Эконометрика

Введение.

Основные понятия

Чекменева Т.Д.

ЭКОНОМЕТРИКА

Эконометрика – устанавливает и исследует количественные закономерности в экономике на основе методов математической статистики.

Эконометрическая модель служит основой для экономического анализа и прогнозирования, дает возможность для принятия обоснованных экономических решений.

Литература

Сведения об учебниках			Количество экземпляров в библиотеке
Наименование, издательство	Автор	Год издания	
Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА	Кремер Н. Ш., Путко Б. А.	2003,311с	10
Эконометрика: Учебник. – М.: Финансы и статистика	Домбровский В. В.	2004,342с	35
Эконометрика: Учебник для вузов. – М.: Финансы и статистика	Елисеева И. И.	2003,342с. 2005,576с.	40 10
Практикум по эконометрике: Учеб. пособие. – М.: Финансы и статистика	Елисеева И. И.	2003	1
Эконометрические методы в макроэкономическом анализе	Замков О. О.	2001	10

В эконометрике исследуются три основных класса моделей:

1) модели временных рядов

$$y_t = f(t) + \varepsilon_t$$

2) регрессионные модели с одним уравнением

где $$y = f(x_1, \ldots\ldots, x_k, \beta_1, \ldots\ldots, \beta_p) + \varepsilon$$

y – зависимая (объясняемая) переменная,

$x_1, \ldots. x_k$ – независимые (объясняющие) переменные,

$\beta_1, \ldots.. \beta_p$ – параметры;

3) системы одновременных уравнений, где каждое уравнение может, кроме объясняющих переменных, включать в себя объясняемые переменные из других уравнений.

Переменные в регрессионной модели делятся на **два класса** в зависимости от того, являются они или нет объектом объяснения.

1. Эндогенные переменные (зависимые) определяется моделируемым явлением.
2. Экзогенные переменные (независимые) входят в модель, но определяются независимо от нее.

Эконометрические данные делят на два типа:

1. Перекрёстные (пространственные) данные (cross-sectional data) – это данные, полученные для разных однотипных объектов и относящиеся к одному периоду времени (например, данные о курсах валют в определенный день по всем обменным пунктам);
2. Временн**ы**е ряды (time series data) – это данные, характеризующие один и тот же объект в различные моменты времени.

Основные понятия теории вероятностей

- случайное событие (A)
- случайная величина (X)
- функция распределения $F(x)$
- числовые характеристики с/в
- случайный вектор ($X_1,\dots,X_n$)

Основные числовые характеристики случайной величины X

1. Математическое ожидание

для дискретной случайной величины:

$$E(x) = \sum_k x_k \cdot p_k$$

2. Дисперсия случайной величины

$$V(X) = E(X - EX)^2 = E\left(X^2\right) - (EX)^2$$

$\sigma = \sqrt{V(X)}$ – стандартное отклонение

3. Квантили и процентные точки распределения

Квантиль уровня q (q-квантиль) непрерывного распределения $F(x)$ – это число Uq (число q задано, $0 < q < 1$) такое, что

$$P\left(X \le U_q\right) = q$$

Q-процентная точка непрерывного распределения $F(x)$ – это число W_Q такое, что

$$P\left(X > W_Q\right) = \frac{Q}{100}$$

Между квантилем и процентной точкой имеется связь: $U_q = W_{100(1-q)}$

Числовые характеристики случайного вектора:

- Математическое ожидание

$$E(X)=\left(EX_1,\ldots,EX_n\right)^T$$

- Дисперсия случайного вектора – квадратная матрица ***V(X)***, составленная из коэффициентов ковариации компонент этого вектора

$$Cov\left(X_i,X_j\right)=E\left[\left(X_i-EX_i\right)\left(X_j-EX_j\right)\right]$$

Парный коэффициент корреляции

$$r_{XY} = \frac{\operatorname{cov}(X,Y)}{\sigma(x)\sigma(y)}$$

Если X, Y – независимы, то $r = 0$:

$$\operatorname{cov}(X,Y) = \mathrm{E}\Big(X - \underbrace{\mathrm{E}X}_{=0}\Big) \cdot \mathrm{E}\Big(Y - \underbrace{\mathrm{E}Y}_{=0}\Big) = 0$$

