ВВЕДЕНИЕ

Учебно-методическое пособие составлено в соответствии с требованиями федеральных государственных образовательных стандартов (ФГОС 3++).

Оно содержит обязательный минимум знаний по эконометрическим исследованиям, который должен усвоить будущий специалист.

СОДЕРЖАНИЕ РАЗДЕЛОВ ДИСЦИПЛИНЫ «ЭКОНОМЕТРИКА»

Тема 1. Основные аспекты эконометрического моделирования

Введение в эконометрическое моделирование. Основные этапы и проблемы эконометрического моделирования. Задачи эконометрики в области социально-экономических исследований. Информационные технологии на базе ПЭВМ в эконометрических исследованиях. Классификация переменных в эконометрических моделях.

Тема 2. Элементы теории вероятностей и математической статистики

Случайные величины и их числовые характеристики. Функции распределения случайной величины. Многомерные случайные величины. Закон больших чисел. Точечные и интервальные оценки параметров. Проверка статистических гипотез.

Тема 3. Парный регрессионный анализ. Показатели качества регрессии

Функциональная, статистическая и корреляционная зависимости. Линейная парная регрессия. Коэффициент корреляции. Основные положения регрессионного анализа. Интервальная оценка функции регрессии и ее параметров. Оценка значимости уравнения регрессии.

Тема 4. Линейная модель множественной регрессии

Классическая нормальная линейная модель множественной регрессии. Матричная форма модели множественной регрессии.

Предпосылки для множественного регрессионного анализа. Оценка значимости множественной регрессии.

Тема 5. Метод наименьших квадратов (МНК)

Метод наименьших квадратов. Допущения классической линейной модели регрессии. Теорема Гаусса — Маркова.

Тема 6. Свойства оценок МНК

Свойства оценок: состоятельность, несмещенность и эффективность.

Тема 7. Линейные регрессионные модели с гетероскедастичными и автокорреляционными остатками

Последствия нарушения допущений классической модели линейной регрессии. Гомоскедастичность и гетероскедастичность. Тесты на гетероскедастичность: Голфельда — Квандта, Уайта, Глейзера. Устранение гетероскедастичности.

Автокорреляция регрессионных остатков. Проверка уравнения регрессии на автокорреляцию: тесты Дарбина — Уотсона, Бреуша — Годфри, Льюинга — Бокса. Устранение автокорреляции.

Тема 8. Обобщенный метод наименьших квадратов (ОМНК)

Обобщенная линейная модель множественной регрессии (ОЛММР) и обобщенный метод наименьших квадратов (ОМНК). Теорема Айткена. ОЛММР с гетероскедастичными остатками. Сравнение ОМНК и МНК: оценки в моделях регрессии с гетероскедастичными остатками. ОЛММР с автокоррелированными остатками. Искажения характеристик точности МНК: оценки, обусловленные автокоррелированностью остатков.

Тема 9. Вопросы практического использования регрессионных моделей. Регрессионные модели с переменной структурой. Фиктивные переменные

Мультиколлинеарность: полная и частичная. Методы устранения мультиколлинеарности. Использование регрессионных моделей с переменной структурой. Фиктивные переменные, их влияние на оценку регрессионной модели. Критерий Г. Чоу. Частная корреляция.

Тема 10. Нелинейные модели регрессии и их линеаризация

Некоторые виды нелинейных зависимостей, поддающиеся непосредственной линеаризацией: экспоненциальные, логарифмические, гиперболические, степенные.

Тема 11. Характеристики временных рядов

Временной ряд и этапы его анализа. Составляющие временного ряда: тренд, сезонная, циклическая, случайная компоненты. Аналитическое выравнивание временного ряда. Прогнозирование на основе моделей временных рядов.

Тема12. Модели стационарных и нестационарных временных рядов, их классификация

Модели стационарных временных рядов и их идентификация: модели авторегрессии порядка «p» ($AR(p)$), скользящего среднего порядка «q» ($MA(q)$) и авторегрессионные модели скользящими средними в остатках ($ARMA$ (p, q, k)).

