КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

ИНСТИТУТ УПРАВЛЕНИЯ, ЭКОНОМИКИ И ФИНАНСОВ

Кафедра экономико-математического моделирования

М.Я.ИБРАГИМОВ, И.И.ИСМАГИЛОВ, Г.И. ЛИСОГОР

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ЭКОНОМИСТОВ

Учебное пособие

Казань – 2015

УДК 330.43
ББК Ув631я73-1

Принято на заседании кафедры
экономико-математического моделирования
Протокол № 2 от 12 октября 2015 года

Рецензент:
кандидат технических наук,
доцент кафедры экономико-математического моделирования,
В.А. Талызин

$\boxed{\text{Ибрагимов М.Я.,}}$ **Исмагилов И.И., Лисогор Г.И.**
Элементы математической статистики для экономистов/ М.Я. Ибрагимов, И.И.Исмагилов, Г.И.Лисогор. – Казань: Казан. ун-т, 2015. – 49 с.

Учебное пособие представляет собой краткий теоретический курс материалов по представленным темам теории вероятности и математической статистики, задачи, упражнения и тесты. Рекомендовано также к использованию в учебном процессе по дисциплинам «Экономико-математические модели» и «Эконометрика».

СОДЕРЖАНИЕ

1. Случайные величины и их числовые характеристики

Основная задача математической статистики – оценивание параметров экономических явлений, проверка статистических гипотез, количественная оценка взаимосвязей экономических величин.

Примеры случайных величин.

1. Совокупный доход страны складывается из доходов отдельных граждан, которые склонны утаивать размер дохода. Поэтому совокупный доход – величина случайная.
2. Вкладывая деньги в покупку акций на финансовом рынке, мы точно не знаем, повысится ли их стоимость в будущем или нет. Поэтому стоимость акций – величина случайная.
3. Страховая компания, продавая вам страховой полис, не знает наверняка, случится ли что-либо с вашим здоровьем или имуществом.
4. Однако можно заметить, что в случае движения молекул, из которых состоят окружающие нас тела (например, воздух), скорость каждой отдельно взятой молекулы совершенно случайная, но тем не менее средняя скорость всех молекул тела имеет определенное значение, определяемое температурой тела.
5. Точно также в повседневной жизни: спрос индивидуума на какой-либо товар может быть совершенно случайным, но совокупный спрос на товар может быть спрогнозирован.

Таким образом, события, случайные на микроуровне (на уровне отдельных субъектов или объектов) могут стать закономерными на макроуровне (на уровне экономики в целом), то есть достаточно большое число случайностей порождает закономерность, называемую статистической закономерностью. Эта статистическая закономерность и есть то, ради чего стоит изучать теорию вероятностей и математическую статистику.

Случайная величина (СВ). Формальное определение: *случайная величина – величина, принимающая под воздействием случайных обстоятельств то или иное значение из заданного множества.*

Виды случайных величин:

- дискретная (принимает конечное или счетное множество значений);
- непрерывная (множество значений – интервал, конечный или бесконечный промежуток).

Распределение СВ. Для описания случайной величины необходимо указать не только множество ее возможных значений, но и охарактеризовать вероятности, с которыми принимаются те или иные значения. Такое полное описание случайной величины называется ее законом распределения.

Пример 1. Пусть дискретная случайная величина X принимает значения $x_1,\ldots,x_n$ с вероятностями $P(X=x_1)=p_1,\ldots,P(X=x_n)=p_n$. Тогда таблица

X	x_1	…	x_n
P	p_1	…	p_n

задает закон распределения. Так как события $X=x_1,\ldots,X=x_n$ несовместны и образуют полную группу событий (то есть группу несовместных событий, сумма которых есть достоверное событие), то $\sum_{k=1}^{n} p_k = 1$.

Пример 2. Пусть X – случайная величина, обозначающая сумму выигрыша на лотерейный билет, если на каждые 100 билетов разыгривается 1 выигрыш в \$1000, 10 по \$100 и 25 по \$5. Тогда закон распределения величины X имеет вид:

X	0	5	100	1000
P	0.64	0.25	0.1	0.01

Закон распределения дискретной случайной величины можно представить графически. Для этого в прямоугольной системе координат строят точки с координатами (x_i,p_i), а затем соединяют их последовательно отрезками прямой. Получается многоугольник распределения дискретной СВ.

Функцией распределения $F(x)$ случайной величины X называется вероятность того, что X примет значение меньше данного числа x, то есть $F(x)=P(X<x)$.

Свойства функции распределения:

1) $P(x_1 \le X < x_2) = F(x_2) - F(x_1)$;

2) $0 \le F(x) \le 1$;

3) $F(x)$ – монотонно неубывающая функция.

Функцией плотности распределения вероятностей случайной величины X называется функция $f(x)$, такая, что $P(x_1 \le X < x_2) = \int_{x_1}^{x_2} f(x)dx$ (вводится для непрерывных случайных величин).

Таким образом, $F(x) = \int_{-\infty}^{x} f(x)dx$, то есть $\int_{-\infty}^{\infty} f(x)dx = 1$ для любой функции плотности распределения вероятностей $f(x)$. Легко видеть, что $f(x)=F'(X)$.

Основные характеристики случайных величин:

Математическое ожидание – мера среднего значения СВ:

$M(X) = \sum_{k=1}^{n} p_k x_k$ - дискретный случай; $M(X) = \int_{-\infty}^{\infty} x f(x)dx$ - непрерывный случай.

Другие меры среднего значения:

Медиана: $M_e(X)=x_e$ такому, что $F(x_e)=0.5$.

Мода: $M_o(X)=x_o$ такому, что $f(x_o)=max\, f(x)$.

Дисперсия – мера разброса СВ вокруг среднего значения:

$$D(X) = M[(X - M(X))^2] = \int_{-\infty}^{\infty} (x - M(X))^2 f(x)dx.$$

Для дискретной СВ: $D(X)=\sum_{k=1}^{n}(x_k-M(X))^2 p_k$. Если вероятности p_k одинаковы, то

$$D(X)=\frac{\sum_{k=1}^{n}(x_k-M(X))^2}{n}.$$

Стандартное (среднеквадратическое) отклонение: $\sigma(X)=\sqrt{D(X)}$.

Коэффициент вариации: $V(X)=\dfrac{\sigma(X)}{|M(X)|}$.

Центральные моменты порядка *k*: $m_k=M[(X-M(X))^k]$.

m_3 – мера асимметрии;

m_4 – мера эксцесса (остроты пика графика функции плотности распределения вероятностей) случайной величины.

Свойства числовых характеристик СВ (*a*,*b* - постоянные величины)**:**

1. $M(aX+b)=aM(X)+b$
2. $D(aX+b)=a^2D(X)$
3. $D(X)=M(X^2)-[M(X)]^2$

Пример 3. Пусть X – случайная величина, равная сумме выплат страхового органа, имеет распределение:

X	0	50	250	300
P	0.9965	0.002	0.001	0.0005

Тогда $M(X)$=50x0.002+250x0.001+300x0.0005=0.5. Таким образом, страховой орган должен получить с клиентов сумму, не меньшую, чем 0.5.

Пример 2. Пусть $f(x)=F'(x)=\begin{cases} 0 & для \quad x<0; \\ 1 & для \quad 0\le x\le 1; \\ 0 & для \quad x>1. \end{cases}$

Тогда $M(X)=\int_{-\infty}^{\infty} f(x)x\,dx=\int_{0}^{1} x\,dx=\left.\frac{x^2}{2}\right|_0^1=\frac{1}{2}$.

$$D(X)=M(X^2)-[M(X)]^2=\int_{-\infty}^{\infty} x^2 f(x)dx-\frac{1}{4}=\left.\frac{x^3}{3}\right|_0^1-\frac{1}{4}=\frac{1}{12}$$

<u>Пример 3.</u> Пусть даны распределения случайных величин *X* и *Y*:

X	-0.1	-0.001	0	0.001	0.1
P	0.1	0.2	0.4	0.2	0.1

Y	-20	-10	0	10	20
P	0.3	0.1	0.2	0.1	0.3

Тогда *M*(*X*)=0, *M*(*Y*)=0, *D*(*X*)=0.00204, *D*(*Y*)=260. Функции распределения:

$$F(x)=\begin{cases} 0, & x<-0.1; \\ 0.1, & -0.1\le x<-0.001; \\ 0.3, & -0.001\le x<0; \\ 0.7, & 0\le x<0.001; \\ 0.9, & 0.001\le x<0.1; \\ 1, & x\ge 0.1 \end{cases};\qquad F(y)=\begin{cases} 0, & x<-20; \\ 0.3, & -20\le x<-10; \\ 0.4, & -10\le x<0; \\ 0.6, & 0\le x<10; \\ 0.7, & 10\le x<20; \\ 1, & x\ge 20 \end{cases}.$$

2. Статистическое исследование взаимосвязей экономических переменных

В приложениях математической статистики важны прежде всего взаимосвязи случайных величин. Для первичного анализа таких взаимосвязей используются показатели ковариации и корреляции.

Коэффициент ковариации случайных величин *X* и *Y* – абсолютная мера взаимосвязи переменных (зависит от масштаба): *cov*(*X*,*Y*)=*M*[(*X*-*M*(*X*))((*Y*-*M*(*Y*)].

Коэффициент корреляции – относительная мера взаимосвязи (безразмерная величина): $\rho(X,Y)=\frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}$.

1) Для любых случайных величин *X* и *Y*: $-1\le\rho(X,Y)\le 1$.

Если $Y=a+bX$ – функциональная линейная зависимость, то $\rho(X,Y)=1$ при $b>0$, $\rho(X,Y)=-1$ при $b<0$. Обратное тоже верно, то есть если $\rho(X,Y)=1$, то $Y=a+bX$ при $b>0$ и если $\rho(X,Y)=-1$, то $Y=a+bX$ при $b<0$.

Таким образом, коэффициент корреляции есть мера линейности взаимосвязи переменных, степень близости к линейной зависимости.

Определение. Случайные величины X и Y называются *независимыми,* если $P(X<x,Y<y)=P(X<x)P(Y<y)$.

2) Если случайные величины X и Y независимы, то $\rho(X,Y)=0$. Обратное в общем случае неверно. Например, пусть X и Y принимают соответствующие значения

X	-2	-1	0	1	2
Y	4	1	0	1	4

то есть $Y=X^2$. Однако $cov(X,Y)=0$, $\rho(X,Y)=0$.