Основные понятия и задачи математической статистики

- генеральная совокупность;
- выборка (выборочная совокупность);
- оценивание параметров;
- проверка гипотез;
- специальные распределения:

χ^2- распределение, *t*- распределение Стьюдента, *F*- распределение Фишера

Нормальный закон распределения (Гаусса)

$$X \sim N(m, \sigma^2)$$

произвольная нормальная величина

$m = E(X),\ \sigma^2 = V(X)$

$$X_0 \sim N(0, 1)$$

стандартная нормальная величина

$$X_0 = \frac{X - m}{\sigma}$$

Статистические оценки делятся на *точечные* и *интервальные*

Свойства оценок (требования):

- несмещенность: $\mathrm{E}\hat{\theta} = \theta$
- состоятельность: оценка $\hat{\theta}$ гарантирует приближение к истинному значению при $n \to \infty$
- эффективность: $\mathrm{E}\left(\hat{\theta} - \theta\right)^2$ - минимальная из всех возможных оценок параметра θ

Точечные оценки (выборочные статистики)

- Оценка математического ожидания – среднее значение признака:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Оценка дисперсии признака:

- выборочная дисперсия:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

- исправленная дисперсия:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

Точечные оценки (выборочные статистики)

- Оценка коэффициента корреляции:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$$

$$|\boldsymbol{r}| \leq \mathbf{1}$$

$|r| > 0{,}7$ - связь сильная

$|r| < 0{,}3$ - связь слабая

Проверка статистических гипотез

$H_0 : \theta \in Z_0$ основная (нулевая) гипотеза

$H_1 : \theta \in Z_1$ альтернативная гипотеза

Задача: на основе наблюдений $X_1, ..., X_n$ проверить нулевую гипотезу, т.е. принять её либо отвергнуть в пользу альтернативной гипотезы

Процедура проверки гипотезы называется статистическим критерием (тестом).

Статистический критерий формулируется в виде неравенства:

$$t_n > C$$

t_n – некоторая функция от выборки:

$$t_n = t_n(X_1, \ldots, X_n)$$ - *проверочная статистика*

$C = t_{n,\alpha}$ - критическое значение (порог), определяющее критическую область (K_α)

.

Алгоритм проверки статистической гипотезы:

1. на основании наблюдений $X_1, ..., X_n$ вычисляется значение статистики t_n;
2. при заданном уровне значимости α находится критическая область K_α (т. е. её критическая точка $t_{n,\alpha}$);
3. если $t_n \in K_\alpha$ (т. е. $t_n > t_{n,\alpha}$), то гипотеза H_0 отвергается в пользу H_1; в противном случае принимается гипотеза H_0.

Ошибки проверки гипотез

1) *Ошибка 1-го рода*: отвергнуть H_0, когда она верна (ложная тревога).

Вероятность ошибки 1-го рода: $P(H_1 \mid H_0) = \alpha$

или $\alpha = P_{H_0}(X \in K_\alpha)$

α – уровень значимости

2) *Ошибка 2-го рода*: принять гипотезу H_0, когда верна гипотеза H_1.

Вероятность: $P(H_0 \mid H_1) = \beta$ или $\beta = P_{H_1}(X \in \overline{K}_\alpha)$

Односторонние и двусторонние критические области

1) *Односторонняя*: $\alpha = P\{t_n > t_{n,\alpha}\}$

2) *Двусторонняя*: $\alpha = P\{|t_n| > t^*_{n,\alpha}\}$

(для симметричных распределений)

Если распределение симметрично, то двусторонние критические точки связаны с односторонними критическими точками соотношением: $t^*_{n,\alpha} = t_{n,\alpha/2}$

p-value

Для рассмотренных распределений статистики t_n (нормальное, χ^2 , Стьюдента, Фишера) можно использовать еще одну критическую величину:

p-значение (значимость), p-value.

Если ***найдено значение*** статистики t_n, например, $t_n = C$, то p-значение вычисляется как вероятность:

- а) $P(|t_n| > |C|)$ – для симметричных распределений;
- б) $P(t_n > C)$ – для несимметричных распределений.

Очевидно, если p-значение $< \alpha$, то $t_n > t_{n,\alpha}$.