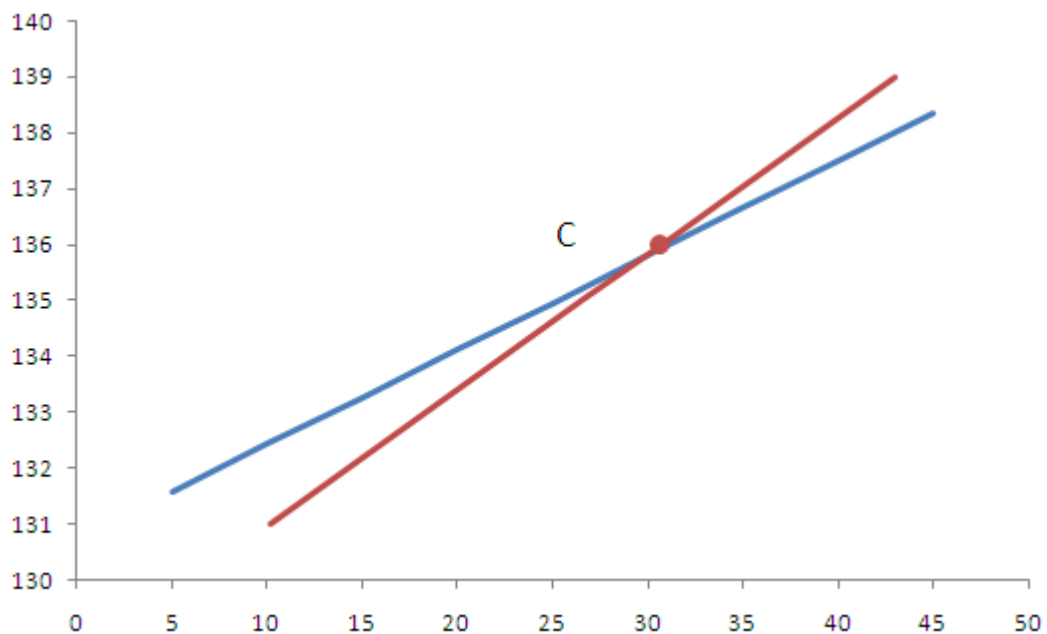


Рис. 4. Графики найденных уравнений регрессий

Контроль: точка пересечения прямых линий регрессии имеет координаты $(\bar{x}; \bar{y})$. В нашем примере: $C(29,8; 135,78)$.

4. Найдем среднее значение \bar{Y} при $x=40$ чел., используя уравнение регрессии \bar{Y} на X . Подставим в это уравнение $x=40$, получим

$$\bar{y}_x = 0,17 \cdot 40 + 130,71 = 137,51.$$

Ожидаемое среднее значение объема складских реализаций при заданном количестве работников ($x=40$) составляет 137,51 д.е.

Замечание 1. Если в корреляционной таблице даны интервальные распределения, то за значения вариант надо брать середины частичных интервалов.

Замечание 2. Если данные наблюдений над признаками X и Y заданы в виде корреляционной таблицы с равноотстоящими вариантами, то целесообразно перейти к условным вариантам:

$$U_i = \frac{x_i - C_1}{h_1}, \quad V_j = \frac{y_j - C_2}{h_2},$$

где h_1 - шаг, т.е. разность между двумя соседними вариантами x_i ;

C_1 - «ложный нуль» вариант x_i (в качестве «ложного нуля» удобно принять варианту, которая расположена примерно в середине ряда);

h_2 - шаг вариант Y;

C_2 - «ложный нуль» вариант Y.

В этом случае выборочный коэффициент корреляции

$$r_{\sigma} = \frac{\overline{U \times V} - \overline{U} \times \overline{V}}{\sigma_u * \sigma_v}, \quad \text{где} \quad \overline{U} = \frac{\sum U_i n_x}{n}, \quad \overline{V} = \frac{\sum V_j n_y}{n},$$

$$\sigma_u = \sqrt{\overline{U^2} - (\overline{U})^2}, \quad \sigma_v = \sqrt{\overline{V^2} - (\overline{V})^2}.$$

Зная эти величины, определим

$$\bar{x} = \overline{U}h_1 + C_1, \quad \bar{y} = \overline{V}h_2 + C_2, \quad \sigma_x = \sigma_u h_1, \quad \sigma_y = \sigma_v h_2.$$

Так в данном примере $C_1=25$, $h_1=10$, $C_2=136$, $h_2=2$;

$$U_i = \frac{x_i - 25}{10}, \quad V_j = \frac{y_j - 136}{2}.$$

U \ V	-2	-1	0	1	2	n _y
-3	7	1				8
-2	2	7	1			10
-1	1	5	4	1		11
0		1	15	10	8	34
1			3	12	15	30
2				1	6	7
n _x	10	14	23	24	29	n=100

$$\bar{U} = \frac{-2 \cdot 10 - 1 \cdot 14 + 0 \cdot 23 + 1 \cdot 24 + 2 \cdot 29}{100} = 0,48$$

$$\bar{V} = \frac{-3 \cdot 8 - 2 \cdot 10 - 1 \cdot 11 + 0 \cdot 34 + 1 \cdot 30 + 2 \cdot 7}{100} = -0,11;$$

$$\overline{U^2} = \frac{4 \cdot 10 + 1 \cdot 14 + 0 \cdot 23 + 1 \cdot 24 + 4 \cdot 29}{100} = 1,94;$$

$$\overline{V^2} = \frac{9 \cdot 8 + 4 \cdot 10 + 1 \cdot 11 + 0 \cdot 34 + 1 \cdot 30 + 4 \cdot 7}{100} = 1,81;$$

$$\begin{aligned} \overline{U \cdot V} &= \frac{1}{100} ((-3) \cdot (-2) \cdot 7 + (-3) \cdot (-1) \cdot 1 + (-2) \cdot (-2) \cdot 2 + (-2) \cdot (-1) \cdot 7 + \\ &+ (-1) \cdot (-2) \cdot 1 + (-1) \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 12 + 1 \cdot 2 \cdot 15 + 2 \cdot 1 \cdot 1 + 2 \cdot 2 \cdot 6) = 1,4 ; \end{aligned}$$

$$\sigma_u = \sqrt{1,94 - 0,2304} = 1,308; \quad \sigma_v = \sqrt{1,81 - 0,012} = 1,34;$$

$$r_{\theta} = \frac{1,4 - 0,48 \cdot (-0,11)}{1,308 \cdot 1,34} = 0,84;$$

$$\bar{x} = \bar{U}h_1 + C_1 = 0,48 \cdot 10 + 25 = 29,8;$$

$$\bar{y} = \bar{V}h_2 + C_2 = -0,11 \cdot 2 + 136 = 135,78;$$

$$\sigma_x = \sigma_u h_1 = 1,308 \cdot 10 = 13,08; \quad \sigma_y = \sigma_v h_2 = 1,34 \cdot 2 = 2,68;$$

$$\bar{y}_x = 0,17x + 130,71; \quad \bar{x}_y = 4,1y - 526,9.$$

Тема 4. Статистическая проверка гипотез. Критерий согласия Пирсона

В экономических исследованиях часто возникает необходимость знать закон распределения генеральной совокупности. С этой целью производят наблюдения и получают опытное (или эмпирическое) распределение случайной величины в виде вариационного ряда. Поставленная задача сводится к оценке закона распределения признака в генеральной совокупности на основе выборочных данных.

Для точной формулировки проблемы дадим основные определения.

Определение 1. Распределение признака в выборке называется эмпирическим распределением.

Определение 2. Распределение признака в генеральной совокупности называется теоретическим распределением.

Определение 3. Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Определение 4. Нулевой (основной) называют выдвинутую гипотезу H_0 .

Определение 5. Конкурирующей (альтернативной) называют гипотезу H_1 , которая противоречит основной.

В результате проверки гипотезы могут быть допущены ошибки двух родов.