3. Свойства суммы и произведения случайных величин

1) $M(X\pm Y)=M(X)\pm M(Y)$
2) $M(XY)=M(X)M(Y)+cov(X,Y)$[1]
3) $D(X\pm Y)=D(X)+D(Y)+2cov(X,Y)$[2]

4. Основные типы распределений случайных величин

4.1. Равномерное распределение (на отрезке $[a,b]$):

$$f(x)=\begin{cases} 0, & x<a; \\ c, & a\le x\le b; \\ 0, & x>b \end{cases} \qquad F(x)=\begin{cases} 0, & x<a; \\ cx-ca, & a\le x\le b; \\ 1, & x>b \end{cases} \qquad c=\frac{1}{b-a},$$

$$M(X)=\frac{a+b}{2}, \qquad D(X)=\frac{(b-a)^2}{12}.$$

[1] $cov(X,Y)=M[(X-M(X))(Y-M(Y))]=M[XY-YM(X)-XM(Y)+M(X)M(Y)]=M(XY)-M(Y)M(X)-M(X)M(Y)+M(M(X)M(Y))=M(XY)-M(X)M(Y)$

[2] $D(X+Y)=M((X+Y)^2)-(M(X+Y))^2=M(X^2)+2M(XY)+M(Y^2)-(M(X)+M(Y))^2=M(X^2)-(M(X))^2+M(Y^2)-(M(Y))^2+2M(XY)-2M(X)M(Y)=D(X)+D(Y)+2cov(X,Y)$.

4.2. Нормальное распределение (N(m,σ^2))

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(y-m)^2}{2\sigma^2}}dy.$$

$M(X)=m$, $D(X)=\sigma^2$, $m_3(X)=0$, $m_4(X)=3\sigma^2$.

Для нормально распределенной случайной величины X:
$P(|X-M(X)|\le\sigma)\approx0.68$; $P(|X-M(X)|\le2\sigma)\approx0.95$; $P(|X-M(X)|\le3\sigma)\approx0.997$.

4.3. Стандартное нормальное распределение ((N(0,1): m=0, σ=1))

Если случайная величина X имеет нормальное распределение с параметрами m и σ, то случайная величина $Z=\frac{X-m}{\sigma}$ имеет стандартное нормальное распределение. При этом

$P(x_1\le X<x_2)=F\left(\frac{x_2-m}{\sigma}\right)-F\left(\frac{x_1-m}{\sigma}\right)$, где $F(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{x^2}{2}}dx$ – функция распределения стандартно нормально распределенной случайной величины Z. Отметим, что $F(z)=0.5+\Phi(z)$, где $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{0}^{z}e^{-\frac{x^2}{2}}dx$ - так называемая функция Лапласа.

Теорема Ляпунова. Распределение суммы n произвольно распределенных и независимых СВ при $n\to\infty$ стремится к нормальному распределению, если вклад отдельных слагаемых в сумму равномерно мал.

Неравенство Чебышева. $P(|X-M(X)|\le a\sigma)\ge1-\frac{1}{a^2}$ для любого a>0 и любой случайной величины X.

Например, при a=2: $P(|X-M(X)|\le2\sigma)\ge3/4$ для любого распределения.

Для нормального распределения: $P(|X-M(X)|\le\sigma)\approx0.68\ge0$; $P(|X-M(X)|\le2\sigma)\approx0.95\ge3/4$; $P(|X-M(X)|\le3\sigma)\approx0.997\ge8/9$.

Значения функции распределения $F(z)$ стандартно нормально распределенной случайной величины Z табулированы. Эти таблицы можно использовать для определения значения функции распределения любой нормально распределенной случайной величины X, имея ввиду, что если случайная величина X имеет нормальное распределение $N(m,\sigma^2)$, то

$$P(x_1 \le X < x_2) = F\left(\frac{x_2 - m}{\sigma}\right) - F\left(\frac{x_1 - m}{\sigma}\right).$$

Другими словами, $P(|X\text{-}m| \le \lambda\sigma)=2F(\lambda)\text{-}1$.

4.4. Логарифмически нормальное распределение

Положительная случайная величина X называется распределенной *логарифмически нормально*, если ее логарифм $Y=lnX$ распределен по нормальному закону.

В основе применения логнормального распределения лежит предельная теорема, согласно которой распределение произведения n независимых положительных случайных величин при условии их равномерной малости в пределе стремится к логнормальному распределению. Практически логнормальное распределение используется для описания совокупного мультипликативного действия многих случайных факторов, когда их влияние на изменение конечного результата примерно пропорционально изменению самих факторов. Свидетельством близости распределения к логнормальному является значительная асимметрия, обусловленная ограничением на случайную величину слева от нуля. После логарифмирования левая часть распределения растягивается, приближаясь к нормальному.

Логнормальное распределение используется, в частности, для описания распределения доходов, банковских вкладов, месячной заработной платы, посевных площадей на разные культуры и т.д.

4.5. Распределение χ^2 (хи-квадрат)

Сумма квадратов n попарно независимых случайных величин $Z_1, \ldots Z_n$, имеющих стандартное нормальное распределение образует случайную величину с χ^2 распределением с n степенями свободы: $\chi^2 = Z_1^2 + \ldots + Z_n^2$. Если переменные $Z_1, \ldots Z_n$ связаны s линейными уравнениями, то их сумма имеет распределение χ^2 с $\nu = n\text{-}s$ степенями свободы.

$$f_{\chi^2}(x, \upsilon) = K e^{-\frac{x}{2}} x^{\frac{\upsilon}{2}-1}, \quad K = \frac{1}{\int_0^\infty e^{-\frac{x}{2}} x^{\frac{\upsilon}{2}-1} dx}, \quad M(\chi^2) = \upsilon, \quad D(\chi^2) = 2\upsilon.$$

4.6. Распределение Стьюдента

Пусть Z – стандартная нормально распределенная случайная величина и U – независимая от нее случайная величина, имеющая χ^2 распределение с ν степенями свободы. Тогда величина $T = \frac{Z\sqrt{\nu}}{U}$ подчиняется t-распределению Стьюдента с ν степенями свободы. Для случайной величины T имеет место:

$$f_t(x, \nu) = B\left(1 + \frac{x^2}{\upsilon}\right)^{-\frac{\nu+1}{2}}, \qquad M(T) = 0, \qquad D(T) = \frac{\nu}{\nu - 2}.$$

Величина B определяется из условия нормируемости, то есть из равенства

$$\int_{-\infty}^{\infty} f_t(x, \nu) dx = B \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{\upsilon}\right)^{-\frac{\nu+1}{2}} dx = 1.$$

С ростом числа степеней свободы ν распределение Стьюдента быстро приближается к стандартному нормальному, поэтому при больших ν (практически начиная с $\nu = 20$) вероятность $P(T<t)$ можно определять с помощью функции Лапласа: $P(T<t) \approx 0.5 + \Phi(t)$.

4.7. *F*-распределение Фишера

Пусть U и V – независимые случайные величины, имеющие χ^2-распределение со степенями свободы ν_1 и ν_2. Тогда величина $F = \dfrac{U/\nu_1}{V/\nu_2}$ имеет *F*-распределение Фишера со степенями свободы ν_1 и ν_2. Функция плотности распределения вероятностей $f_F(x, \upsilon_1, \nu_2) = Cx^{\frac{\upsilon_1}{2}-1}\left(1+\dfrac{\nu_1 x}{\nu_2}\right)^{-\frac{\nu_1+\nu_2}{2}}$, где коэффициент C определяется из условия нормируемости.

При больших значеняих ν_1 и ν_2 распределение Фишера приближается к нормальному.

5. Линейная регрессионная модель. Оценка параметров и качества модели

В регрессионном анализе важную роль играют два понятия:

1. **Генеральная совокупность** – все возможные наблюдения показателя X.
2. **Выборка объема *n*** из генеральной совокупности – случайный вектор $(x_1,\ldots,x_n)$, где случайные величины x_i независимы и одинаково распределены, распределение каждой из которых совпадает с распределением исследуемой случайной величины X. Если все элементы генеральной совокупности имеют равную возможность оказаться в выборке, то выборка называется репрезентативной.

Построение регрессионных моделей обычно проводится с использованием 2-х видов экономических данных:

1. *Временные ряды (Time series)* – данные, характеризующие один и тот же объект, но в различные, следующие, как правило, с одинаковым интервалом, моменты времени (например, объем ВВП страны в 90-е годы).

2. *Перекрестные данные (Cross-section data)* – это данные о значении какого-либо показателя, полученные для разных однотипных объектов (фирм, регионов, стран…). При этом данные относятся к одному моменту времени, либо их временная принадлежность несущественна (например, объем ВВП по странам мира в 2000 г., данные бюджетных исследований).

Задача статистического оценивания состоит в том, чтобы по данным случайной выборки оценить неизвестные значения параметров известного закона распределения генеральной совокупности значений СВ. *Оценкой* числового параметра θ называется функция выборочных значений $\theta^*=\theta^*(x_1,\ldots,x_n)$, которая в определенном статистическом смысле близка к истинному значению этого параметра. Важнейшими статистическими свойствами оценки, определяющими ее близость к истинному значению числовой характеристики рассматриваемой СВ, являются свойства несмещенности, состоятельности и эфективности.

1) Оценка называется *несмещенной*, если ее математическое ожидание как случайной вличины равно истинному значению числовой характеристики: $M(\theta^*)=\theta$.

2) Оценка называется *состоятельной*, если предел оценки по вероятности равен истинному значению числовой характеристики, то есть

lim P($|\theta^*-\theta|<\varepsilon$)=1 для любого $\varepsilon>0$.

3) Оценка называется *эффективной*, если она имеет наименьшую дисперсию среди всех несмещенных оценок.