Тема 13. Система линейных одновременных уравнений. Косвенный, двухшаговый и трехшаговый метод наименьших квадратов

Система одновременных уравнений (СОУ). Косвенный метод наименьших квадратов.

Структурная и приведенная формы модели систем одновременных уравнений. Рекурсивные системы одновременных уравнений. Модель спроса — предложения как пример системы одновременных уравнений. Основные структурные характеристики моделей. Условия идентифицируемости уравнений системы. Идентификация рекурсивных систем.

Статистическое оценивание неизвестных значений параметров. Двухшаговый метод наименьших квадратов (2 МНК) оценивания структурных параметров отдельного уравнения системы.

Трехшаговый метод наименьших квадратов (3 МНК) одновременного оценивания всех параметров системы уравнений.

ТЕМА № 1
ОСНОВНЫЕ АСПЕКТЫ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

План

1. Введение в эконометрическое моделирование.
2. Основные этапы и проблемы эконометрического моделирования.

1.1. Введение в эконометрическое моделирование

Последние десятилетия эконометрика как научная дисциплина стремительно развивается. Свидетельством признания эконометрики является присуждение за наиболее выдающиеся разработки в этой области Нобелевских премий по экономике Р. Фришу и Я. Тинбергену, Л. Клейну, Дж. Тобину, Р. Лукасу, Дж. Хекману и Д. Мак-Фаддену, Р. Энглу и К. Грейнджеру.

«Современное экономическое образование, — утверждает директор ЦЭМИ РАН академик В. Л. Макаров, — *держится на трех китах: макроэкономике, микроэкономике и эконометрике».*

Дадим определение эконометрики:

> Эконометрика — это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе экономической теории, экономической статистики, экономических измерений и математико-статистического инструментария придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

Таким образом, эконометрика занимается разработкой и применением статистических методов для измерений взаимосвязей между экономическими переменными.

В любой эконометрической модели зависимая переменная разбивается на две части: объясненную и случайную. Задача эконометрического моделирования может быть сформулирована следующим образом: *на основании экспериментальных данных* *определить объясненную часть и, рассматривая случайную составляющую как случайную величину, получить оценки параметров ее распределения*.

Эконометрическая модель имеет следующий вид:

$$Y = f(X) + e, \tag{1.1}$$

где: Y — наблюдаемое значение зависимой переменной;

$f(X)$ — объясненная часть, зависящая от значений объясняющих переменных;

e — случайная составляющая.

1.2. Основные этапы и проблемы эконометрического моделирования

Выделяют шесть основных этапов эконометрического моделирования:

1. *Постановочный.*

На этом этапе формируется *цель* исследования и *набор* участвующих в модели экономических переменных.

В качестве практических целей эконометрической модели (1.1) можно выделить:

* анализ исследуемого экономического объекта (процесса);
* прогноз экономических показателей, характеризующих состояние и развитие анализируемой системы;
* имитация возможных сценариев социально-экономического развития анализируемой системы, когда статистически выявленные взаимосвязи между ее различными характеристиками используются для прослеживания того, как возможные изменения тех или иных параметров повлияют на значения интересующих нас характеристик;
* выработка управленческих решений.

При работе с экономическими показателями необходимо учитывать то, что многие из них неотрицательны (поэтому их надо описывать неотрицательными случайными величинами), неопределенны (поэтому их надо описывать в терминах теории нечетких множеств, математики и статистики интервальных данных), имеют нечисловую природу (поэтому к ним больше применимы методы статистики объектов нечисловой природы).

При выборе экономических переменных необходимо теоретическое обоснование каждой переменной, они не должны быть связаны функциональной или корреляционной зависимостью.

2. Априорный.

На данном этапе проводится *анализ сущности изучаемого объекта*, формирование и формализация априорной (известной до начала моделирования) информации.

3. Параметризация.

На этом этапе осуществляется *моделирование*, то есть выбор общего вида функции f(X) в эконометрической модели (1.1), выявление входящих в нее связей.