Определение 6. Ошибка 1 рода состоит в том, что будет отвергнута правильная нулевая гипотеза. Вероятность ошибки 1 рода называется уровнем значимости и обозначается α .

Определение 7. Ошибка 2 рода состоит в том, что будет принята неправильная нулевая гипотеза. Вероятность ошибки 2 рода обозначается β .

Определение 8. Критерием согласия называется критерий проверки гипотезы о предполагаемом законе неизвестного распределения. Это численная мера расхождения между эмпирическим и теоретическим распределением.

Основная задача. Дано эмпирическое распределение (выборка). Сделать предположение (выдвинуть гипотезу) о виде теоретического распределения и проверить выдвинутую гипотезу на заданном уровне значимости α .

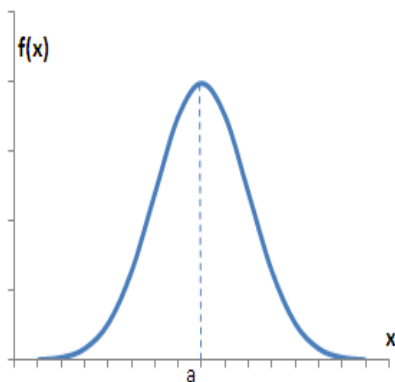
Решение основной задачи состоит из двух частей:

I. Выдвижение гипотезы.

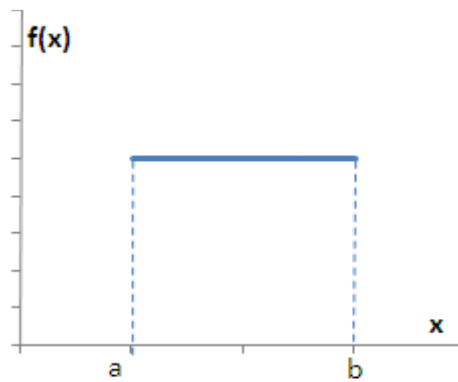
II. Проверка гипотезы на заданном уровне значимости. Рассмотрим подробно эти части.

I. **Выбор гипотезы** о виде теоретического распределения удобно делать с помощью полигонов или гистограмм частот. Сравнивают эмпирический полигон (или гистограмму) с известными законами распределения и выбирают наиболее подходящий.

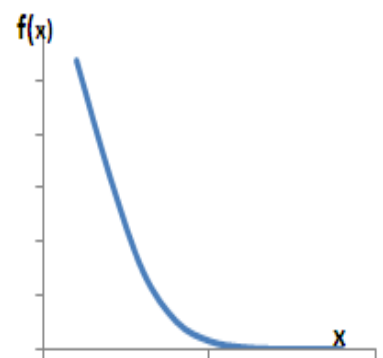
Приведем графики важнейших законов распределения:



Нормальное
распределение $N(a, \sigma)$



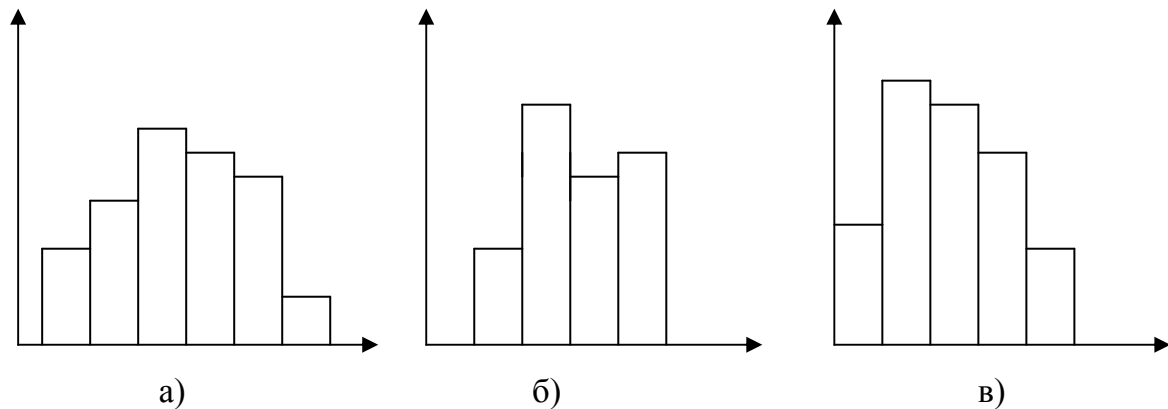
Равномерное
распределение $[a, b]$



Распределение
Пуассона

$$P_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

Пример эмпирических законов распределения



В случае (а) выдвигается гипотеза о нормальном распределении, в случае (б) — гипотеза о равномерном распределении, в случае (в) гипотеза о Пуассоновском распределении.

Основанием для выдвижения гипотезы о теоретическом распределении могут быть теоретические предпосылки о характере изменения признака. Например, выполнение условий теоремы Ляпунова позволяет сделать гипотезу о нормальном распределении. Равенство средней и дисперсии наводит на гипотезу о распределении Пуассона.

На практике чаще всего приходится встречаться с нормальным распределением, поэтому в наших задачах требуется проверить только гипотезу о нормальном распределении.

II. Проверка гипотезы о теоретическом распределении отвечает на вопрос: можно ли считать расхождение между предполагаемыми теоретическим и эмпирическим распределениями случайным, несущественным, объясняемым случайностью попадания в выборку тех или иных объектов, или же это расхождение говорит о существенном расхождении между распределениями. Для проверки существуют различные методы (критерии согласия) – χ^2 (хи-квадрат), Колмогорова, Романовского и др. В наших задачах рассматривается метод χ^2 .

Алгоритм метода

Эмпирическое распределение задано в виде последовательности интервалов одинаковой длины и соответствующих им частот.

1. Находим \bar{x}_g , σ_g . В качестве вариант принимают среднее арифметическое концов интервала.

2. Переходим к случайной величине Z , $Z = \frac{X - \bar{x}_g}{\sigma_g}$. Вычисляем концы интервалов $Z_i = \frac{x_i - \bar{x}_g}{\sigma_g}$, $Z_{i+1} = \frac{x_{i+1} - \bar{x}_g}{\sigma_g}$, причем наименьшее значение Z полагают равным $-\infty$, а наибольшее $+\infty$.

3. Вычисляем теоретические частоты.

$n' = n \cdot P_i$, где n – объем выборки, $P_i = \Phi(Z_{i+1}) - \Phi(Z_i)$, $\Phi(Z)$ – интегральная функция Лапласа.

4. Сравниваем эмпирические и теоретические частоты. Для этого:

а) находим наблюдаемое значение критерия Пирсона

$$\chi_{набл}^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i}$$

б) по таблице критических точек распределения χ^2 , по заданному уровню значимости α и числу степеней свободы $R=S-3$ (S – число интервалов в выборке) находим критическую точку $\chi_{кр}^2(\alpha; R)$.

Если $\chi_{табл}^2 < \chi_{кр}^2(\alpha; R)$, нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности. Если $\chi_{табл}^2 > \chi_{кр}^2(\alpha; R)$, гипотезу отвергают.

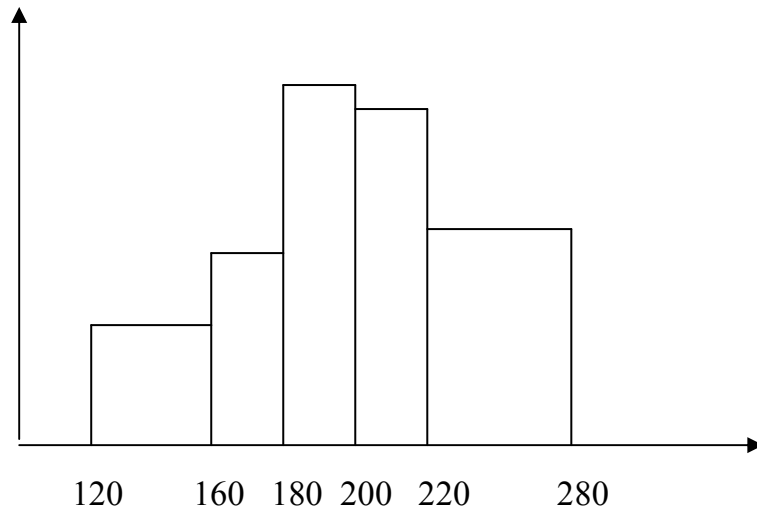
Пример. Результаты исследования спроса на товар представлены в таблице:

120-160	160-180	180-200	200-220	220-280
5	10	14	12	9

Выдвинуть гипотезу о виде распределения и проверить ее на уровне значимости $\alpha=0,01$.