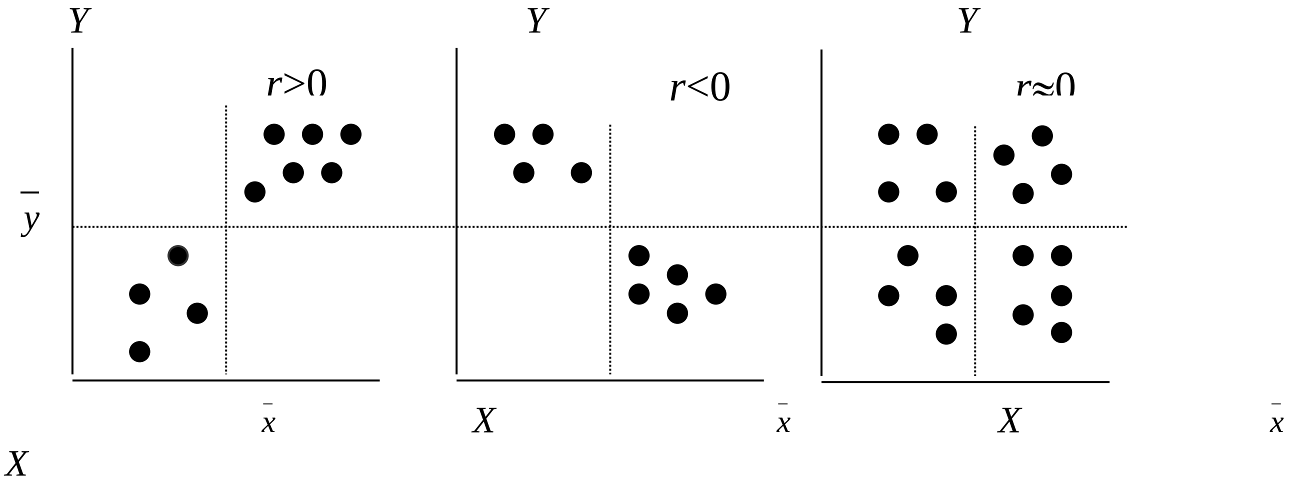
Экономические переменные свзязаны друг с другом множеством взаимосвязей. Рассмотрим экономические переменные Y и X. Пусть заданы выборки их значений объема n:

Y	X
y_1	x_1
y_2	x_2
…	…
y_n	x_n

Пусть ρ - коэффициент корреляции для генеральной совокупности значений X и Y, то есть $\rho(X,Y)=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$. Оценкой для ρ является выборочный коэффициент корреляции $r(X,Y)=\dfrac{\sum\limits_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y})}{\sqrt{\sum\limits_{k=1}^{n}(x_k-\bar{x})^2}\sqrt{\sum\limits_{k=1}^{n}(y_k-\bar{y})^2}}$, где

$\bar{x}=\dfrac{\sum\limits_{k=1}^{n}x_k}{n}$ - оценка математического ожидания $M(X)$ переменной X, $\bar{y}=\dfrac{\sum\limits_{k=1}^{n}y_k}{n}$ - оценка математического ожидания $M(Y)$ переменной Y.

Смысл коэффициента корреляции:



Выборочный коэффициент корреляции двух СВ является случайной величиной. Как статистическая оценка, он отклоняется от истинного значения коэффициента корреляции в генеральной совокупности, но чем больше такое отклонение, тем менее оно вероятно.

Статистическая гипотеза о равенстве нулю коэффициента корреляции («нулевая гипотеза») проверяется следующим образом:

- Предполагается, что коэффициент корреляции ρ в генеральной совокупности равен нулю.
- При ρ=0 выборочный коэффициент корреляции r (оценка для ρ) при данном числе наблюдений имеет определенное распределение.
- Оценки, сильно отличающиеся от нуля, имеют малую вероятность. Для конкретной величины r вычисляется вероятность получить в выборке такую или большую по модулю ее величину.
- Если эта вероятность мала, то есть случилось маловероятное событие, то гипотеза о том, что ρ=0, отвергается. «Критическое», то есть граничное значение вероятности называется *уровнем значимости*[3]. В эконометрических исследованиях уровень значимости α выбирается обычно равным 1% или 5%.

Пусть выдвинуты два предположения:

H_0: ρ=0 (нулевая гипотеза);

H_1: $\rho \neq 0$ (альтернативная гипотеза).

Вероятность отклонить нулевую гипотезу и принять гипотезу H_1, когда в действительности верна H_0, называется *ошибкой 1-го рода*. Вероятность принять гипотезу H_0, когда в действительности верна H_1, называется *ошибкой 2-го рода*.

Если возможными считаются только положительные или только отрицательные значения коэффициента корреляции, то рассматривается односторонняя альтернативная гипотеза H_1: $\rho>0$ или H_1: $\rho<0$.

При проверке нулевой гипотезы рассматривается не сама величина r выборочного коэффициента корреляции, а ее функция, имеющая известное

[3] *Уровень значимости статистического критерия* - вероятность ошибочно отвергнуть основную проверяемую гипотезу, когда она верна. В теории статистической проверки гипотез уровень значимости также называют *вероятностью ошибки первого рода*.

распределение. Такой функцией для выборочного коэффициента корреляции является t-статистика, рассчитываемая по формуле: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. Она имеет распределение Стьюдента с (n-2) степенями свободы.

Пример 4. Коэффициенты корреляции между темпами инфляции и уровнем безработицы в 1980-1990 гг. (n=11) оказались равными: во Франции r=–0.81; в Германии r=–0.73. Зададим уровень значимости α=0.01, то есть 1%. Критическая область состоит из двух равных «хвостов», площадь каждого из которых равна $\alpha/2$=0.005. Из таблиц распределения Стьюдента для числа степеней свободы n-2=9 находим $t_{кр}$=3.25. Это означает, что нулевая гипотеза отвергается при $|t|$>3.25. В этом случае признается наличие линейной зависимости между переменными (принимается альтернативная гипотеза).

Для Франции: $t = \frac{-0.81\sqrt{11-2}}{\sqrt{1-0.81^2}} \approx -4.1$; для Германии: $t = \frac{-0.73\sqrt{11-2}}{\sqrt{1-0.73^2}} \approx -3.20$.

Значит, при данном уровне значимости нулевая гипотеза принимается для Германии, отвергается для Франции. Однако, если бы мы использовали одностороннюю альтернативную гипотезу, или же задали бы уровень значимости больше, чем 0.01, то и для Германии нулевая гипотеза была бы отвергнута, что вполне согласуется с выводом кривой Филлипса о наличии отрицательной связи между уровнями инфляции и безработицы.

Итак, корреляционный[4] анализ позволяет проверить гипотезу о наличии линейной связи между переменными. Однако этого недостаточно для экономического анализа, поскольку возникают и другие проблемы:

- Какие переменные определяют поведение других величин, то есть могут быть использованы в качестве объясняющих переменных?
- Какова формула зависимости между переменными и каков экономический смысл ее коэффициентов?

Оценивание параметров (коэффициентов) уравнения зависимости решается с помощью того или иного математико-статистического метода обработки данных. Так как параметры уравнений представляют собой выборочные характеристики, то процесс оценивания параметров должен сопровождаться статистической проверкой их значимости (существенности).

6. Линейное уравнение парной регрессии[5]

Пусть между переменными X и Y теоретически существует некоторая линейная зависимость. Однако в действительности отдельные наблюдения X и Y будут отклоняться от линейной зависимости в силу воздействия различных причин. Обычно зависимая переменная находится под влиянием целого ряда факторов, в том числе и не известных исследователю, а также случайных причин. Одним из существенных источников отклонений являются ошибки измерения. Отклонения от предполагаемой формы связи могуть возникнуть также из-за неправильной *спецификации*[6], то есть неправильного выбора вида самого уравнения, описывающего зависимость между переменными.

Предположим, что спецификация выполнена правильно и, учитывая возможные отклонения, линейное уравнение связи (парную регрессию) двух

---

[4] от лат. *Correlatio* – соотношение, взаимосвязь.

[5] От лат. *Regressio* – движение назад. Термин введен английским статистиком Ф.Гальтоном, который, изучая зависимость между ростом родителей и их детей, обнаружил явление «регрессии к среднему» – у детей, родившихся у очень высоких родителей, рост имел тенденцию быть ближе к средней его величине.

[6] *Спецификация* в регрессионном анализе – выбор независимых переменных, существенно влияющих на зависимую величину и определение вида уравнения регрессии. Данная задача решается с помощью качественного анализа изучаемой взаимосвязи.

переменных X и Y представим в виде $Y = \alpha + \beta X + \varepsilon$, где: α и β - неизвестные параметры (коэффициенты) регрессии; ε - случайная переменная, характеризующая отклонение от теоретически предполагаемой регрессии – возмущение.

Таким образом, в данном уравнении значение Y представлено как сумма двух частей – систематической ($\alpha+\beta X$) и случайной ε.

Относительно возмущения ε сделаем следующие предположения:

(I) Возмущение ε является нормально распределенной случайной переменной.

(II) Математическое ожидание ε равно нулю: $M(\varepsilon)=0$.

(III) Дисперия возмущений постоянна: $D(\varepsilon)=const$.

(IV) Последовательные значения ε не зависят друг от друга.

Таким образом, при построении регрессии принимается гипотеза о том, что для каждого наблюдения i справедлива следующая зависимость:

$$y_i = \alpha + \beta x_i + \varepsilon_i .$$

В результате статистического наблюдения исследователь имеет ряд значений независимой переменной X и соответствующих значений зависимой переменной Y:

Y	X
y_1	x_1
y_2	x_2
…	…
y_n	x_n

Задача заключается в определении параметров α и β. Однако истинные значения этих параметров по имеющейся выборке наблюдений ограниченного объема получить нельзя. Поэтому делается попытка определить статистические оценки a и b неизвестных параметров α и β.

7. Метод наименьших квадратов (МНК) для парной регрессии

Согласно МНК, параметры уравнения регрессии подбираются так, чтобы сумма квадратов отклонений наблюдений переменной у от линии регрессии была минимальной:

$$S(a,b)=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2\to\min .$$

Для решения этой задачи найдем частные производные функции $S(a,b)$ по неизвестным параметрам a и b и приравняем их к нулю (необходимое условие существования минимума функции):

$$\begin{cases} \dfrac{\partial S}{\partial a}=-2\sum\limits_{i=1}^{n}\left(y_i-a-bx_i\right)=0; \\ \dfrac{\partial S}{\partial b}=-2\sum\limits_{i=1}^{n}\left(y_i-a-bx_i\right)x_i=0. \end{cases}$$

Поделив уравнение на n получим: $\begin{cases} \dfrac{\sum\limits_{i=1}^{n}y_i}{n}-a-b\dfrac{\sum\limits_{i=1}^{n}x_i}{n}=0; \\ \sum\limits_{i=1}^{n}x_iy_i-a\sum\limits_{i=1}^{n}x_i-b\sum\limits_{i=1}^{n}x_i^2=0, \end{cases}$

$$\begin{cases} \overline{y}=a+b\,\overline{x}; \\ \dfrac{\sum\limits_{i=1}^{n}x_iy_i}{n}-a\,\overline{x}-b\dfrac{\sum\limits_{i=1}^{n}x_i^2}{n}=0 \end{cases},$$

где $\overline{y}=\dfrac{\sum\limits_{i=1}^{n}y_i}{n}$, $\overline{x}=\dfrac{\sum\limits_{i=1}^{n}x_i}{n}$ - выборочные средние случайных величин Y и X.