4. Информационный.

На этом этапе осуществляется *сбор* необходимой статистической информации — наблюдаемых значений экономических переменных:

$$\left(x_{i_1}, x_{i_2} \ldots x_{i_p}\ y_{i_1}, y_{i_2}, \ldots . y_{i_q}\right), \text{где } i = 1, 2 \ldots n.$$

Статистическая информация может быть получена в условиях активного или пассивного эксперимента.

5. Идентификация модели.

На этом этапе осуществляется *статистический анализ модели и оценка ее параметров*.

6. Верификация модели.

На данном этапе проводится *проверка адекватности модели*. Выясняется, насколько удачно решены проблемы идентификации модели, какова точность расчетов по данной модели, насколько построенная модель (1.1) соответствует моделируемому реальному экономическому объекту (процессу).

Контрольные вопросы к теме № 1

1. Что такое эконометрика.
2. Каков математический инструментарий эконометрики.
3. Каковы специфические особенности экономических данных.
4. Назовите прикладные цели эконометрического исследования.
5. Какие ученые внесли наибольший вклад в эконометрику.

6. Сформулируйте задачу эконометрического моделирования.

7. Назовите основные этапы эконометрического моделирования и дайте их характеристику.

Задача № 1.1. Торговое предприятие имеет сеть, состоящую из 12 магазинов, информация о деятельности которых представлена в таблице № 1.

Таблица № 1

№ магазина	Годовой товарооборот, млн руб.	Торговая площадь, тыс. $м^2$	Среднее число посетителей в день, тыс. чел.
1	19,76	0,24	8,25
2	38,09	0,31	10,24
3	40,95	0,55	9,31
4	41,08	0,48	11,01
5	56,29	0,78	8,54
6	68,51	0,98	7,51
7	75,01	0,94	12,36
8	89,05	1,21	10,81
9	91,13	1,29	9,89
10	91,26	1,12	13,72
11	99,84	1,29	12,27
12	108,55	1,49	13,92

Нужно:

1. Построить диаграммы рассеяния годового товарооборота (Y) в зависимости от торговой площади (X_1) и среднего числа посетителей в день (X_2).
2. Определить форму связи между результирующим показателем (Y) и каждым из факторов (X_1 и X_2).

Задача № 1.2. На основании информации, приведенной в таблице № 1, построено двухфакторное уравнение годового товарооборота (Y) в зависимости от торговой площади (X_1) и среднего числа посетителей в день (X_2), которое имеет вид:

$$Y = -10,8153 + 61,6583X_1 + 2,2748X_2 .$$

Нужно:

Дать экономическую интерпретацию коэффициентов уравнения.

ТЕМА № 2
ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

План

1. Случайные величины и их числовые характеристики.
2. Функция распределения случайной величины.
3. Многомерные случайные величины.
4. Закон больших чисел.
5. Точечные и интервальные оценки параметров.
6. Проверка статистических гипотез.

2.1. Случайные величины и их числовые характеристики

Приведем основные понятия теории вероятностей и математической статистики, которые используются в эконометрике.

Случайный эксперимент — процесс регистрации наблюдения на единице обследуемой совокупности.

Каждый из возможных исходов случайного эксперимента называется *элементарным событием*, а совокупность таких исходов — *пространством элементарных событий*:

$$\Omega = \left(\omega_1, \omega_2 \ldots \varpi_n\right) \tag{2.1}$$

Случайным событием «A» называют любое подмножество ($\omega_1, \omega_2, \ldots.., \omega_k$) пространства элементарных событий, то есть

$$A = \left(\omega_1, \omega_2, \ldots, \omega_k\right) \tag{2.2}$$

Это означает, что осуществление любого из элементарных событий, входящих в «*A*», влечет за собой осуществление тия «*A*».