I. Выдвижение гипотезы.

Для указания вида эмпирического распределения построим гистограмму



По виду гистограммы можно сделать предположение о нормальном распределении генеральной совокупности.

II. Проверим выдвинутую гипотезу о нормальном распределении, используя критерий согласия Пирсона.

1. Вычисляем \bar{x}_g, σ_g . В качестве вариант возьмем среднее арифметическое концов интервалов:

$$\bar{x}_g = \frac{140 \cdot 5 + 170 \cdot 10 + 190 \cdot 14 + 210 \cdot 12 + 250 \cdot 9}{50} = 196,6;$$

$$\sigma_g = 32,1.$$

2. Найдем интервалы $(Z_i; Z_{i+1})$: $Z_i = \frac{x_i - 196,6}{32,1}$; $Z_{i+1} = \frac{x_{i+1} - 196,6}{32,1}$.

Левый конец первого интервала примем равным $-\infty$, а правый конец последнего интервала $+\infty$. Результаты представлены в табл. 5

3. Найдем теоретические вероятности P_i и теоретические частоты n'_i .

$P_i = \Phi(Z_{i+1}) - \Phi(Z_i)$, где $\Phi(Z)$ – интегральная функция Лапласа.

$n'_i = n \cdot P_i = 50 \cdot P_i$ (см.табл.5).

Таблица 1

i	Граница интервалов				$\Phi(Z_i)$	$\Phi(Z_{i+1})$	$P_i = \Phi(Z_{i+1}) - \Phi(Z_i)$	$n'_i = 50 \cdot P_i$
	x_i	x_{i+1}	Z_i	Z_{i+1}				
1	120	160	$-\infty$	-1,14	-0,5	-0,3729	0,1271	6,36
2	160	180	-1,14	-0,52	-0,3729	-0,1985	0,1744	8,72
3	180	200	-0,52	0,11	-0,1985	0,0438	0,2423	12,12
4	200	220	0,11	0,73	0,0438	0,2673	0,2235	11,18
5	220	280	0,73	$+\infty$	0,2673	0,5	0,2327	11,64

4. Сравним эмпирические и теоретические частоты. Для этого:

а) вычислим наблюдаемое значение критерия Пирсона.

$$\chi^2_{набл} = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i}. \text{ Вычисления представлены в таблице 2.}$$

Таблица 2

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$(n_i - n'_i)^2/n'_i$
1	5	6,36	-1,36	1,8496	0,291
2	10	8,72	1,28	1,6384	0,188
3	114	12,12	1,88	3,5344	0,292
4	12	11,18	0,82	0,6724	0,060
5	9	11,64	-2,64	6,9696	0,599
Σ	50	50			$\chi^2_{набл} = 1,43$

б) найдем число степеней свободы $R = S - 3 = 5 - 3 = 2$.

По таблице критических точек распределения χ^2 , по уровню значимости $\alpha = 0,01$ и числу степеней свободы $R = 2$ находим критическую

$$\text{точку } \chi^2_{кр}(\alpha; R): \quad \chi^2_{кр}(0,01; 2) = 9,2$$

$$\text{Сравниваем } \chi^2_{набл} \text{ с } \chi^2_{кр}(\alpha; R). \quad \chi^2_{набл} = 1,43 < \chi^2_{кр} = 9,2.$$

Следовательно, нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности. Вывод: расхождение между эмпирическими и теоретическими частотами незначимо (случайно).

Замечание. Интервалы, содержащие малочисленные эмпирические частоты ($n_i < 5$), следует объединить, а частоты этих интервалов сложить. Если производилось объединение интервалов, то при определении числа степеней свободы по формуле $R = S - 3$ следует в качестве S принять число оставшихся после объединения интервалов.

Контрольные задания

Правила выполнения и оформления контрольной работы

Вариант для контрольной работы студент выбирает в соответствии с двумя последними цифрами номера зачетной книжки по следующему правилу: вторая цифра номера варианта должна совпадать с последней цифрой шифра. Далее, если предпоследняя цифра шифра четная, то первая цифра номера варианта должна быть равна 0 или 2; если же предпоследняя цифра нечетная, то первая цифра номера варианта должна быть 1. Например, при шифре 955027 студент решает 07 вариант; при шифре 953054 – вариант 14 и т.д.

При выполнении контрольной работы надо строго придерживаться указанных ниже правил. Работы, выполненные без соблюдения этих правил, не засчитываются и возвращаются студенту для переработки.

1. Контрольные работы выполнять в тетради пастой или чернилами любого цвета, кроме красного, оставляя поля для замечаний преподавателя.
2. На обложке тетради должны быть ясно написаны фамилия студента, его инициалы, название дисциплины и номер контрольной работы.
3. В работу должны быть включены все задачи, указанные в задании, строго по своему варианту. Контрольные работы, содержащие не все задания, а также содержащие задачи не своего варианта, не засчитываются.
4. Решение задач надо располагать в порядке, указанном в заданиях, сохраняя номера задач.
5. Перед решением каждой задачи надо выписать полностью ее условие. Если несколько задач имеют общую формулировку, следует, переписывая условие задачи, заменить общие данные конкретными из соответствующего номера.

6. В конце работы следует указать литературу, которую изучал студент, выполняя данную работу.

7. Зачтенные работы обязательно предъявляются на зачете и экзамене.

8. Перед сдачей зачета и экзамена студент обязан защитить контрольную работу.

Номера вариантов	Номера задач для контрольного задания		
01	1	21	41
02	2	22	42
03	3	23	43
04	4	24	44
05	5	25	45
06	6	26	46
07	7	27	47
08	8	28	48
09	9	29	49
10	10	30	50
11	11	31	51
12	12	32	52
13	13	33	53
14	14	34	54
15	15	35	55
16	16	36	56
17	17	37	57
18	18	38	58
19	19	39	59
20	20	40	60

Задачи для контрольных заданий

В задачах 1–20 выборочные совокупности заданы из соответствующих генеральных совокупностей. Требуется: 1) по несгруппированным данным найти выборочную среднюю; 2) найти доверительный интервал для оценки неизвестного математического ожидания признака X генеральной совокупности (генеральной средней), если признак X распределен по нормальному закону; известно γ – надежность и σ – среднее квадратическое отклонение; 3) составить интервальное распределение выборки с шагом h , взяв за начало первого интервала x_0 ; 4) построить гистограмму частот; 5) дать экономическую интерпретацию полученных результатов.

1. Получены результаты о фондовооруженности в 25 производственных объединениях (тыс. руб.):

16,8 17,2 17,6 17,6 17,9 18,0 18,2 18,4 18,6 18,9 18,9
19,0 19,1 19,2 19,2 19,3 19,7 19,9 20,0 20,0 20,2 20,3
20,4 20,8 21,5

$\gamma=0,95$; $\sigma=1$; $h=1$; $x_0=16,5$.

2. Для определения удельного веса активных элементов основных производственных фондов было выборочно обследовано 25 производственных объединений и получены следующие результаты:

22,3 23,7 24,5 25,9 26,1 26,6 27,3 27,9 28,2 28,5 28,8
29,1 29,2 29,9 30,5 30,7 31,4 32,2 32,3 33,5 34,2 34,4
34,9 35,7 38,9

$\gamma=0,95$; $\sigma=4$; $h=5$; $x_0=20$.