Итак, $\dfrac{\sum\limits_{i=1}^{n}x_iy_i}{n}-\overline{y}\,\overline{x}+b\,\overline{x}^2-b\dfrac{\sum\limits_{i=1}^{n}x_i^2}{n}=0$, $b\left(\dfrac{\sum\limits_{i=1}^{n}x_i^2}{n}-\overline{x}^2\right)=\dfrac{\sum\limits_{i=1}^{n}x_iy_i}{n}-\overline{y}\,\overline{x}$, то есть $b = \dfrac{\textit{выборочный коэффициент } \operatorname{cov}(X,Y)}{\textit{выборочная } D(X)}$.

Таким образом, $a = \bar{y} - b\bar{x}$; $b = r\dfrac{s(Y)}{s(X)}$, где $s(Y)$, $s(X)$ – выборочные среднеквадратические отклонения случайных величин Y и X.

Свойства оценок параметров, получаемых МНК. Оценки параметров a и b, полученные методом МНК, при условии, что сделанные выше предположения относительно возмущений ε справедливы, обладают свойствами:

1) Оценки параметров являются *несмещенными*, то есть математическое ожидание параметра равно истинному его значению: $M(a)=\alpha$; $M(b)=\beta$.

2) Оценки *состоятельны*, то есть дисперсия оценки параметра стремится к нулю с возрастанием объема выборки n: $\lim\limits_{n\to\infty} \sigma_a^2 = 0$, $\lim\limits_{n\to\infty} \sigma_b^2 = 0$.[7]

3) Оценки являются *эффективными*, то есть они имеют минимальную дисперсию по сравнению с любыми другими линейными и несмещенными оценками параметров α и β.

Величины y_i и x_i в выборке значений переменных Y и X являются случайными. Следовательно, оценки a и b также случайны. Их математические ожидания равны, соответственно, α и β. Чем меньше разброс оценок a и b вокруг их истинных значений α и β, тем более они значимы. Формулы для дисперсий оценок следующие:

$$D(b) = s_b^2 = \frac{s^2}{\sum\limits_{i=1}^{n}\left(x_i - \bar{x}\right)^2}; \qquad D(a) = s_a^2 = \frac{s^2 \sum\limits_{i=1}^{n} x_i^2}{n\sum\limits_{i=1}^{n}\left(x_i - \bar{x}\right)^2},$$

[7] Свойство состоятельности означает, что при увеличении объема наблюдений оценки параметров становятся более надежными, то есть с ростом n оценки все плотнее концентрируются вокруг истинных неизвестных значений параметров.

где $s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$ - мера разброса значений зависимой переменной Y около линии регрессии (*необъясненная дисперсия*); $e_i=y_i-a-bx_i$ - отклонения наблюдаемых значений случайной величины Y от линии регрессии; s_a и s_b - стандартные отклонения оценок коэффициентов a и b.

Проверка статистического качества оцененного уравнения регрессии включает следующие шаги:

- проверка статистической значимости каждого коэффициента регрессии;
- проверка общего качества уравнения регрессии;
- проверка наличия свойств данных, предполагавшихся при оценивании уравнения регрессии.

Если с помощью уравнения регрессии анализируется взаимосвязь экономических переменных, то результаты оценивания должны иметь разумную экономическую интерпретацию, в частности, должны быть получены ответы на следующие вопросы:

- являются ли статистически значимыми объясняющие факторы, важные с теоретической точки зрения;
- являются ли коэффициенты, показывающие направление воздействия этих факторов, положительными или отрицательными и соответствуют ли знаки коэффициентов экономическому смыслу;
- лежат ли оценки коэффициентов регрессии внутри интервалов, предполагаемых из теоретических соображений.

Формальный метод проверки значимости коэффициента регрессии b использует величину отношения b к его стандартной ошибке s_b. Величина $t_b = \frac{b}{s_b}$ (t-статистика для коэффициента b) при условии, что предположения **(I)-(IV)** о возмущениях ε выполнены, имеет t-распределение Стьюдента с n-2 степенями свободы. При заданном уровне значимости α для t-статистики проверяется нулевая гипотеза H_0 о равенстве ее нулю (и, следовательно, о равенстве нулю коэффициента b). Нулевая гипотеза H_0 при заданном уровне значимости отвергается, если $|t_b| > t_{кр}(\alpha;n\text{-}2)$, где $t_{кр}(\alpha;n\text{-}2)$–граница критической области распределения Стьюдента для числа степеней свободы n-2 и уровне значимости α.

Аналогично проверяется гипотеза о равенстве нулю свободного члена α уравнения регрессии $Y=\alpha+\beta X$.

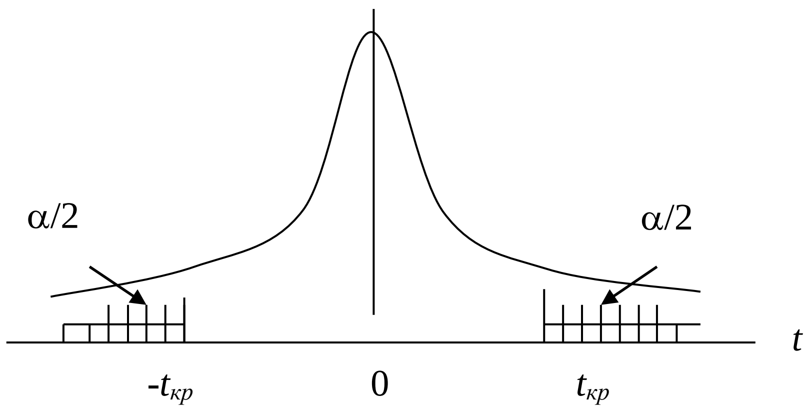


Рис.1. Графическое представление решения

8. Модель множественной линейной регрессии

На значения экономических переменных обычно влияют многие факторы. В этом случае предполагается, что зависимая переменная Y является функцией m объясняющих факторов $X_1,\dots,X_m$ и оцениваются параметры функции $Y=\alpha_0+\alpha_1X_1+\dots+\alpha_mX_m+\varepsilon$, где α_j- коэффициенты регрессии, ε - случайная ошибка. Как и в случае парной регрессии предполагается, что

(I) Возмущение ε является нормально распределенной случайной величиной.

(II) Математической ожидание ε равно нулю: $M(\varepsilon)=0$.

(III) Дисперия возмущений постоянна: $D(\varepsilon)=const$.

(IV) Последовательные значения ε не зависят друг от друга.

Таким образом, задача состоит в нахождении оценок $a_0, a_1, \ldots, a_m$ коэффициентов $\alpha_0, \alpha_1, \ldots, \alpha_m$. Для ее решения обычно используется также метод наименьших квадратов, заключающийся в минимизации функции

$$S(a_0, a_1, \ldots, a_m) = \sum_{i=1}^{n} e_i \to \min,$$

где $e_i = y_i - a_0 - a_1 x_{i1} - \ldots - a_m x_{im}$ – отклонения зависимой переменной Y от линии регрессии; n – объем выборок переменных.

Для обеспечения статистической надежности оценок требуется, чтобы число наблюдений (объем выборок) как минимум в три раза превосходило число оцениваемых параметров.

Если предпосылки 1-4 выполняются, то МНК обеспечивает несмещенность, состоятельность и эффективность оценок.

Приравнивая к нулю частные производные функции S по всем a_j, мы получаем так называемую *систему нормальных уравнений*, состоящую из (m+1) линейных уравнений с (m+1) неизвестными. Такая система обычно имеет единственное решение.

При проверке нулевой гипотезы отдельно для каждого коэффициента a_j рассчитываются t-статистики: $t_{a_j} = \frac{a_j}{s_{a_j}}$, где s_{a_j} - стандартная ошибка для коэффициента a_j. Они имеют распределение Стьюдента с (n-m-1) степенями свободы. Процедура проверки статистической значимости коэффициентов множественной линейной регрессии такая же, как и в случае парной регрессии.

9. Проверка общего качества уравнения регрессии

Для проверки общего качества уравнения регрессии обычно используется коэффициент детерминации R^2, который в случае парной регрессии равен квадрату коэффициента корреляции между переменными X и Y. Коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 1 - \frac{\textit{Остаточная дисперсия}}{\textit{Общая дисперсия}} =$$

$$= \frac{\textit{Общая дисперсия} - \textit{Остаточная дисперсия}}{\textit{Общая дисперсия}} = \frac{\textit{Объясненная дисперсия}}{\textit{Общая дисперсия}}$$

Таким образом, коэффициент детерминации представляет собой долю объясненной дисперсии в общей дисперсии зависимой переменной Y.

Для определения статистической значимости коэффициента детерминации проверяется нулевая гипотеза для F-статистики, вычисляемой по формуле:

$$F = \frac{R^2}{1 - R^2}\frac{n - m - 1}{m}.$$

Смысл проверяемой гипотезы заключается в равенстве нулю всех коэффициентов регрессии за исключением свободного члена. Если они действительно равны нулю в генеральной совокупности, то уравнением регрессии является $Y = \bar{y}$, и коэффициент детерминации R^2 и F-статистика также равны нулю. Итак, нулевая гипотеза H_0: $R^2=0$ равносильна гипотезе H_0: $\alpha_1 = \ldots = \alpha_n = 0$.

Если предположения **(I)-(IV)** имеют место, то при выполнении нулевой гипотезы величина F имеет распределение Фишера с (m; n-m-1) степенями свободы. При принятом уровне значимости α для распределения Фишера находится критическое значение $F_{крит}$ такое, что $P(F>F_{крит})=\alpha$. Нулевая гипотеза отвергается, если $F>F_{крит}$. В случае парной линейной регрессии проверка нулевой гипотезы для t-статистики коэфициента регрессии эквивалентна проверке нулевой гипотезы для F-статистики.