Каждому элементу ω_i пространства элементарных событий соответствует некоторая неотрицательная числовая характеристика p_i шансов его появления, называемая вероятностью события ω_i, причем:

$$p_1 + p_2 + ... + p_n = \sum p = 1 \qquad (2.3)$$

Вероятностью $P(A)$ *события «A»* называется численная мера степени объективной возможности появления этого события. Вероятность события A равна:

$$P(A) = m/n, \qquad (2.4)$$

где m — число случаев, благоприятствующих событию «A»;
n — общее число случаев.

При определенных условиях в качестве оценки вероятности события $P(A)$ используется статистическая вероятность $P^*(A)$, то есть частность $W(A)$ появления события «A» в «n» произведенных испытаниях.

Случайной величиной называется переменная, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений. Можно сказать, что случайная величина (X) это функция, определенная на множестве элементарных исходов, ее возможные значения и их общее число определяются структурой соответствующего пространства Ω элементарных событий, то есть

$$X = f(\omega) \qquad (2.5)$$

где ω — элементарное событие, принадлежащее пространству Ω, то есть $\omega \subset \Omega$.

Случайные величины бывают дискретные и непрерывные. Для *дискретной* случайной величины множество возможных значений конечно или счетно, для *непрерывной* — бесконечно и несчетно.

Случайные величины описываются законами распределения. *Законом распределения* случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями. Например,

X_1	X_2	………………	X_i	…………….	X_n
P_1	P_2	………………	P_i	………………	P_n

Такая таблица называется *рядом распределения* дискретной случайной величины.

Закон распределения дискретной случайной величины позволяет вычислить вероятности любых событий, связанных со случайной величиной, однако он не всегда удобен для анализа.

Поэтому для описания случайных величин часто используют их числовые характеристики. *Числовые характеристики* — числа, в сжатой форме выражающие наиболее существенные черты распределения случайной величины.

А) *Числовые характеристики центра группирования значений случайной величины.*

* *Математическое ожидание* (*среднее значение*) дискретной случайной величины (X) представляет собой сумму произведений ее значений (x_i) на соответствующие им вероятности (p_i):

$$M(x) = \sum x_i \times p_i$$

Для непрерывного случая:

$$M(x) = \int xf(x)dx \tag{2.7}$$

Свойства математического ожидания:

1) $M(c) = c$, где c — постоянная величина.

2) $M(cx) = cM(x)$.

3) $M(x_1 + x_2 + ... + x_n) = M(x_1) + M(x_2) + ... + M(x_n)$.

4) $M(xy) = M(x)M(y)$, где $(x; y)$ — независимые случайные величины.

5) $M(x + c) = M(x) + c$.

• *Медиана* ($X_{\text{мед}}$) случайной величины (X). Представляет значение случайной величины, обладающее свойством: вероятность того, что случайная величина окажется больше ($X_{\text{мед}}$) равна вероятности того, что она окажется меньше ($X_{\text{мед}}$):

$$P(X \geq X_{\text{мед}}) = P(X \leq X_{\text{мед}}) = 0,5 \tag{2.8}$$

• *Мода* ($X_{\text{мод}}$) случайной величины (X) — наиболее часто наблюдаемое ее значение.

В) *Числовые характеристики степени рассеяния значений случайной величины.*

Они дают представление о том, как сильно могут отклоняться от своего центра группирования значения случайной величины.

• *Дисперсия* $D(X)$ *случайной величины* (X) есть математическое ожидание квадрата ее отклонения от математического ожидания:

$$D(x) = M\left(X - M(X)\right)^2;$$

$$D(x) = \sum\left(x_i - M(x)\right)^2 \cdot p_i \text{ (для дискретного случая);} \qquad (2.9)$$

$$D(x) = \int\left(x - M(x)\right)^2 f(x)d(x) \text{ (для непрерывного случая)} \qquad (2.10)$$

Свойства дисперсии:

1) $D(c) = 0$, где c — постоянная величина.

2) $D(cx) = c^2 D(x)$.

3) $D(x+y) = D(x) + D(y)$ где $(x; y)$ — независимые случайные величины.

4 $D(x) = M(x^2) - M(x)^2$.

5) $D(c+bx) = b^2 D(x)$.