3. Произведено выборочное обследование 25 магазинов по величине товарооборота. Получены следующие результаты (в тыс. руб.):

42,5 60,0 63,5 70,5 82,0 83,5 92,0 95,5 100,0 101,0 105,0

108,5 110,0 115,5 120,0 120,5 122,0 130,0 138,5 140,0

142,0 150,5 160,0 162,1 180,5

$\gamma=0,96$; $\sigma=31$; $h=20$; $x_0=42,5$.

4. При изучении уровня инфляции за некоторый период времени было обследовано 25 стран, получены следующие результаты:

0,35 0,41 0,53 0,59 0,64 0,68 0,71 0,73 0,77 0,78 0,82

0,83 0,85 0,86 0,88 0,89 0,92 0,93 0,97 1,01 1,07 1,08 1,14 1,25 1,28

$\gamma=0,98$; $\sigma=0,22$; $h=0,2$; $x_0=0,3$.

5. Для определения себестоимости продукции было произведено выборочное обследование 25 предприятий пищевой промышленности и получены следующие результаты (руб.)

15,0 16,4 17,8 18,0 18,4 19,2 19,8 20,2 20,6 20,6

20,6 21,3 21,4 21,7 22,0 22,2 22,3 22,7 23,0 24,2

24,2 25,1 25,3 26,0 26,5 27,1.

$\gamma=0,95$; $\sigma=2,8$; $h=2,5$; $x_0=15$.

6. Проведено выборочное обследование 25 частных фирм по количеству занятых в них служащих, получены следующие результаты (чел.):

266 278 315 336 347 354 368 369 391 408

411 416 427 437 444 448 457 462 481 483

495 512 518 536 576

$\gamma=0,96$; $\sigma=65$; $h=50$; $x_0=250$.

7. Проведена случайная выборка личных заемных счетов в банке, получены следующие результаты (тыс. руб.):

1850 2200 2400 2450 2500 2550 2800 2900 2950 3100 3150 3200 3200
3300 3350 3400 3450 3550 3550 3600 3800 3900 4100 4300 4550

$\gamma=0,96$; $\sigma=690$; $h=500$; $x_0=1550$.

8. Выборочно исследовано 25 предприятий для определения объема выпущенной продукции в месяц на одного рабочего, получены следующие результаты:

773 792 815 827 843 854 861 869 877 886 889 892 885
901 903 905 911 918 919 923 929 937 941 955 981

$\gamma=0,92$; $\sigma=50$; $h=40$; $x_0=760$.

9. В Сбербанке проведено выборочное обследование 25 вкладов, которое дало следующие результаты (тыс.руб.):

75 210 350 350 400 520 540 560 590 680 700 700 720
750 780 790 810 850 875 890 1000 1000 1100 1200 1250

$\gamma=0,95$; $\sigma=280$; $h=250$; $x_0=50$.

10. При исследовании объема продаж некоторого товара в супермаркете за 25 дней были получены следующие результаты (шт.):

69 76 77 79 83 86 87 88 89 89 90 91 91 92 93 93
94 94 96 96 99 101 103 107 108

$\gamma=0,98$; $\sigma=9,5$; $h=10$; $x_0=65$.

11. Получены следующие данные о государственных закупках товаров и услуг (усл.д.ед.):

331 346 362 385 404 411 419 429 435 437 441 445 458
468 469 477 481 491 507 518 536 542 543 544 544

$\gamma=0,95$; $\sigma=55$; $h=50$; $x_0=325$.

12. Проведено выборочное обследование 16 компаний по индексу акций нефтяных предприятий и получены следующие результаты:

630 645 652 656 678 687 693 694 697 704

715 716 729 732 745 772

$\gamma=0,95$; $\sigma=55$; $h=40$; $x_0=600$.

13. Проведено выборочное обследование объема промышленного производства за 16 месяцев и получены следующие результаты:

750 950 1000 1050 1050 1150 1150 1150 1200 1200 1250

1250 1350 1400 1400 1550

$\gamma=0,98$; $\sigma=200$; $h=200$; $x_0=700$.

14. Проведено выборочное обследование 25 предприятий, состоящих на самостоятельном балансе, по объему валовой продукции и получены следующие результаты:

627 645 651 664 666 675 679 684 687 693 694 699 702

708 709 711 715 716 725 728 737 744 751 768 781

$\gamma=0,92$; $\sigma=60$; $h=40$; $x_0=608$.

15. При изучении производительности труда (тыс.руб.) на одного рабочего было обследовано 25 однотипных предприятий и получены следующие результаты:

2,5 3,0 3,6 3,8 4,0 4,1 4,2 4,2 4,4 4,6 4,7 4,85 5,2

5,25 5,3 5,4 5,4 5,45 5,6 5,8 5,8 5,85 6,0 6,5 7,0

$\gamma=0,98$; $\sigma=1$; $h=1$; $x_0=25$.

16. Получены результаты выборочного обследования по выполнению плана выработки на одного рабочего (в %):

90,0 96,0 980 980 98,5 99,0 101,5 102 102 102,5 103
103 103,5 104 104 104 104,5 105,5 106 108 108,2 108,7
109 112 113,5

$\gamma=0,98$; $\sigma=4,7$; $h=5$; $x_0=90$.

17. Для определения себестоимости строительно-монтажных работ было произведено выборочное обследование 25 строительно-монтажных управлений и получены следующие результаты (тыс.руб.):

1250 1450 1550 1700 1760 1820 1880 1960 2100 2175
2190 2200 2220 2275 2280 2310 2400 2550 2580 2600
2670 2800 2950 3000 3075

$\gamma=0,94$; $\sigma=446$; $h=400$; $x_0=1100$.

18. Было проведено обследование 25 частных фирм по вкладу в национальный доход. Получены следующие результаты:

159 1625 164 164,5 165,5 166 168,5 169 169 170,5 171
171 171 173 174,5 174,5 176 176,5 178 179 182 183,5
184 185 188

$\gamma=0,95$; $\sigma=7$; $h=5$; $x_0=155$.

19. В сборочном цехе завода было произведено выборочное обследование заработной платы рабочих и получены следующие результаты (в тыс.руб.):

136 155 160 169 175 175 180 188 189 192 195 200 202
205 205 205 208 212 215 220 225 234 242 245 260

$\gamma=0,95$; $\sigma=31$; $h=20$; $x_0=130$.

20. Получены выборочные данные об индексе потребительских цен за 25 лет:

31 33,5 34,5 35 36,5 37 37 38,5 38,5 39 39,5 40 40

40,5 40,5 41 41,5 42 43 43 44 45 46,5 48 49

$\gamma=0,9$; $\sigma=7,5$; $h=4$; $x_0=30$.

В задачах 21-40 по корреляционной таблице требуется:

1) в прямоугольной системе координат построить эмпирические ломаные регрессии Y на X и X на Y , сделать предположение о виде корреляционной связи; 2) оценить тесноту линейной корреляционной связи; 3) составить линейные уравнения регрессии Y на X и X на Y , построить их графики в одной системе координат; 4) используя полученное уравнение регрессии, оценить ожидаемое среднее значение признака Y при $x=x_0$. Дать экономическую интерпретацию полученных результатов.

21. В таблице дано распределение 50 производственных объединений по выработке на одного работника X тыс.руб. и по фондоотдаче Y руб.:

$Y \backslash X$	8	13	18	23	28	n_y
1,25				2	6	8
1,5			4	7	4	15
1,75	1	1	7	5		14
2	2	4	1			7
2,25	3	3				6
n_x	6	8	12	14	10	$n=100$

$x=15$.

22. В таблице дано распределение 100 предприятий по еженедельным издержкам X и по отработанному времени в отделах Y :

Y \ X	30	40	50	60	70	n_y
12	8	8			4	20
18	7	16	7			30
24		15	10	1		26
30		4	9	5	3	21
36				2	1	3
n_x	15	43	26	8	8	$n=100$

$x=57$.