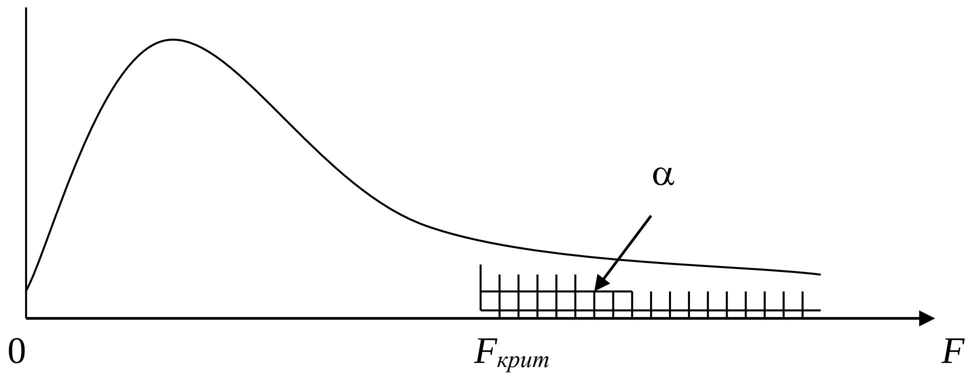


Рис. 2.Графическое представление

Пример 5. Рассмотрим уравнение кривой Филлипса для США, параметры которого оценены на основе годовых данных 1994-1966 гг., 13 наблюдений:

π=6.29-0.76u, R^2=0.66

(здесь π – темп инфляции, u - уровень безработицы).

Отсюда следует, что $F = \frac{0.66}{0/34}\frac{11}{1} \approx 21.4$. Для степеней свободы (1;11) при 5% уровне значимости (доверительная вероятность 95%) $F_{крит}$=4.84, при 1% уровне значимости $F_{крит}$=9.65. Поскольку F=21.4>$F_{крит}$, то нулевая гипотеза H_0: R^2=0 отвергается в обоих случаях.

10. Проверка некоторых предполагавшихся свойств случайных возмущений

Необходимо отметить, что близкое к единице значение коэффициента детерминации R^2 еще не является свидетельстом высокого качества уравнения регрессии. Коэффициент детерминации может быть относительно высоким, однако модель может оказаться непригодной для прогнозирования. Это обычно происходит в результате того, что взаимосвязь между зависимым экономическим показателем Y и объясняющими (независимыми) переменными X_i оказывается явно нелинейным. В этом случае можно утверждать, что не выполнены необходимые предпосылки об отклонениях e_i значений показателя Y от линии регрессии. Если эти отклонения не являются взаимно независимыми и дисперсия их непостоянна, то такое нарушение исходных предпосылок, сделанных выше, свидетельствует о неточной спецификации (определения вида зависимости) уравнения регрессии, а также свидетельствуют о неточности полученных оценок коэффициентов регрессии и их стандартных ошибок. Поэтому при оценке качества уравнения регрессии существенна проверка его некоторых важных свойств, которые предполагались при оценке параметров этого уравнения. Если, например, реальная связь между экономическим показателем Y и объясняющими переменными $X_1,X_2,,...,X_m$ нелинейна, то анализ статистической значимости коэффициентов регрессии неточен и оценки этих коэффициентов не обладают желательными для исследователя свойствами, как несмещенность, состоятельность и эффективность.

Закономерность в поведении остатков $e_i=y_i-a_0-a_1x_{i1}-a_2x_{i2}-...-a_mx_{im}$, $i=1,...,n$, - отклонений выборочных значений показателя Y от линии регрессии) выражается, как правило, в знаке каждых соседствующих отклонений, что может являться следствием нелинейного характера связи переменных или воздействием какого-то фактора, не включенного в уравнение регрессии. Это может быть причиной того, что существует возможность улучшить уравнение регрессии путем оценивания другой нелинейной формулы уравнения или путем включения нового объясняющего экономического фактора.

Одним из основных предполагаемых свойств отклонений ε_i выборочных значений y_i от регрессионной формулы является их статистическая независимость между собой. Поэтому после оценки параметров уравнения регрессии необходимо проверить статистическую независимость отклонений e_i. При этом проверяется обычно их некоррелированность, которая, вообще говоря, является необходимым, но недостаточным условием независимости[8]. Для проверки

[8] В предположении, что отклонения e_i подчиняются нормальному закону распределения, некоррелированность остатков является также и достаточным условием независимости.

некоррелированности соседних значений остатков e_i можно использовать их коэффициент автокорреляции первого порядка[9]

$$r_{i,i-1} \approx \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sqrt{\sum_{i=1}^{n} e_i^2 \sum_{i=2}^{n} e_{i-1}^2}}$$

и проверить его статистическую значимость, например, с помощью t-статистики $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, которая имеет t-распределение Стьюдента с n-2 степенями свободы. При этом, используя закон распределения Стьюдента, для заданного уровня значимости α проверяется нулевая гипотеза о равенстве нулю коэффициента корреляции.

На практике эконометрических исследований для проверки наличия или отсутствия автокорреляции остатков обычно используют тесно связанную с коэффициентом автокорреляции первого порядка $r_{i,i-1}$ статистику Дарбина-Уотсона, рассчитываемую по формуле

$$DW = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}.$$

Так как $DW = \frac{\sum_{i=2}^{n} e_i^2 - 2\sum_{i=2}^{n} e_i e_{i-1} + \sum_{i=2}^{n} e_{i-1}^2}{\sum_{i=1}^{n} e_i^2}$ и при больших объемах выборки n имеет место $\sum_{i=1}^{n} e_i^2 \approx \sum_{i=2}^{n} e_i^2 \approx \sum_{i=2}^{n} e_{i-1}^2$, то $DW \approx 2(1-r_{i,i-1})$. Если e_i в точности равны e_{i-1}, то DW=0. Если $e_i=-e_{i-1}$, то DW=4. В остальных случаях $0<DW<4$.

[9] Первый порядок означает, что вычисляется коэффициент автокорреляции между соседними остатками.

Таким образом, когда каждое отклонение e_i примерно совпадает с предыдущим отклонением e_{i-1}, то $\sum_{i=2}^{n} e_i^2 \approx \sum_{i=2}^{n} e_{i-1}^2 \approx \sum_{i=2}^{n} e_i e_{i-1}$. Следовательно, числитель статистики Дарбина-Уотсона в этом случае близок к нулю. В этом случае говорят о наличии положительной автокорреляции остатков первого порядка. Положительная автокорреляция остатков первого порядка имеет место, как правило, в случае, если связь между зависимым экономическим показателем Y и независимыми переменными $X_1, X_2, \ldots, X_n$ является нелинейной.

Другой крайний случай возникает, когда точки наблюдений показателя Y поочередно отклоняются в разные стороны от линии регрессии, то есть когда каждое следующее отклонение e_i имеет, как правило, противоположный знак, чем предыдущее отклонение e_{i-1}. Тогда $e_i - e_{i-1} \approx 2e_i$ и $DW \approx \frac{\sum_{i=2}^{n} (2e_i)^2}{\sum_{i=1}^{n} e_i^2} = 4\frac{\sum_{i=2}^{n} e_i^2}{\sum_{i=1}^{n} e_i^2} \approx 4$.

В этом случае говорят о наличии отрицательной автокорреляции остатков первого порядка. Отрицательная автокорреляция остатков первого порядка имеет место, как правило, в случае, если существует сезонный характер изменений. Поэтому при работе с годовыми данными отрицательная автокорреляция остатков первого порядка, как правило, не наблюдается. Теоретически отрицательная автокорреляция остатков первого порядка может возникнуть при работе с квартальными или месячными данными. Однако опыт показывает, что в экономических исследованиях отрицательная автокорреляция остатков имеет место крайне редко.

Если характер поведения отклонений e_i случаен, можно предположить, что в половине случаев знак соседних отклонений совпадает, а в половине случаев - различен. Тогда:

$$DW \approx \frac{\sum_{i=2}^{n} 0.5(2e_i)^2}{\sum_{i=1}^{n} e_i^2} \approx 2\,.$$

Это показывает, что близость статистистики Дарбина-Уотсона DW к двум является необходимым условием случайного характера отклонений от линии регрессии[10]. Если статистика Дарбина-Уотсоеа близка к двум, можно считать отклонения наблюдений экономического показателя Y от линии регрессии случайными[11]. Это значит, что оцененная функция отражает реальную взаимосвязь и не осталось существенных неучтенных факторов, влияющих на показатель Y.

При заданном уровне значимости α для статистики Дарбина-Уотсона существуют два критических значения, меньших двух: нижнее $d_н$ как граница для принятия гипотезы о наличии положительной автокорреляции остатков первого порядка и верхнее $d_в$ для признания отсутствия положительной автокорреляции. Для проверки гипотезы об отрицательной автокорреляции остатков используется критическая область ($4\text{-}d_н$; 4). Гипотеза же об отсутствии автокорреляции остатков первого порядка принимается, если расчетное значение статистики Дарбина-Уотсона попадает в интервал ($d_в$; $4\text{-}d_в$):

```
|_____|______|_____|______|_______|_______|
0    d_н    d_в    2    4-d_в   4-d_н     4
```

В случае, если коэффициент Дарбина-Уотсона DW попадает в интервалы ($d_н$; $d_в$), ($4\text{-}d_в$; $4\text{-}d_н$), никакая из трех гипотез (об отсутствии автокорреляции остатков, о положительной автокорреляции остатков, об отрицательной

---

[10] В показателе Дарбина-Уотсона DW сравниваются только соседние значения отклонений от линии регрессии. Циклы же изменения исследуемых экономических показателей могут быть более или менее длительными, чем единица времени, с которой работает исследователь (год, квартал, месяц, неделя и т.д.). Поэтому в зависимости от периода времени статистика Дарбина-Уотсона может принимать разные статистические значения и к этим значениям следует относиться с известной долей осторожности, интерпретируя их не только по количественному, но и по качественному признаку.

[11] Хотя в действительности они могут и не быть таковыми, так как равенство двум коэффициента Дарбина-Уотсона не гарантирует независимость остатков первого порядка. Это лишь необходимое (но не достаточное) условие независимости.