* *Среднее квадратическое отклонение* δ_x случайной величины (X) представляет собой арифметическое значение корня квадратного из ее дисперсии:

$$\delta_x = \sqrt{D(x)}. \qquad (2.11)$$

* *Коэффициент вариации* V_x:

$$V_x = \sqrt{\frac{D(x)}{M(x)}} \cdot 100\%. \qquad (2.12)$$

С) Показатели связи между двумя переменными.

Наиболее распространенными показателями линейной связи между двумя переменными являются:

* *Ковариация* cov(*x, y*) случайных величин (*x*) и (*y*). Представляет собой математическое ожидание произведения отклонений этих величин от своих математических ожиданий:

$$\operatorname{cov}(x,y)=M\left(\left[x-M(x)\right]\left[y-M(y)\right]\right). \tag{2.13}$$

Отметим основные свойства ковариации двух случайных величин:

1) $\operatorname{cov}(x,y)=0$, если (*x*), (*y*) независимы.

2) $\operatorname{cov}(a+вx,\ c+дy)=вд\operatorname{cov}(x,y)$, где *a*, *в*, *c*, *д* — константы.

• *Коэффициент корреляции* P_{xy} двух случайных величин (*x*) и (*y*). Представляет собой отношение их ковариации к произведению средних квадратических отклонений этих величин:

$$P_{xy}=\frac{\operatorname{cov}(x,y)}{\delta_x\delta_y}. \tag{2.14}$$

Отметим свойства коэффициента корреляции:

1) $-1\le P_{xy}\le 1$.

2) $P_{xy}=0$, если случайные величины (*x*) и (*y*) независимы.

3) Если $P_{xy}=1$, то между случайными величинами (*x*) и (*y*) существует линейная функциональная зависимость.

2.2. Функция распределения случайной величины

Формами закона распределения случайной величины (*x*) являются *функция и плотность распре деления.* Плотность распределения $\varphi(x)$ существует только для непрерывных случайных величин.

Функция распределения случайной величины (*x*) представляет собой функцию F(*x*), выражающую для каждого значения (*x*) вероятность того, что случайная величина (*x*) примет значение, меньше (*x*):

$$F(x)=P(X\le x). \tag{2.15}$$

Назовем основные свойства функции распределения случайной величины:

1) Функция распределения случайной величины есть неубывающая функция на всей числовой оси, то есть при $x_2 \geq x_1$

$$F(x_2) \geq F(x_1).$$

2) Функция распределения случайной величины есть неотрицательная функция, заключенная между нулем и единицей:

$$0 \leq F(x) \leq 1.$$

3) Вероятность попадания случайной величины (X) в интервал (x_1, x_2) равна приращению ее функции распределения на этом интервале, то есть:

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1).$$

Случайная величина (X) называется непрерывной, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме, быть может, отдельных точек.

Плотностью вероятности $\varphi(x)$ непрерывной случайной величины (X) называется производная ее функции распределения:

$$\varphi(x) = F'(x). \tag{2.16}$$

2.3. Многомерные случайные величины

Упорядоченный набор $x = (x_1, x_2, \dots x_n)$ случайных величин называется *многомерной случайной величиной*. Пример, показатель «качество жизни россиян» в определенный период времени может быть охарактеризован многомерной случайной величиной $x = (x_1, x_2, \dots, x_n)$ где x_1 — продолжительность жизни, x_2 — уровень потребления социально значимых благ и услуг, x_3 — уровень развития науки и т. д.