23. В таблице дано распределение 100 торговых предприятий по затратам X тыс.руб. и по ежемесячным объемам продаж Y:

Y \ X	1,0-3,5	1,5-2,0	2,0-2,5	2,5-3,0	3,0-3,5	n_y
100-150	4	.				4
150-200	12	4	2			18
200-250	2	9	10	4		25
250-300		9	18	9	3	39
300-350				11	3	14
n_x	18	22	30	24	6	$n=100$

$x=44$

24. В таблице дано распределение 200 коммерческих предприятий по цене товара X д.ед. и по количеству проданного товара Y тыс.шт.:

Y \ X	0,4-0,8	0,8-1,2	1,2-1,6	1,6-2,0	2,0-2,4	2,4-2,8	n_y
7,25-9,25	14	22					36
9,25-11,25		10	38	6			54
11,25-13,25			30	30	4		64
13,25-15,25			10	12	8		30
15,25-17,25				2	8	6	16
n_x	14	32	78	50	20	6	$n=200$

$x=1,3$.

25. В таблице дано распределение 100 производственных объединений по фондовооруженности основных промышленных фондов на одного работника X тыс. руб. и по выработке на одного работника Y тыс. руб.

Y \ X	4-8	8-12	12-16	16-20	20-24	n_y
5-15	1	2				3
15-25	3	6	1	1		11
25-35		7	9			16
35-45		1	16	8		25
45-55			21	4	5	30
55-65			10	3	2	15
n_x	4	16	57	16	7	$n=100$

$x=22$.

26. В таблице дано распределение 100 однотипных предприятий по основным фондам X млн. руб. и себестоимости единицы продукции Y руб.

Y \ X	20	30	40	50	60	n_y
1	8	2				10
3	12	20	8			40
5			10	1		11
7			9	6	2	17
9			10	4	8	22
n_x	20	22	37	11	10	$n=100$

$x=35$.

27. В таблице дано распределение 100 предприятий производительности труда X и по объемам производства Y:

Y \ X	10	15	20	25	30	35	40	45	n_y
15	2	4							6
20	1	6	5	8		3			23

25		3	13	4	2	1			23
30			4	11	5	7			27
35				2	1	4	3	1	11
40				1	2	5	1	1	10
n_x	3	13	22	26	10	20	4	2	$n=100$

$x=21$.

28. В таблице дано распределение 100 заводов по объему основных производственных фондов X млн.руб. и по суточной выработке продукции Y:

X \ Y	40-50	50-60	60-70	70-80	80-90	90-100	n_y
0-6					4	6	10
6-12			6	6	8		20
12-18	1	2	14	3			20
18-24	6	18	2				26
24-30	4	10	2				16
30-36	6	2					8
n_x	17	32	24	9	12	6	$n=100$

$x=42$

29. В таблице дано распределение 100 предприятий по величине основных фондов X млн. руб. и себестоимости продукции Y млн. руб.:

X \ Y	98-100	100-102	102-104	104-106	106-108	108-110	n_y
15,5-16,5	2	3	1				6
16,5-17,5	3	6	4	1			14
17,5-18,5		4	13	14	10		41
18,5-19,5			5	10	8	6	29
19,5-20,5				2	5	3	10
n_x	5	13	23	27	23	9	$n=100$

$x=103$.

30. В таблице дано распределение 100 заводов по энерговооруженности X и по стоимости продукции Y:

Y \ X	30	40	50	60	70	80	n_y
30	3	6	12	7	2		30
36		2	8	10	2	1	23
42			1	4	16	6	27
48				2	3	5	10
54					4	6	10
n_x	3	8	21	23	27	18	$n=100$

$x=77$.

31. В таблице дано распределение 55 компаний по возрасту X и заработной плате Y усл. ден. ед.

Y \ X	25-35	35-45	45-55	55-65	65-75	n_y
50-80	5	4				9
80-110		12	8	1		21
110-140			5	5		10
140-170			4	7		11
170-200				2	1	3
200-230					1	1
n_x	5	16	17	15	2	$n=55$

$x=28$.

32. В таблице дано распределение 100 предприятий по производительности труда одного рабочего X (в руб.) и по валовой продукции Y тыс.руб.

Y \ X	80	90	100	110	120	n_y
100	2	3	5			10

110	2	6	20	7		35
120	1	3	10	9	5	28
130	1	2	5	4	7	19
140			2	3	3	8
n_x	6	14	42	23	15	$n=100$

$x=95$.

33. В таблице дано распределение 200 заводов по основным фондам X млн. руб. и по готовой продукции Y млн. руб.:

$Y \backslash X$	20	30	40	50	60	70	80	90	100	n_y
12	4									4
18	6	10	2							18
24		8	13	1	1					23
30		4	7	9	3	4	2			29
36		1	2	3	12	4	8			30
42				1	3	18	24	1		47
48							7	12	3	22
54								9	18	27
n_x	10	23	24	14	19	26	41	22	21	$n=200$

$x=98$.

34. В таблице дано распределение 80 рабочих по объемам выпускаемой продукции в месяц на одного рабочего X и по среднемесячной зарплате Y руб.:

$Y \backslash X$	325-375	375-425	425-475	475-525	525-575	n_y
1250-1750	3					3
1750-2250	2	8	2			12
2250-2750		7	5	13		25
2750-3250		1	10	10	7	28
3250-3750				7	5	12
n_x	5	16	17	30	12	$n=80$

$x=463$.

35. В таблице дано распределение 60 семей по доходу семьи X д.е. и уровню потребления Y:

Y \ X	15-30	30-45	45-60	60-75	75-90	n_y
16-24			1	4	1	6
24-32			7	7	2	16
32-40		4	12	2		18
40-48		8	6			14
48-56	2	4				6
n_x	2	16	26	13	3	$n=60$

$x=80$

36. В таблице дано распределение 100 предприятий по производственным средствам X млн. руб. и суточной выработки Y т.

Y \ X	20	30	40	50	60	n_y
10	8	7	2			17
20	2	16	8	6	2	34
30		9	12	12	4	37
40			2	4	5	11
50					1	1
n_x	10	32	24	22	12	$n=100$

$x=45$.

37. В таблице дано распределение 80 снабженческо-сбытовых организаций по складским площадям X тыс. m^2 и по объему складских реализаций Y млн. руб.:

Y \ X	8-16	16-24	24-32	32-40	40-48	n_y
30-70	2	3				5
70-110	3	4	8	1		16
110-150	1	5	16	8	1	31
150-190			12	3	2	17
190-230			1	4	6	11
n_x	6	12	37	16	9	$n=80$

$x=44$.

38. В таблице дано распределение 50 заводов по объему валовой продукции X млн. руб. и себестоимости Y:

Y \ X	1500	2500	3500	4500	5500	n_y
2				1	6	7
2,5			4	6	3	13
3		3	6	4		13
3,5	2	6	3	1		12
4	3	2				5
n_x	5	11	13	12	9	$n=50$

$x=3783$.

39. В таблице дано распределение 50 малых предприятий по выпуску продукции X тыс.ед. в день и по издержкам Y тыс.руб. за день.

Y \ X	4-6	6-8	8-10	10-32	12-14	n_y
0,5-2,0			2	3	1	6
2,0-3,5		4	5	1		10
3,5-5,0		8	5	5		18
5,0-6,5	3	8	2			13
6,5-8,0	2	1				3
n_x	5	21	14	9	1	$n=50$

$x=13$.

40. В таблице дано распределение 200 предприятий по основным фондам X млн. руб. и по готовой продукции Y млн. руб.:

Y \ X	40	50	60	70	80	n_y
15	5					5
20	7	4	8			19
25		16	20	11		47
30		23	32	29	9	93
35			27	2	7	36
n_x	12	43	87	42	16	$n=200$

$x=63$.