автокорреляции остатков) при заданном уровне значимости α не может быть ни принята, ни отвергнута.

При объясняющих переменных не более 3 и достаточном числе наблюдений (не меньше 12), если статистика Дарбина-Уотсона составляет 1.5-2.5 для уровня значимости α=0.05 принимается гипотеза об отсутствии автокорреляции остатков первого порядка.

В случае наличия автокорреляции остатков полученная формула регресии обычно считается неудовлетворительной. В этом случае приходится искать другую формулу функции, включить неучтенные факторы, применить к данным уменьшающее автокорреляцию остатков преобразование (например, метод скользящих средних) или разбить наблюдения на части, с тем, чтобы для первого периода получить одно уравнение регрессии, а для второго периода - другое.

<u>Пример 6.</u> Для уравнения кривой Филлипса для США, параметры которого оценены на основе годовых данных 1994-1966 гг., 13 наблюдений: π=6.29-0.76u статистика Дарбина-Уотсона DW=0.965. При уровне значимости α=5% $d_н$=0.89, $d_в$=1.21. Гипотеза об отсутствии автокорреляции остатков первого порядка здесь принимается при $d_в=1.21<DW<2.79=4-d_в$ и отвергается при $DW<0.89=d_н$ или $DW>3.11=4-d_н$. Поскольку в нашем случае DW лежит между $d_н$ и $d_в$, гипотеза об отсутствии автокорреляции остатков не может быть ни принята, ни опровергнута.

11.Задачи, упражнения и тестовые задания

11. 1 Случайные величины и их числовые характеристики

Вопросы для обсуждения:

1. Что такое случайная величина? Что такое распределение случайной величины? Как может быть задан закон распределения дискретной (непрерывной) случайной величины? Приведите простые примеры случайных величин; опишите законы их распределения.

2. В чем смысл анализа экономических показателей как случайных величин? Приведите примеры случайных событий и процессов в экономике н микро- и макроуровне; охарактеризуйте законы их распределения.

3. Как определяется функция распределения случайной величины? Когда закон распределения случайной величины называется равномерным? Каковы его функции распределения и плотности вероятности?

4. Какое распределение случайной величины называется нормальным? Каковы его статистические характеристики? Почему нормальное распределене часто встречается в экономических задачах? Приведите примеры.

5. Каковы определение и свойства математического ожидания случайной величины?

6. Каковы определение и свойства дисперсии случайной величины?

7. Выведите формулу связи дисперсии с математическими ожиданиями случайной величины и ее квадрата.

8. Как рассчитать вероятность попадания дискретных и непрерывных случайных величин в интервал: $Prob(a \le X < b)$:

а) с помощью функции распределения;

б) с помощью плотности распределения вероятности для непрерывной СВ и с помощью функции вероятности для дискретной СВ?

9. В каких задачах эконометрики применяется распределение Стьюдента?

10. В каких задачах эконометрики применяется распределение Фишера?

11. Что такое ковариация и коэффициент корреляции двух случайных величин? Какой тип связи случайных величин они характеризуют? Какова их размерность?

12. Как соотносятся понятия некоррелированности и независимости двух случайных величин?

13. О чем говорит знак коэффициента корреляции?

14. Что такое генеральная совокупность? Что такое выборка? Может ли выборка совпадать с генеральной совокупностью? Какие виды выборок Вы знаете? Приведите примеры для экономических показателей “число произведенных автомобилей”, “объем произведенного валового внутреннего продукта”. Какие виды эмпирических зависимостей могут исследоваться с помощью этих выборок? Чем выборочное среднее отличается от математического ожидания? Чем выборочная дисперсия отличается от дисперсии генеральной совокупности?

15. Какие составляющие временных рядов экономических показателей Вы знаете? Приведите примеры показателей, динамические ряды которых могут включать все перечисленные компоненты. Могут ли какие-либо из этих компонентов отсутствовать? Приведите соответствующие примеры.

16. Каков закон распределения выборочного коэффициента корреляции, если истинное его значение равно нулю? Как проверяется нулевая гипотеза для коэффициента корреляции?

17. Охарактеризуйте общую схему проверки статистических гипотез на примере проверки нулевой гипотезы для коэффициента корреляции.

18. Каковы преимущества работы со средними величинами по сравнению с работой с отдельными наблюдениями?

а) выше надежность получаемых результатов;

б) может быть использован закон больших чисел;

в) меньше разброс получаемых величин;

г) все перечисленное верно.

Задачи и тестовые задания.

1. Случайная величина X имеет распределение

X	-2	-1	0	1	2
P_X	1/12	1/6	1/2	1/6	1/12

Укажите неверное утверждение среди перечисленных:

1) M[X]=0 2) D[X]=1 3) $M[X^2]=2/3$ 4) $M[X^3]=0$ 5) $M[X^5]=0$

2. Случайная величина X имеет распределение

X	0	1	2	3
P	1/12	1/6	1/4	1/2

Укажите неверное утверждение.

а) M[X]=13/6

б) $M[(X-1)^2]=8/3$

в) $M[(X+1)^2]=11$

г) $M[(X/2)^2]=17/12$

д) $M[X^5]\geq 0$.

3. Случайная величина X имеет следующую функцию распределения

$$F(x)=\begin{cases} 0, & x\leq 0; \\ x^2, & 0<x\leq 1; \\ 1, & x>1. \end{cases}$$

Укажите неверное утверждение среди перечисленных:

1) Prob{1/2≤X<1}=3/4; 2) Prob{0≤X<1/2}=1/4;

3) Prob{1/3≤X<2/3}=1/3; 4) Prob{2/3≤X<2}=4/9;

5) Prob{ -1≤X<0}=0.

4. Случайная величина X имеет следующую функцию распределения

$$F(x)=\begin{cases} 0, & x<1; \\ 1-\dfrac{1}{x}, & x\geq 1. \end{cases}$$

Значение функции плотности вероятности при x=2 равно

1) 1/4 2) 1/2 3) 0 4) -1/2 5) 1/8

5. Вычислите математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X, распределение которой приведено в таблице.

X	0	1	2	3
P	0.1	0.4	0.4	0.1

Постройте график функции распределения случайной величины X. Найдите вероятность того, что $x\leq 2$.

6. Случайная величина X имеет следующую функцию распределения

$$F(x)=\begin{cases} 0, & x\leq 0; \\ \dfrac{x^2}{4}, & 0<x\leq 2; \\ 1, & x>2. \end{cases}$$

Найдите: а) Prob{X≥1}, б) Prob{0.5<X≤1}. Запишите аналитически функцию плотности распределения вероятностей для случайной величины X. Вычислите математическое ожидание и дисперсию.

7. Инвестиции приносят доход в $600 с вероятностью 1/3 и $1200 с вероятностью 2/3. Каковы ожидаемое значение и дисперсия получаемого дохода (в тысячах долларов)?

8. Из тысячи лотерейных билетов на 250 выпадает выигрыш $1, на 10 билетов - $10, на 1 билет - $150. Остальные билеты без выигрыша. Найдите математическое ожидание и дисперсию выигрыша, если Вы купили 1 билет.

9. Пусть X - стоимость пакета акций некоторой компании, которая ведет себ как случайная, нормально распределенная величина. За длительный период наблюдения за стоимостью акций этой компании было установлено, что средняя стоость пакета акций (m_X) составила \$150, а выборочное стандартное отклонение (s_X) составило \$25. Оцените, в каком интервале будет находиться стоимость пакета акций в текущем году с вероятностью 0.8?

10. Оцените вероятность того, что краткосрочная процентная ставка r в стране X в очередном году превысит 20%, если за достаточно длительный период установлено, что m_r=4%. Оцените вероятность вышеупомянутого события, если доподлинно известно, что:

а) σ_r=1 процентному пункту;

б) σ_r=4 процентным пунктам.

11. Могут ли совпадать средние значения, рассчитанные для генеральной совокупности и выборки из этой совокупности?

а) да, это возможно;

б) нет, это не может случиться никогда;

в) это возможно только в том случае, когда выборка совпадает с генеральной совокупностью;

г) это верно в любом случае.

11.2 Статистические взаимосвязи экономических переменных

Вопросы для обсуждения:

1. Приведите примеры статистических взаимосвязей макроэкономических переменных. Сформулируйте проблему их эконометрического анализа.

2. Что такое ковариация и коэффициент корреляции двух случайных величин? Какой тип связи случайных величин они характеризуют? Какова их размерность?

3. Как соотносятся понятия некоррелированности и независимости двух случайных величин?

4. О чем говорит знак коэффициента корреляции?

5. Каков закон распределения выборочного коэффициента корреляции, если истинное его значение равно нулю? Как проверяется нулевая гипотеза для коэффициента корреляции?

6. Какое распределение называется распределением Стьюдента и как оно используется в корреляционном анализе?

7. Опишите общую схему проверки статистических гипотез на примере проверки нулевой гипотезы для коэффициента корреляции.

8. Что такое линейная регрессия?

9. Что такое спецификация и параметризация уравнения регрессии? Как они осуществляются?

Задачи и тестовые задания:

1. Пусть X и Y - две независимые нормально распределенные случайные величины. Пусть M[X]=1; M[Y]=2; D[X]=1; D[Y]=2. Коэффициент ковариации между двумя случайными переменными U=X+Y и V=Y-X равен:

a) 3 b) 1 c) 0 d) -3

2. В случае, когда значение статистики выходит за границы критического интервала, делается вывод о том, что:

a) истинна нулевая гипотеза;

b) истинна альтернативная гипотеза;

c) ложна альтернативная гипотеза;

d) ничто из перечисленного выше не выполняется.

3. Из генеральных совокупностей дохода и потребления получены соответствующие выборки объемом 66 наблюдений. Выборочный коэффициент корреляции между доходом и потреблением оказался равным r=-0.159. Можно ли при уровне значимости α=0.05 считать, что наблюдаемые переменные отрицательно коррелированы?

4. Укажите верное утверждение среди приведенных ниже:

a) если коэффициент корреляции двух случайных величин равен единице, то они статистически независимы;

b) если коэффициент корреляции двух случайных величин близок к нулю, то они статистически независимы;

c) если коэффициент корреляции двух случайных величин равен нулю, то они статистически независимы;

d) если две случайные величины статистически независимы, то их коэффициент корреляции равен единице;

e) если две случайные величины статистически независимы, то их коэффициент корреляции равен нулю;

5. Оценена функция потребления для России за 1990-1996 годы:

$$C=2270+0.6Y.$$

Здесь C - объем реального личного потребления, Y - реальный ВВП. Оценки свободного члена уравнения и коэффициента регрессии статистически значимы. Показатели общего качества регрессии вполне приемлемы. Что можно сказать о соответствии этой зависимости теории потребления Кейнса?