Функцией распределения многомерной случайной величины $(X_1, \dots, X_n)$ называется функция $F(x_1, x_2, \dots x_n)$, выражающая вероятность совместного выполнения (n) неравенств $X_1 \le x_1,\ X_2 \le x_2, \ldots, X_n \le x_n$, то есть

$$F\left(x_1,\, x_2, \ldots x_n\right) = P(X_1 \le x_1,\, X_2 \le x_2, \ldots,\, X_n \le x_n)\,. \tag{2.17}$$

Плотностью распределения непрерывной двумерной случайной величины (X, Y) называется вторая смешанная частная производная ее функции распределения, то есть

$$\varphi\left(x, y\right) = F''_{x,y}\left(x, y\right). \tag{2.18}$$

Условным законом распределения одной из одномерных составляющих двумерной случайной величины (X, Y) называется ее закон распределения, вычисленный при условии, что другая составляющая приняла определенное значение. Условные плотности распределения двумерной случайной величины (X, Y) определяются по формулам:

$$\begin{gathered}\varphi_y\left(x\right) = \frac{\varphi\left(x, y\right)}{\int\limits_{-\infty}^{+\infty} \varphi\left(x, y\right) d\left(x\right)};\\ \varphi_x\left(y\right) = \frac{\varphi\left(x, y\right)}{\int\limits_{-\infty}^{\infty} \varphi\left(x, y\right) d\left(y\right)}.\end{gathered} \tag{2.19}$$

Условное математическое ожидание $M_x(Y)$ случайной величины (Y) при $X = x$ есть функция от (x), называемая *функцией регрессии* (Y) по (X).

2.4. Закон больших чисел

Рассмотрим некоторые результаты теории вероятностей, используемые в эконометрике.

4.1. Закон больших чисел

Под *законом больших чисел* понимается принцип, согласно которому совокупное действие большого числа случайных факторов приводит (при общих условиях) к результату, почти не зависящему от случая.

4.2. Теорема Бернулли

Частость события в (n) повторных независимых испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью (p), при неограниченном увеличении числа (n) сходится по вероятности к вероятности (p) этого события в отдельном испытании. То есть

$$\lim P = \left((\frac{m}{n} - p) \leq \varepsilon \right) = 1\,, \qquad (2.20)$$

где ε — сколь угодно малая положительная величина.

4.3. Теорема Ляпунова

Если независимые случайные величины $x_1, x_2, .., x_n$ имеют конечные математические ожидания и дисперсии, по своему значению ни одна из этих случайных величин резко не выделяется среди остальных, то при $n \to \infty$ закон распределения их суммы $\sum_{i=1}^{n} X_i$ неограниченно приближается к нормальному.

В частности, если $x_1, x_2, .., x_n$ одинаково распределены, то закон распределения их суммы $\sum_{i=1}^{n} x_i$ при $n \to \infty$ неограниченно приближается к нормальному.

2.5. Точечные и интервальные оценки параметров

Теперь рассмотрим точечные и интервальные оценки случайных величин.

Оценкой ($\widetilde{Q}_n$) *параметра* (Q) называют всякую функцию результатов наблюдений над случайной величиной (X), с помощью которой судят о значениях параметра (Q). Оценка ($\widetilde{Q}_n$) есть величина случайная, и она должна обладать наименьшим разбросом относительно оцениваемого параметра (Q).

Оценка ($\widetilde{Q}_n$) параметра (Q) называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, то есть М($\widetilde{Q}_n$) = (Q). В противном случае оценка называется *смещенной.*

Оценка ($\widetilde{Q}_n$) параметра (Q) называется *состоятельной*, если она удовлетворяет закону больших чисел, то есть сходится по вероятности к оцениваемому параметру:

$$\lim P\left(\left|\widetilde{Q}_n - Q\right| \le \varepsilon\right) = 1 . \qquad (2.21)$$

Несмещенная оценка ($\widetilde{Q}_n$) параметра (Q) называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра (Q), вычисленных по выборкам одного и того же объема (n).

Полный набор данных, относящихся к исследуемому явлению, называется *генеральной совокупностью*. В реальности редко возможны ситуации, когда исследователю доступны все данные для исследования. Обычно приходится иметь дело с ограниченным их количеством — *выборкой*. Статистические показатели, полученные на основе выборок, называются *выборочными показателями*. Для решения задач статистического оценивания рассчитывают показатели на основе некоторой совокупности выборок, в результате получается *выборочное распределение выборочного показателя*.