В задачах 41-60 даны эмпирические значения случайной величины. Требуется: 1) выдвинуть гипотезу о виде распределения; 2) проверить гипотезу с помощью критерия Пирсона при заданном уровне значимости α . За значения параметров a и σ принять среднюю выборочную и выборочное среднее квадратичное отклонение, вычисленные по эмпирическим данным.

41-50. $\alpha=0,01$. Сумма банковских вкладов имеет следующее распределение:

41.

2-6	6-10	10-14	14-18	18-22
7	15	29	18	11

42.

8-10	10-12	12-14	14-16	16-18
9	17	33	14	7

43.

0,2-0,4	0,4-0,6	0,6-0,8	0,8-1,0	1,0-1,2	1,2-1,4
5	17	23	16	7	2

44.

0,4-1	1-1,6	1,6-2,2	2,2-2,8	2,8-3,4
6	12	21	14	7

45.

0,3-0,7	0,7-1,1	1,1-1,5	1,5-1,9	1,9-2,3
10	22	42	18	8

46.

1,7-2,1	2,1-2,5	2,5-2,9	2,9-3,3	3,3-3,7
12	16	21	15	6

47.

3-4	4-5	5-6	6-7	7-8
6	8	18	11	7

48.

1-5	5-9	9-13	13-17	17-21	21-25
6	10	17	12	4	1

49.

3-5	5-7	7-9	9-11	11-13
7	10	18	10	5

50.

12-22	22-32	32-42	42-52	52-62
14	20	25	13	8

51-60. $\alpha=0,05$. В таблицах дано распределение дохода от реализации некоторого товара:

51.

7-9	9-11	11-13	13-15	15-17
5	23	41	20	11

52.

7-17	17-27	27-37	37-47	47-57
5	11	13	12	9

53.

8-12	12-16	16-20	20-24	24-28	28-32
6	11	25	13	4	1

54.

4-8	8-12	12-16	16-20	20-24
7	25	38	21	9

55.

20-24	24-28	28-32	32-36	36-40
10	21	30	17	12

56.

3-4	4-5	5-6	6-7	7-8
5	11	13	12	9

57.

0-2	2-4	4-6	6-8	8-10
6	11	16	12	5

58.

5-11	11-17	17-23	23-29	29-35
7	12	18	15	8

59.

0-6	6-12	12-18	18-24	24-30
5	11	23	13	8

60.

2-14	14-26	26-38-	38-50	50-62
6	13	19	15	7

Список литературы по дисциплине «Статистика»

1. Гмурман В.Е. Теория вероятностей и математическая статистика.-М.: Высшая школа, 2003.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике.-М.: Высшая школа, 2003.
3. Елисеева И.И. Общая теория статистики.-М.: Финансы и статистика, 2005.
4. Ефимова М.Р. Общая теория статистики.-М.: Инфра-М, 2004.
5. Ефимова М.Р., Ганченко О.И., Петрова Е.В. Практикум по общей теории статистики.-М.: Финансы и статистика, 2008.
6. Шмойлова Р.А. Теория статистики.-М.: Финансы и статистика, 2006.
7. Шмойлова.Р.А. Практикум по теории статистики.-М.: Финансы и статистика, 2006.

Методические указания для выполнения самостоятельной практической работы по дисциплине «Эконометрика»

Цель работы.

Цель СПР – проанализировать стоимость определенной категории квартир на вторичном рынке жилья, построив и проанализировав модель множественной регрессии. Предлагается выбрать один из восьми районов г.Нижнего Новгорода и категорию квартир – однокомнатные, двухкомнатные, трех- и более комнатные, то есть каждый студент проводит уникальное исследование.

1. Сбор данных

Предлагается осуществить поиск и сбор данных о ценах квартир на вторичном рынке жилья с сайтов риэлтерских компаний, газеты «Из рук в руки», иных Интернет-ресурсов. Объем выборки должен быть достаточно большим ($n > 7m$, где n – объем выборки, m – число факторов в модели. Например, если число факторов, от которых зависит цена квартиры и значения которых удалось найти, равно 7, то необходимо иметь информацию не менее чем о 50 квартирах). Данные соответствуют актуальным на момент исследования значениям характеристик продаваемых квартир.

2. Устранение эксцессов

Выявление и удаление данных, резко отличающихся и вносящих возмущения в выборку. Предлагается вычислить стоимость одного квадратного метра общей площади квартиры, упорядочить (отсортировать в порядке возрастания) данные и проанализировать их в области минимальных и максимальных значений. Рекомендуется построить график зависимости цены одного квадратного метра от номера квартиры и

удалить из дальнего анализа «хвосты», резко отличающиеся от остальной части выборки.

3. Построение модели парной линейной регрессии

Предлагается выбрать один, наиболее значимый фактор и построить регрессионную модель $\hat{y} = a + bx$, оценить ее качество (в целом и отдельных параметров). Необходимо проанализировать значение индекса детерминации R^2 , значения F-критерия Фишера и t-критерия Стьюдента, проверить условия применимости МНК.

Для определения параметров регрессионной зависимости и указанных выше характеристик предлагается использовать статистическую функцию ЛИНЕЙН электронных таблиц MS Excel или функцию Регрессия надстройки Пакет анализа. Напомним, что в результате применения функции ЛИНЕЙН получим в качестве результата массив значений, содержащих следующую информацию

b	a
m_b	m_a
R^2	m_y
F	df_2
$S_{регрессион}^2$	$S_{остаточн}^2$

Для проверки качества построенной регрессионной модели анализируем полученное значение R^2 (чем ближе значение R^2 к единице, тем модель лучше соответствует реальным данным).

Далее анализируется значение полученного F-критерия Фишера, в основе которого лежит оценка сумм квадратов регрессионной модели на одну степень свободы. Вычисляются средние квадраты отклонений или дисперсию на одну степень свободы.

$$D_{\text{общая}} = \frac{\sum_{i=1}^n (y - \bar{y})^2}{n-1}; \quad D_{\text{факторная}} = \frac{\sum_{i=1}^n (\hat{y} - y)^2}{1}; \quad D_{\text{остаточная}} = \frac{\sum_{i=1}^n (y - \hat{y})^2}{n-2}$$

Сопоставляя факторную и остаточную дисперсию на одну единицу степени свободы получим величину F-критерия для проверки нулевой гипотезы $H_0: D_{\text{факторная}} = D_{\text{остаточная}}$

$$F_{\text{критерий}} = \frac{D_{\text{факторная}}}{D_{\text{остаточная}}}$$

Для отклонения H_0 необходимо, чтобы F-фактическое было больше F-табличное

$$F_{\text{фактическое}} > F_{\text{табличное}}(\alpha, df_1, df_2),$$

где α – уровень значимости (выбирается значение 1% или 5%).

Для парной регрессии $df_1=1$, а $df_2 = n - 2$.

Если F-фактическое больше F-табличное, то можно сделать вывод о значимости уравнения регрессии с вероятностью $(1 - \alpha)$.

Величина F-критерия связана с индексом детерминации R^2 для парной регрессии зависимостью

$$F = \left(\frac{R^2}{1 - R^2} \right) (n - 2)$$

Величина стандартной ошибки совместно с t-распределением Стьюдента при $(n - 2)$ степенях свободы применяется для проверки существенности коэффициентов регрессии.

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его элементов. С этой целью по каждому из параметров определяется его стандартная ошибка m_a и m_b .

$$m_b = \sqrt{\frac{\sum (y - \hat{y})^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2}{\sum (x - \bar{x})^2}}$$

S^2 – остаточная дисперсия на одну единицу свободы.

Выдвигается гипотеза H_0 о случайной природе показателя b , то есть о незначительности его отличия от 0. Для оценки существенности коэффициентов регрессии, его величина сравнивается с его стандартной ошибкой, то есть определенное значение t-критерия Стьюдента

$$t_b = \frac{b}{m_b},$$

которая затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Аналогично оценивается статистическая значимость свободного члена уравнения регрессии a .