6. Выборочный коэффициент корреляции между темпом инфляции и долей инвестиций в ВВП в Нидерландах в 1973-1992 годах (20 наблюдений) составил 0.56. Критическое значение коэффициента корреляции при 5%-м уровне значимости при двусторонней альтернативной гипотезе равно 0.44; при односторонней 0.38. Выберите один из перечисленх ниже выводов:

a) темп инфляции и доля инвестиций в ВВП в Нидерландах были в 1973-1992 годах независимыми;

b) темп инфляции и доля инвестиций в ВВП в Нидерландах были в 1973-1992 годах линейно независимыми;

c) существовала отрицательная линейная связь темпа инфляции и доли инвестиций в ВВП в Нидерландах в 1973-1992 годах;

d) существовала положительная линейная связь темпа инфляции и доли инвестиций в ВВП в Нидерландах в 1973-1992 годах;

e) так быть не могло, поскольку критическое значение коэффициента корреляции при одном и том же уровне значимости при двусторонней альтернативной гипотезе должно быть меньше, чем при односторонней.

7. Коэффициент корреляции r между долей инвестиций в ВНП и реальной краткосрочной ставкой процента в экономике США за 1971-1990 годы (20 погодовых наблюдений, т.е. n=20) составил -0.35.

Статистика Стьюдента $t_{18} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = -1.585.$ Из таблиц можно видеть:

$P\{|t_{18}|>1.33\}=0.2$; $P\{|t_{18}|>1.73\}=0.1$; $P\{|t_{18}|>2.1\}=0.05$.

Укажите верное утверждение среди перечисленных:

a) если значение r для генеральной совокупности равно нулю, то вероятность получения такого (или большего по моделю) коэффициента корреляции выборки равна здесь примерно 14%;

b) отрицательный коэффициент корреляции свидетельствует об убывании доли инвестиций в ВНП при росте процентной ставки;

c) то, что модуль коэффициента корреляции меньше 0.5, говорит об отсутствии связи рассматриваемых показателей;

d) то, что t-статистика Стьюдента по модулю больше единицы, говорит о наличии статистически значимой связи показателей;

e) то, что t-статистика Стьюдента по модулю меньше двух, говорит об отсутствии статистически значимой связи показателей.

Задачи для самостоятельного решения:

1. Оценена функция потребления для России за 1985-1991 годы:

$$C=-82.4+0.62Y.$$

Здесь C - объем реального личного потребления, Y - реальный ВВП. Оценка свободного члена уравнения статистически незначима, оценка коэффициента регрессии статистически значима. Показатели общего качества регрессии вполне приемлемы. Что можно сказать о соответствии этой зависимости теории потребления Кейнса? Какие шаги можно предпринять в дальнейшем исследовании этой зависимост?

2. Выборочный коэффициент корреляции между темпом инфляции и долей инвестиций в ВВП в Канаде в 1973-1992 годах (20 наблюдений) составил -0.3. Критичкое значение коэффициента корреляции при 5%-м уровне значимости при двусторонней альтернативной гипотезе равно 0.44; при односторонней 0.38. Выберите один из перечисленных ниже выводов:

a) темп инфляции и доля инвестиций в ВВП в Канаде были в 1973-1992 годах независимым;

b) темп инфляции и доля инвестиций в ВВП в Канаде были в 1973-1992 годах линейно независимым;

c) существовала отрицательная линейная связь темпа инфляции и доли инвестиций в ВВП в Канаде в 1973-1992 годах;

d) существовала положительная линейная связь темпа инфляции и доли инвестиций в ВВП в Канаде в 1973-1992 годах;

e) так быть не могло, поскольку критическое значение коэффициента корреляции при одном и том же уровне значимости при двусторонней альтернативной гипотезе должно быть меньше, чем при односторонней.

3. Коэффициент корреляции r между уровнем инфляции (по ВНП) и скоростью обращения денег M1 в экономике США за 1931-1990 годы (60 погодовых наблюдений, т.е. n=60) составил 0.29.

Статистика Стьюдента $t_{58} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = -2.31$. Из таблиц можно видеть:

$P\{|t_{58}|>2\}=0.05$; $P\{|t_{58}|>2.39\}=0.02$; $P\{|t_{58}|>2.66\}=0.01$.

Укажите верное утверждение среди перечисленных:

a) положительный коэффициент корреляции свидетельствует о возрастании скорости обращения денег при росте уровня инфляции;

b) то, что модуль коэффициента корреляции меньше 0.5, говорит об отсутствии значимой связи рассматриваемых показателей;

c) то, что t-статистика Стьюдента по модулю больше двух, говорит о наличии статистически значимой связи показателей;

d) если значение r для генеральной совокупности равно нулю, то вероятность получения такого (или большего по модулю) коэффициента корреляции выборки равна здесь примерно 2.5%;

e) вероятность того, что коэффициент корреляции генеральной совокупности не равен нулю, равна здесь примерно 98.5%.

11.3. Модель линейной регрессии в экономике: оценивание и статистический анализ

Вопросы для обсуждения:

1. Каковы возможности и задачи регрессионного анализа зависимостей экономических переменных?

2. Какими могут быть критерии качества оценки линейной регрессии?

3. В чем сущность метода наименьших квадратов (МНК)?

4. Каковы предпосылки о свойствах отклонений зависимой переменной от теоретической линии регрессии?

5. Обладают ли оценки параметров линейной регрессии, полученные с помощью МНК, свойствами несмещенности, состоятельности и несмещенности?

6. В чем различие, смысловое и количественное, теоретических значений коэффициентов регрессии α и β и их оценок a и b?

7. Какие факторы влияют на величину стандартных ошибок коэффициентов a и b?

8. Как связан коэффициент регрессии b с коэффициентом корреляции величин x и y?

9. Имеют ли коэффициенты a и b размерность?

10. Какой показатель характеризует долю объясненной с помощью регрессии дисперсии в общей дисперсии зависимой переменной?

11. Каким образом проверяется нулевая гипотеза для коэффициента регрессии b?

12. Стандартная ошибка коэффициента b равна b/2. Можно ли в этом случае говорить о наличии зависимости y от x? Если можно, то что именно?

13. Коэффициент детерминации R^2=0.5. Что можно сказать о качестве оцененной формулы в целом? Какая нужна дополнительная информация?

14. Если нулевая гипотеза для статистики Фишера отвергается, то что можно сказать про оцененную парную линейную регрессию?

15. Таблицы каких распределений используются при оценке качества линейной регрессии?

16. В каких случаях наблюдается положительная автокорреляция отклонений e_i? Приведите примеры из экономики.

17. Какова связь статистики Дарбина-Уотсона и коэффициента автокорреляции отклонений e_i?

18. Статистика Дарбина-Уотсона оказалась близкой к четырем. Что это означает?

19. Если y зависит от x как квадратичная функция $y=x^2$, но оценена связывающая их линейная регрессия, то какой окажется величина DW статистики Дарбина-Уотсона?

Задачи и тестовые задания

1. Метод наименьших квадратов обладает, по сравнению с другими метдами, преимуществом:

а) задача сводитсям к решению системы линейных уравнений;

б) он дает наименьшую сумму модулей отклонений переменной y от линии регрессии;

в) задача сводится к решению квадратного уравнения;

г) он обеспечивает некоррелированность отклонений e_i;

д) он обеспечивает совпадение оценок a и b с теоретическими значениями коэффициентов регрессии α и β.

2. Пусть случайные величины X и Y связаны в генеральной совокупности зависимостью $Y=\alpha+\beta X+\varepsilon$, где $\alpha=11$; $\beta=0.8$; ε - случайная ошибка с нулевым математическим ожиданием и постоянной дисперсией. Тогда по данным случайной выборки с помощью МНК будет оценено уравнение:

а) $Y=11+0.8X+e$, где e - оценка ε;

б) $Y=a+bX+e$, где $M[a]=\alpha$; $M[b]=\beta$;

в) $Y=a+bX+e$, где $a=M[\alpha]$; $b=M[\beta]$;

г) $Y=a+bX+e$, где $a\rightarrow\alpha$; $b\rightarrow\beta$ при $n\rightarrow\infty$;

д) $Y=a+bX+e$, где a имеет то же распределение, что и α, а b - то же распределение, что и β.

3. Полученная с помощью МНК оценка коэффициента парной линейной регрессии (b) тем надежнее, то есть среднеквадратическое отклонение s_b меньше, чем:

а) меньше сумма квадратов отклонений x_i от среднего $\bar{x}=\sum x_i/n$, т.е. чем меньше $\sum(x_j-\bar{x})^2$;

б) меньше необъясненная дисперсия $S^2=\sum e_i^2/(n-2)$;

в) больше среднеквадратическая величина независимой переменной $\sum e_i^2/n$;

г) меньше число наблюдений.

4. Если переменная x принимает среднее в выборке значение $\bar{x}=\sum x_i/n$, то:

а) наблюдаемая величина зависимой переменной y равна значению $\bar{y}=\sum y_i/n$;

б) рассчитанная по уравнению регрессии величина y равна значению $\bar{y}=\sum y_i/n$;

в) рассчитанная величина y в среднем равна $\bar{y}$, но необязательно равна ему в каждом случае;

г) равенство $\bar{y}=a+b\bar{x}$ говорит об отсутствии автокорреляции остатков e_i;

д) равенство $\bar{y}=a+b\bar{x}$ говорит о выполнении предпосылок МНК для остатков e_i;

5. По 20 наблюдениям парной линейной регрессии $y=-12.5+0.38x$ стандартная ошибка коэффициента b (равного 0.38) составила 0.25. Уровень значимости для t-статистики равен 0.05. Не прибегая к таблицам, по общему порядку цифр ответьте на вопрос: нулевая гипотеза для коэффициента регрессии b:

а) отвергается;

б) не отвергается;

в) принимается;

г) не принимается и не отвергается;

д) число наблюдений недостаточно для вывода.