Для нахождения оценок параметров генеральной совокупности по данным выборки используется *метод максимального правдоподобия*. Основу этого метода составляет *функция правдоподобия*, которая выражает плотность вероятности совместного появления результатов выборки $\left(x_1, x_2 \ldots x_n\right)$:

$$L\left(x_1, x_2 \ldots x_n, Q\right) = \prod_{i=1}^{n} \varphi\left(x_i, Q\right) . \qquad (2.22)$$

Согласно этого метода в качестве оценки неизвестного параметра (Q) принимается такое значение ($\widetilde{Q}_n$), которое максимизирует функцию (L).

Достоинством метода максимального правдоподобия является то, что получаемые им оценки состоятельны, эффективны и имеют нормальное распределение. Несмещенной, состоятельной и эффективной оценкой математического ожидания (α) является *выборочная средняя:*

$$\overline{x} = \frac{\sum x_i n_i}{n}, \tag{2.23}$$

где x_i, n_i — частоты значений x_i.

Несмещенной, состоятельной оценкой дисперсии (δ^2) является *выборочная дисперсия*:

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 n_i}{n}. \tag{2.24}$$

Наряду с точечными оценками рассматривают интервальные оценки параметров.

Интервальная оценка параметра (Q) представляет собой числовой интервал (Q_1, Q_2), который с заданной вероятностью (γ) накрывает неизвестное значение параметра (Q). Такой интервал называется *доверительным*, а вероятность (γ) — *надежностью* оценки.

Величина доверительного интервала зависит от объема выборки (n) и от значения надежности.

Доверительный интервал для *генеральной* средней (Xo) (математическое ожидание средней всех выборочных средних) на уровне значимости (α) определяется по формуле:

$$\left(\overline{x} - t_{1-\alpha,\, n-1} \frac{s}{\sqrt{n-1}}, \overline{x} + t_{1-\alpha,\, n-1} \frac{s}{\sqrt{n-1}}\right). \tag{2.25}$$

Доверительный интервал для генеральной дисперсии на уровне значимости (α) определяется по формуле:

$$\left(\frac{ns^2}{\chi_{\alpha/2,\, n-1}^{2}}, \frac{ns^2}{\chi_{1-\alpha/2, n-1}^{2}}\right). \tag{2.26}$$

2.6. Проверка статистических гипотез

Исследователь часто располагает априорными догадками относительно величины параметров генеральной совокупности.

Введем следующие понятия:

• *Статистическая гипотеза* — это предположение о величине параметра распределения генеральной совокупности.

Процесс проверки гипотезы базируется на формировании двух типов гипотез: нулевой и альтернативной.

∗ *Нулевая гипотеза* (H_0) — допущение, которое считается верным до тех пор, пока не будет доказано обратное, исходя из результатов статистической проверки.

∗ *Альтернативная гипотеза* (H_1) — гипотеза, которая принимается, если в результате проверки отвергается нулевая гипотеза.

Статистическая проверка гипотезы состоит в использовании *стандартизированного статистического критерия, вычисляемого по данным выборки.*

В процессе проверки гипотезы существует вероятность того, что нулевая гипотеза (H_0) будет отвергнута, когда в действительности она должна быть принята. Это называется *ошибкой 1-го рода*. *Ошибка 2-го рода* имеет место при принятии нулевой гипотезы (H_0) в то время, когда она должна быть отвергнута.

Следует помнить, что принятие гипотезы (H_0) расценивается не как абсолютно верный содержащийся в ней факт, а лишь как достаточное правдоподобное, не противоречащее опыту утверждение.

Контрольные вопросы к теме № 2

1. Дайте определение случайной величины.
2. Назовите числовые характеристики случайной величины.
3. В чем состоит смысл закон больших чисел и предельных теорем Бернулли и Ляпунова.
4. В чем суть проверки статистической гипотезы.

Задача 2.1. Дан ряд распределения случайной величины (*X*):

X	0	1	2	3
P	0,05	0,30	0,45	0,22

Необходимо: а) найти математическое ожидание M(*X*), дисперсию D(*X*), среднее квадратическое отклонение (δ) случайной величины (*X*).