Замечание:

Отметим, что в случае парного регрессионного анализа t-критерий и F-критерий эквивалентны друг другу.

$$t_b = \sqrt{F_{\text{факт}}}$$

Доверительный интервал для коэффициента регрессии b определяется, как $b \pm t * m_b$, где t – табличное значение критерия Стьюдента.

Замечание:

Поскольку b имеет экономическую интерпретацию, то доверительные границы интервала для b не должны содержать противоречивых результатов.

Стандартная ошибка параметра a определяется по формуле:

$$m_a = \sqrt{\frac{\sum (y - \hat{y})^2 * \sum x^2}{(n-2)^n * \sum (x - \bar{x})^2}} = \sqrt{S^2 \frac{\sum x^2}{n * \sum (x - \bar{x})^2}},$$

а соответствующее значение t-статистики $t_a = \frac{a}{m_a}$ (t – статистика для параметра a сравнивается с табличными значениями).

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции m_r .

$$m_r = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}$$

Фактическое значение t-критерия Стьюдента для линейного коэффициента корреляции определяется как

$$t_r = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \sqrt{n - 2}$$

Отметим, что для парной линейной регрессии $t_z^2 = t_b^2 = F$.

Для получения табличных значений F-критерия Фишера и t-критерия Стьюдента используются статистические функции ФРАСПОБР и СТЬЮДРАСПОБР.

Рекомендуется построить график регрессионной прямой, сопоставив ее с реальными данными.

Оценить качество построенной модели можно и с помощью средней ошибки аппроксимации, значение которой принято определять в процентах по следующей формуле

$$A = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Если средняя ошибка аппроксимации $A < 7\%$, то модель хорошо соответствует реальным данным.

Необходимо проанализировать остатки на выполнение условий применимости метода наименьших квадратов (МНК), а именно проверить

все пять условий Гаусса-Маркова (обязательно критерии Гольдфельда-Квандта и Дарбина-Уотсона).

Напомним, что оценки, полученные с использованием МНК, должны быть несмещенными, эффективными и состоятельными.

Несмещенность оценки означает, что математическое ожидание остатков равняется нулю.

Оценки считаются эффективными, если они характеризуются наименьшей дисперсией.

Состоятельность оценки характеризует увеличение их точности с увеличением объемов выборки.

Условия применимости МНК связаны с исследованием поведения остатков (ошибок ε).

Пять предпосылок МНК:

1). Случайный характер остатков. Он проверяется путем построения графика зависимости остатков ε от значения результативного признака y . Если значения остатков разбросаны случайно и находятся в некоторой симметричной относительно оси абсцисс полосе, то первая предпосылка МНК выполняется.

2). Нулевая средняя величина остатков, не зависящая от x_i . Это условие автоматически выполняется для линейных моделей и моделей нелинейных относительно включаемых переменных, но линейных по параметрам.

3). Остатки подчиняются нормальному распределению. Существует критерий размахов (*RS – критерий*). Находим значения размаха остатков R и S по формулам:

$$R = \varepsilon_{\max} - \varepsilon_{\min} ,$$

$$S = \sqrt{\frac{\sum \varepsilon_i^2}{n-1}}$$

и вычисляем их соотношение $RS = R/S$ - это значение сравнивается с табличными: верхними и нижними границами (более подробную информацию предлагается найти самостоятельно)

4). Гомоскедастичность остатков – постоянство дисперсии остатков или дисперсия каждого отклонения одинакова для всех x_i . Невыполнение данного условия означает наличие гетероскедастичности.

Способы выявления гетероскедастичности:

4.1) Графический (построение графика).

4.2) Критерий Гольдфельда-Квандта:

Шаг 1. Упорядочение n наблюдений по мере возрастания значений фактора x .

Шаг 2. Исключение из рассмотрения C центральных наблюдений, при этом $(n - C)/2 > p$, где p - это число оцениваемых параметров.

Шаг 3. Разделение совокупности из $(n - C)$ на две группы (с малыми и большими значениями фактора) и определение по каждой из групп уравнения регрессии.

Шаг 4. Нахождение отношения $R = \frac{S_1^{2ост}}{S_2^{2ост}}$ (большее делим на меньшее).

Найденное значение $F_{табл} = \left(\alpha; \frac{n - C - 2p}{2}; \frac{n - C - 2p}{2} \right)$, если $R > F_{табл}$, то имеет место гетероскедастичность.

5). Отсутствие автокорреляции в остатках. Ищется коэффициент корреляции $r_{\varepsilon_i \varepsilon_j}$ ($j=i+1$), если его значение по модулю значительно (близко к единице), то имеет место автокорреляции остатков. Более точным является критерий для проверки автокорреляции остатков Дарбина-Уотсона.

4. Построение и анализ модели множественной регрессии

- а) Отбор факторов – введение фиктивных переменных и анализ матрицы корреляции между факторами (и показателем). Удаление явно коррелирующих факторов. Проверка на мультиколлинеарность.

Парная регрессия дает хороший результат, если на рассмотрение оказывает один фактор и влияниями других факторов можно пренебречь. Как правило, необходимо учесть влияние двух и более факторов.

Проблемы при построении модели множественной регрессии:

- 1). Отбор факторов, включаемых в модель.
- 2). Объем выборки должен быть достаточно большим ($n > 7m$, где n – объем выборки, а m – число факторов).
- 3). Выбор спецификации модели, то есть функциональной зависимости (линейные и нелинейные).

Отбор факторов.

Факторы, включаемые в модель должны отвечать следующим условиям:

- Факторы должны быть количественно измеримы, если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность одним из двух способов – разработать шкалу и придать количественное значение фактору или использовать фиктивных переменные (вид несущих стен, наличие – отсутствие балкона и т.д.). Фиктивные переменные принимают значения 0 либо 1.

$$z = \begin{cases} 0 & \text{– отсутствие балкона} \\ 1 & \text{– наличие балкона} \end{cases}. \text{ Фиктивные переменные должны}$$

использоваться в модели исследования цены квартиры на рынке жилья.

- Факторы не должны быть интеркоррелированы и тем более находиться в точной функциональной связи. (Интеркоррелированность – это зависимость факторов, когда при построение регрессии вида $\hat{y} = a + b_1x_1 + b_2x_2$ выполняется неравенство, следовательно, связь между x_1 и y менее сильна, чем связь между x_1 и x_2 , в таком случае включать x_1 в регрессионную модель не нужно.)

Если между факторами существует высокая корреляция, то нельзя выявить их изолированное влияние на исследуемый показатель и параметры уравнения регрессии не возможно верно интерпретировать. Считается, две переменные явно коллинеарные, то есть находятся между собой в линейной зависимости, если $|r_{x_i, x_j}| > 0,7$.

Негативным считается явление мультиколлениарности факторов. Это явление проявляется тогда, когда более чем два фактора связаны между собой зависимостью, близкой к линейной, то есть имеет место совокупное воздействие факторов друг на друга. Чем сильнее мультиколлениарность факторов, тем менее надежна оценка параметров регрессии с помощью метода наименьших квадратов, при этом затрудняется интерпретация параметров множественной регрессии. Данную модель нельзя использовать для прогнозирования, так как она ненадежна.

Для оценки мультиколлениарности факторов будем использовать определитель матрицы парных коэффициентов корреляции между факторами. В MS Excel используем функцию МОПРЕД.

Если между факторами существует полная линейная зависимость, то все линейные коэффициенты межфакторной корреляции будут равны единице, $Det|MR| = 0$, если же факторы вообще не связаны, то $Det|MR| = 1$.

При проверке мультиколлинеарности рекомендуется пользоваться упрощенным критерием и сравнивать детерминант матрицы парной корреляции факторов с числом 0,3. Предлагается найти в литературе или в сети Интернет иные критерии проверки мультиколлинеарности. Осуществить процесс определения и устранения факторов, ответственных за мультиколлинеарность.

Самый простой способ устранения мультиколлинеарности – это устранение факторов, ответственных за мультиколлинеарность. Для поиска переменной, ответственной за мультиколлинеарность факторов, строят регрессии, где в качестве зависимой переменной рассматривается каждый из факторов и устраняют тот фактор, связь которого с другими наиболее сильная.