6. Пусть линия $y=2\bar{y}$ (где $\bar{y}$ - среднее выборочное значение зависимой переменой y) является линией регрессии, полученной по методу наименьших квадратов. Тогда:

а) коэффициент детерминации R^2 близок к 0;

б) коэффициент детерминации R^2 близок к 1;

в) статистика Дарбина-Уотсона близка к 2;

г) статистика Дарбина-Уотсона близка к 0;

д) так быть не могло, поскольку получена смещенная оценка свободного члена регрессии.

7. Статистика Дарбина-Уотсона DW для уравнения линейной регрессии оказлась близкой к 2. От сюда вытекает, что:

а) отклонения e_i и e_j статистически независимы для любых $i \neq j$;

б) коэффициент корреляции (e_i, e_{i+k}) близок к нулю для всех $k \geq 1$;

в) коэффициент корреляции (e_i, e_{i+k}) близок к единице;

г) коэффициент корреляции (e_i, e_{i+k}) близок к нулю;

д) ничего из перечисленного сказать нельзя.

8. Пусть модель кривой Филлипса с инфляционными ожиданиями имеет вид: $\pi=\pi(-1)+a(u^*-u)$, где π - темп инфляции (%), u - уровень безработицы (%), (-1) - индекс предыдущего года, u^* - естественный уровень безработицы. Для страны Z оценено статистически значимое уравнение регрессии: $\Delta\pi=9-2u$, где $\Delta\pi=\pi-\pi(-1)$. В этом случае естественный уровень безработицы в данной стране равен:

а) 2% б) 4.5% в) 5% г) 6% д) 9%

9. Оценена функция потребления для России за 1985-1991 годы:

$C=-82.4+0.62Y$; $R^2=0.74$; $DW=2.04$.

Стандартная ошибка свободного члена равна 0.16. Здесь C - объем реального личного потребления, Y - реальный ВВП. В качестве следующего шага можно предложить:

а) использовать метод скользящих средних для устранения афторегрессии остатков;

б) ввести уровень процентной ставки как дополнительную объясняющую переменную;

в) оценить зависимость нелинейной регрессии;

г) оценить линейную регрессию с постоянной средней склонностью к потреблению;

д) полученное уравнение приемлемо по всем параметрам.

10. Оценена парная линейная регрессия y=a+bx+e по 30 годовым наблюдениям. Коэффициент детерминации R^2 оказался близким к 0, статистика Дарбина-Уотсона DW - также близкой к нулю. Затем 30-летний период разбили на два 15-летних и оценили регрессию по той же формуле для каждого из них. Коэффициент R^2 для каждого из периодов оказался близким к 1, статистика Дарбина-Уотсона - к двум. Изобразите на графике в координатах x,y, какой могла быть динамика этих величи в последовательности 30 наблюдений.

11. На рисунках (А)-(Г) показана линия парной регрессии y=a+bx. Отклонения e_i на всех этих рисунках не являются независимыми. На каком из них статистика Дарбина-Уотсона DW≈2?

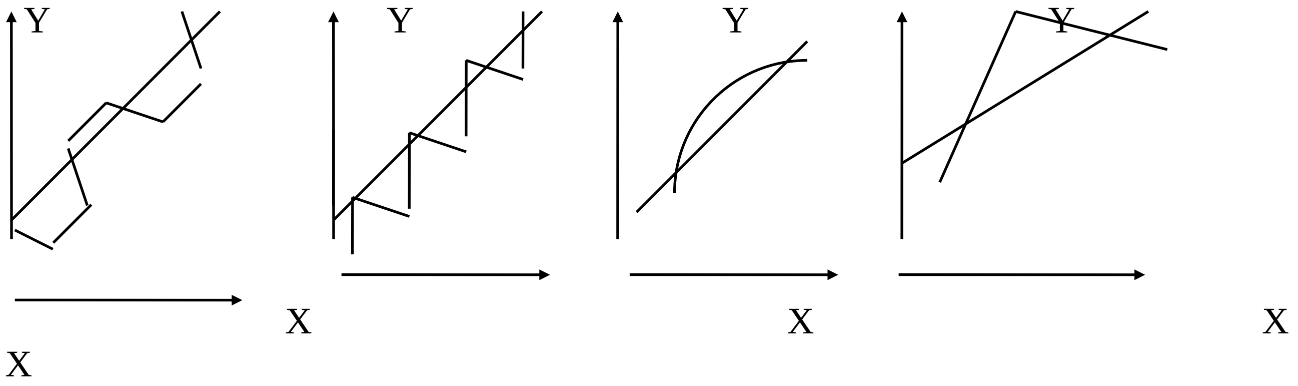


Рисунок А Рисунок Б Рисунок В Рисунок Г

12. Проведите комплексный анализ качества уравнений равновесия денежного рынка в США за 1931-1960 гг. и 1961-1990 гг. (здесь: $\left(\dfrac{M}{P}\right)$ -

реальное предложение денег *M1*, *GNP* - реальный ВНП, *RSR*- реальная краткосрочная ставка процента; в скобках приведены стандартные ошибки коэффициентов).

Для 1931-1960 гг.

$\left(\frac{M}{P}\right)$ = 0.345*GNP* - 7.416*RSR;* R^2=0.83; *DW*=0.35; *F*=137.9.

(σ): (0.009) (1.964)

Для 1961-1990 гг.

$\left(\frac{M}{P}\right)$ = 0.175*GNP* - 4. 857*RSR;* R^2=-1.97; *DW*=0.038; (*F* не рассчитано, т.к. R^2<0).

(σ): (0.007) (6.041)

Какие шаги по уточнению модели Вы предложили бы в каждом из этих случаев?

Задачи и тесты для самостоятельного решения:

1. Гипотеза о наличии автокорреляции остатков первого порядка отвергается, если:

а) коэффициент детерминации R^2 близок к единице;

б) *t*-статистика Стьюдента превышает некоторую критическую величину;

в) статистика Дарбина-Уотсона *DW* близка к 0;

г) статистика Дарбина-Уотсона *DW* близка к 4;

д) статистика Дарбина-Уотсона *DW* близка к 2.

2. Оценивается парная линейная регрессия $y=a+bx$ (x - независимая переменная, y - зависимая переменная). Пусть $e_j=y_j-a-bx_j$; $x_j'=x_j-\bar{x}$; $y_j'=y_j-\bar{y}$; $\bar{x}$, $\bar{y}$ - средние значения x и y. Статистика Дарбина-Уотсона рассчитывается по формуле:

а) $\dfrac{\sum_{j=2}^{n}\left(e_j - e_{j-1}\right)^2}{\sum_{j=1}^{n}\left(e_j\right)^2}$ б) $1-\dfrac{\sum_{j=1}^{n} e_j^2}{\sum_{j=1}^{n}\left(y_j^{'}\right)^2}$

в) $\dfrac{\sum_{j=2}^{n} e_j e_{j-1}}{\sum_{j=1}^{n}\left(e_j\right)^2}$ г) $\dfrac{\sum_{j=2}^{n} x'_j\, y'_{j-1}}{\sqrt{\sum_{j=1}^{n}\left(x'_j\right)^2 \sum_{j=1}^{n}\left(y'_j\right)^2}}$ д) $\dfrac{\sum_{j=1}^{n} x'_j\, y'_j}{\sum_{j=1}^{n}\left(x'_j\right)^2}$

3. Производство некоторого вида животноводческой продукции относительно велико в третьем и четвертом кварталах и мало - в первом и втором кварталах каждого года. Общегодовой его уровень за 10 наблюдений (лет) хорошо описывается как множественная линейная регрессия от затрат ресурсов ($x_1,...,x_n$). Ресурсы в течение года тратятся относительно равномерно. Если ту же регрессию оценить по квартальным данным, по месячным данным и по полугодовым данным, о в каком из этих случаев можно ожидать значение статистики Дарбина-Уотсона DW длизким к нулю, в каком - близким к двум и в каком - к четырем? Колебания риводства внутри квартала и между кварталами с высоким (или с низким) уровнем производства при данных затратах ресурсов невилики и случайны.

а) поквартально $\Rightarrow$ DW близко к 2; помесячно - к 4; по полугодиям - к 0;

б) поквартально $\Rightarrow$ DW близко к 2; помесячно - к 0; по полугодиям - к 4;

в) поквартально $\Rightarrow$ DW близко к 0; помесячно - к 2; по полугодиям - к 4;

г) поквартально $\Rightarrow$ DW близко к 0; помесячно - к 4; по полугодиям - к 2;

д) поквартально $\Rightarrow$ DW близко к 4; помесячно - к 0; по полугодиям - к 2.

ЛИТЕРАТУРА

1. Башина О.Э. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: учебник для вузов; под ред. О.Э. Башиной, А.А. Спирина. - М.: Финансы и статистика, 2008.
2. Доугерти Кристофер. Введение в эконометрику. - М.: ИНФРА-М, 2009.
3. Ефимова М.Р. Практикум по общей теории статистики: учебное пособие для вузов / М.Р. Ефимова и др. – М.: Финансы и статистика, 2007.
4. Колемаев, В. А. Теория вероятностей и математическая статистика: учеб. для вузов / В. А. Колемаев, В. Н. Калинина; под ред. В. А. Колемаева. — М.: ИНФРА-М, 2005.
5. Кочетков Е.С., Смерчинская С.О., Соколов В.В.. Теория вероятностей и математическая статистика [Электронный ресурс]: Учебник - 2-е изд., испр. и перераб. - М.: Форум: НИЦ ИНФРА-М, 2014. (http://znanium.com/bookread2.php?book=447828)
6. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2010.
7. Мацкевич, И. П. Высшая математика: Теория вероятностей и математическая статистика: учеб. для студ. экон. вузов / И. П. Мацкевич, Г. П. Свирид. — Мн.: Выш. шк., 2003.
8. Новиков А.И. Эконометрика [Электронный ресурс]: Учебное пособие - 3-е изд., перераб. и доп. - М.: НИЦ ИНФРА-М, 2014. (http://znanium.com/bookread2.php?book=437118)
9. Салин В.Н. Курс теории статистики для подготовки специалистов финансово-экономического профиля: учебник / В.Н. Салин, Э.Ю. Чурилова. – М.: Финансы и статистика, 2007.
10. Фигурин, В. А. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В. А. Фигурин, В. В. Оболонкин. — Мн.: Новое знание, 2007.