Задача 2.2. Случайная величина (*X*) сосредоточена на интервале (–1,3), задана функцией распределения $F(x)=\frac{1}{4}x+\frac{1}{4}$. Найти вероятность попадания случайной величины (*X*) в интервал (0, 2).

Задача 2.3. Случайная величина (*X*) задана функцией распределения:

$$F(x)=\begin{array}{rl} 0 & \text{при } x\le 0 \\ x^2 & \text{при } 0\le x\le 1 \\ 1 & \text{при } x\ge 1 \end{array}$$

Найти плотность распределения $\varphi(x)$.

Задача 2.4. Дан ряд распределения случайной величины:

X	1	4	5
P	0,4	0,1	0,5

Найти и изобразить графически функцию ее распределения.

ТЕМА № 3
ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ. ПОКАЗАТЕЛИ КАЧЕСТВА РЕГРЕССИИ

План

1. Функциональная, статистическая и корреляционная зависимости.
2. Линейная парная регрессия.
3. Коэффициент корреляции.
4. Основные положения регрессионного анализа.
5. Интервальная оценка функции регрессии и ее параметров.
6. Оценка значимости уравнения регрессии.

3.1. Функциональная, статистическая и корреляционная зависимости

Регрессионный анализ занимает центральное место в математическом аппарате эконометрики. *Задачами регрессионного анализа* являются установление формы зависимости между переменными, оценка функции регрессии, прогноз значений зависимой переменной. Различают функциональную, статистическую и корреляционную зависимости. При *функциональной* зависимости каждому значению одной переменной соответствует вполне определенное значение другой. При *статистической* зависимости каждому значению одной переменной соответствует определенное распределение другой переменной (например, зависимость производительности труда на заводе от его энерговооруженности). При *корреляционной* зависимости каждому значению одной переменной соответствует определенное математическое ожидание другой.

Регрессионный анализ применим тогда, когда изучаемые зависимости:

1) имеют *стохастическую* (*вероятностную*) природу;

2) выявляются на основании *статистического* наблюдения за анализируемыми событиями.

Корреляционная зависимость представляется в виде

$$M_x(Y) = \varphi(x), \tag{3.1}$$

где (Y) — эндогенная переменная (результирующая, объясняемая):

X — экзогенная переменная (объясняющая, регрессор).

Эндогенная переменная характеризует некий результат функционирования анализируемой экономической системы. В регрессионном анализе эндогенная переменная выступает в роли *функции*, значения которой всегда стохастичны по своей природе. Экзогенная переменная описывает функционирование изучаемой экономической системы и задается как бы «извне». В регрессионном анализе она играет роль *аргумента* той функции, в качестве которой рассматривается эндогенная переменная. По своей природе она может быть как случайной, так и неслучайной.

Уравнение (3.1) называют *уравнением регрессии*, а $\varphi(x)$ — *функцией регрессии*.

С помощью уравнения регрессии (3.1) можно решить следующие практические задачи:

1) установить факт наличия (отсутствия) связи между переменными, проверить статистическую значимость этой связи;

2) осуществить прогноз неизвестных значений эндогенной переменной по заданным значениям экзогенных переменных;

3)выявить причинно-следственные связи между эндогенными и экзогенными переменными;

4) осуществить регулирование значений экзогенных показателей с целью управления эндогенными показателями.

Выбор вида функции регрессии — наиболее важная и наименее теоретически обоснованная часть регрессионного анализа. В практике регрессионного анализа сложились определенные традиции, которых обычно придерживаются при выборе вида искомой функции:

1) прежде всего, необходимо опираться на имеющуюся априорную информацию об изучаемом явлении;

2) исследователь проводит визуальный анализ данных, использует разные приемы математической статистики при обработке полученных данных;

3) главное правило при построении регрессионной модели — это движение от простого к сложному. Поэтому анализ принято начинать с простейшего случая — линейной модели.