- б) Построение модели множественной линейной регрессии, вычисление средней ошибки аппроксимации. Проверка качества модели по критерию Фишера, частному критерию Фишера и критерию Стьюдента. Множественный индекс детерминации и коэффициент корреляции.

Оценка параметров уравнения множественной регрессии производится с помощью метода наименьших квадратов. При этом оцениваются параметры линейного уравнения; если используется нелинейная модель, то производим сначала ее линеаризацию. При

использовании функции ЛИНЕЙН выделяем $(m + 1)$ столбцов и 5 строчек и получаем искомые значения

b_m	b_{m-1}	...	b_1	b_0
m_{b_m}	$m_{b_{m-1}}$...	m_{b_1}	m_{b_0}
R^2	m_y			
F	df_2			
$S^2_{регрессии}$	$S^2_{остаточная}$			

Не зависимо от форм связи показатель множественной корреляции может быть найден, как индекс множественной корреляции:

$$R_{yx_1x_2\dots x_n} = \sqrt{1 - \frac{\sigma^2_{остаточная}}{\sigma_y^2}}$$

Отметим, что $R^2_{yx_1x_2\dots x_m} \geq R^2_{yx_i}(\max)$, то есть индекс детерминации больше или равен максимальному индексу парной корреляции $|r_{yx_1x_2\dots x_m}| \geq |r_{yx_{i\max}}|$.

Статистическая значимость уравнения множественной регрессии в целом оценивается с помощью F-критерия Фишера, значение которого может быть найдено по формуле

$$F = \frac{D_{факт}}{D_{\varepsilon}} = \frac{R^2}{1 - R^2} \times \frac{n - m - 1}{m}$$

где m - число факторов, включенных в модель.

Наряду с этим критерием используется частный F-критерий Фишера для обоснования включения фактора в модель.

$$F_{x_i} = \frac{R^2_{yx_1x_2x_i\dots x_m} - R^2_{yx_1\dots x_{i-1}x_{i+1}\dots x_m}}{1 - R^2_{yx_1\dots x_i\dots x_m}} \times \frac{n - m - 1}{1}$$

$$y_{x_1} | x_2, x_3, \dots, x_m = a + b_1 x_1 + b_2 \bar{x}_2 + \dots + b_m \bar{x}_m$$

$$y_{x_2} | x_1, x_3, \dots, x_m = a + b_1 \bar{x}_1 + b_2 x_2 + \dots + b_m \bar{x}_m$$

...

или в виде

$$y_{x_1} | x_2, \dots, x_m = A_1 + b_1 x_1$$

$$y_{x_2} | x_1, \dots, x_m = A_2 + b_2 x_2$$

...

где $A_1 = \sum a + b_2 \bar{x}_2 + \dots + b_m \bar{x}_m$, $A_2 = \sum a + b_1 \bar{x}_1 + \dots + b_m \bar{x}_m$ и т.д.

В отличие от парной регрессии, частные уравнения регрессии характеризуют изолированное влияние фактора на результат, при этом иные факторы закреплены на неизменном уровне.

Это позволяет на основе частных уравнений регрессии определять частные коэффициенты эластичности.

$$\mathcal{E}_{y_{x_i}} = b_i \frac{x_i}{\hat{y}_{x_i} | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}$$

Средние показатели коэффициента эластичности:

$$\mathcal{E}_{y_i} = b_i \frac{\bar{x}_i}{\hat{y} | x_1, x_2, \dots, x_m}$$

d) Построение уравнения регрессии в стандартизированном масштабе. Анализ коэффициентов модели.

Уравнение регрессии в стандартизированном масштабе имеет вид:

$$\hat{t}_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m},$$

где $t_y = \frac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \bar{x}}{\sigma_{x_i}}$ - это стандартизированные переменные. Они

обладают следующими свойствами - $\bar{t}_y = \bar{t}_{x_i} = 0$, $\sigma_{t_y}^2 = \sigma_{t_{x_i}}^2 = 1$.

β - это стандартные коэффициенты регрессии, их находят при помощи МНК. В силу того, что все переменные t заданы, как центрированные, стандартные коэффициенты β сравнимы между собой.

По значениям стандартизированных коэффициентов регрессии можно ранжировать факторы по силе их воздействия на результат.

е) Расчет коэффициентов частной корреляции, сопоставление с коэффициентами модели в стандартизированном масштабе.

Частные коэффициенты (индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

С помощью частных коэффициентов корреляции можно проранжировать факторы по их степени влияния, а так же они могут быть использованы при решении проблемы отбора факторов.

Показатели частной корреляции представляют собой отношения сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевший место до введения его в модель.

Сопоставление коэффициентов частной корреляции разного порядка по мере увеличения числа включаемых факторов показывает процесс «очищения» зависимости результативного признака с исследуемым фактором. При наличии m факторов коэффициент частной корреляции, измеряющий влияние на Y фактора x_i при неизменном уровне других факторов определяется по формуле:

$$r_{yx_i|x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m} = \sqrt{1 - \frac{1 - R_{yx_1, x_2, \dots, x_i, \dots, x_m}^2}{1 - R_{yx_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}^2}}$$

Ранжирование факторов по их воздействию на результат с помощью частных коэффициентов корреляции должно согласовываться со

значениями параметров уравнения регрессии в стандартизированном масштабе β_i .

- f) Проверка условий применимости МНК (обязательно критерии Гольдфельда-Квандта и Дарбина-Уотсона)

После проведенного анализа для построенной модели множественной регрессии предлагается исключить (добавить) один или несколько факторов и осуществить построение новой, «улучшенной» модели. Указанные действия повторять до выбора наилучшей модели по имеющимся данным. Выбор наилучшей модели обосновать.

- g) Прогноз цены для ряда квартир, не включенных в выборку и сопоставление прогнозных и реальных данных.

5. Выводы по работе

Изложить основные результаты проделанного исследования, описать полученные модели, в том числе и с точки зрения их практического применения для прогнозирования цены квартир на вторичном рынке жилья.

Список литературы по дисциплине «Эконометрика»

1. Эконометрика: Учеб. для студентов вузов по спец. 061700 "Статистика" / Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2002. – 343 с.: ил. (и последующие издания)
2. Практикум по эконометрике: Учеб. пособие / Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2003. – 192 с.: ил. (и последующие издания)
3. Магнус, Я. Р. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике: Пер. с англ. / Я. Р. Магнус, Х. Нейдеккер. - Перераб. изд. – М.: Физматлит, 2002. – 496 с.
4. Доугерти, К. Введение в эконометрику: Учеб. для студентов экон. спец. вузов: Пер. с англ. / К. Доугерти. – М.: ИНФРА –М, 1999, 2001. XIV, 402 с.: ил.
5. Магнус, Я. Р. Эконометрика: Нач. курс: Учеб. для студентов вузов по экон. спец. / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий; Акад. нар. хоз-ва при Правительстве РФ. – 3-е изд., перераб. и доп. – М.: Дело, 2000. – 399 с.: ил.

Александр Владимирович Елесин
Наталья Юрьевна Прокопенко

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ВЫПОЛНЕНИЯ САМОСТОЯТЕЛЬНЫХ ПРАКТИЧЕСКИХ
РАБОТ ПО ДИСЦИПЛИНАМ «СТАТИСТИКА» И «ЭКОНОМЕТРИКА»

Подписано в печать _ _ _ _ _ Формат 60x90 1/16 Бумага газетная. Печать трафаретная.
Уч. изд. л. 3,2 Усл. печ. л. 4 Тираж 200 экз. Заказ №
Государственное образовательное учреждение высшего профессионального образования
«Нижегородский государственный архитектурно-строительный университет»
603950, Н.Новгород, Ильинская, 65.
Полиграфцентр ННГАСУ, 603950, Н.Новгород, Ильинская, 65