**Эконометрика. Основные понятия**

Это наука, которая занимается изучением и математическим описанием взаимосвязей экономических показателей, а так же прогнозированием возможных значений экономических показателей в зависимости от тех факторов, которые воздействуют на экономические процессы. Эконометрика опирается на математическую экономику, которая рассматривает теорию вероятностей и статистику. Задача эконометрики состоит в разработке математических моделей принятия решений в условиях неопределенности, численной оценки параметров этой модели и получение прогнозов и предсказаний на основе проведенного анализа. В эконометрике принято разделять экономические показатели на результирующие (зависимые) и факторные (независимые):

* У – зависимая переменная;
* Х – независимая переменная.

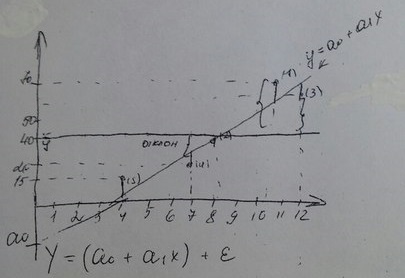
У = f(х) + Е, где

Е - случайная компонента, которая определяется мерой незнания об объекте.

**Рассмотрим пример**. Имеются два показателя экономических:

* Капитальные затраты на ремонт однотипного оборудования;
* Срок службы оборудования. i – номер наблюдения.

|  |  |  |
| --- | --- | --- |
| **i** | **Затраты (у)** | **Срок службы (х)** |
| 1 | 70 | 10 |
| 2 | 40 | 8 |
| 3 | 60 | 12 |
| 4 | 20 | 7 |
| 5 | 15 | 4 |

****

**Выборочный коэффициент корреляции**

* **Kyx –** выборочный корреляционный момент;
* **SxSy** – выборочное стандартное отклонение по х.

– выборочная дисперсия СВ Х

- выборочная дисперсия СВ У (общая дисперсия СВ У)

Посчитать дисперсию, стандартное отклонение, коэффициент корреляции:

**Дисперсия**

=1, когда зависимость не случайная, а функциональная.

отклонение

Т.е. мы находим численные значения коэффициентов и .

уравнение простой линейной регрессии.

Задавая *x* мы можем рассчитывать *y* для того, чтобы вычислить возможные последствия принятия решений.

**Оценка качества подгонки уравнения регрессии к наблюдаемым данным.**

Выборочная дисперсия характеризует разброс значений случайной величины.

Доля отклонений\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ объясняемое уравнением регрессии и доля отклонений случайной величины y от средней величины объясняемое случайными факторами в сумме равняются 1.

Чем больше значение 1го слагаемого, тем меньше значение 2го.

Коэффициент детерминации:

Значение индекса детерминации показывает, какая доля отклонений случайной величины y от среднего значения объясняется уравнением регрессии.

Чем выше индекс детерминации, тем выше качество подгонки.

показывает, какая доля отклонения y от среднего значения объясняется случайными факторами.

С признаком детерминации связаны показатели, которые определяют тесноту связи между результирующим \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Множественный коэффициент корреляции:

Для регрессионной модели с одной факторной переменной множ-ый коэф-ент корреляции равняется выборочному коэф-ту корреляции.

Свойства множественного коэф-нта корреляции:

* Если между результирующей и факторной переменной чистая линейная зависимость, коэф-т корреляции равен 1, если прямая и -1, если обратная.

**Проверка статистической значимости индекса детерминации и множества коэффициентов корреляции.**

Для проверки статистической значимости индекса детерминации используют F-критерий Фишера

Расчетное значение критерия Фишера сравнивается с критическим значением.

Критическое значение:

где α – уровень значимости – вероятность совершения ошибки первого уровня

dt1,dt2 – число степеней свободы

Если , то мы отвергаем проверяемую гипотезу.

**Проверка статистической значимости коэффициентов уравнения регрессии.**

- выборочные параметры

Для проверки используется критерий Стьюдента.

Проверяется гипотеза

Двусторонняя гипотеза

Рассчитывается t-критерий Стьюдента:

H0 отклоняется – статистически значимое

H0 принимается

– уровень критерия Стьюдента

Значение p-уровня сравниваем с

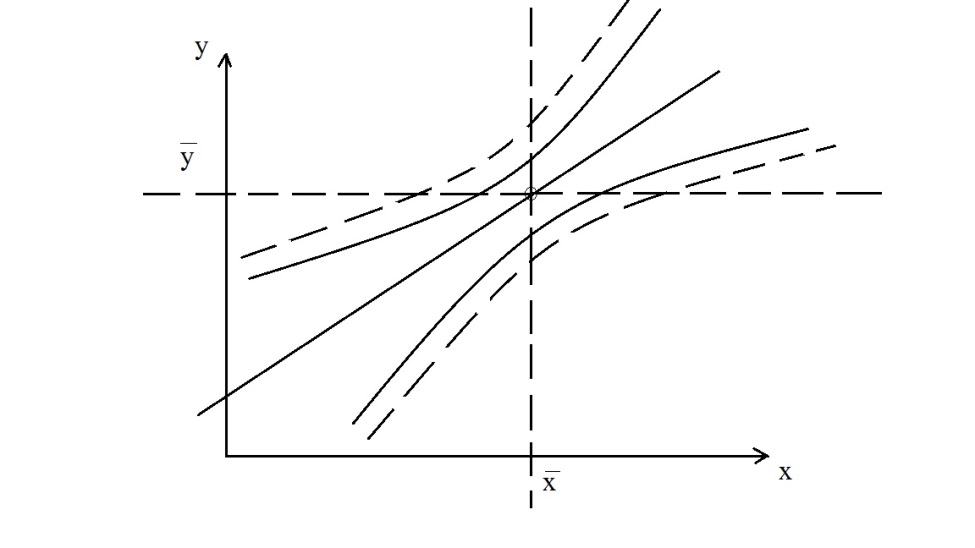
нулевую гипотезу отклоняем полученное выборочное значение статистически значимое

Стандартная ошибка уровня регрессии

**Прогнозирование на основе уровня простой линейной регрессии.**

Полученное прогнозное значение – случайная величина, поэтому необходимо определить доверительный интервал, в котором с заданной вероятностью будет находиться истинное прогнозное значение.

Первый интервал для среднего значения прогнозной величины:



– квантиль распределения Стьюдента уровня

– стандартная ошибка коэффициента

Условия Гаусса-Маркова

1. Математическое ожидание
2. Дисперсия ;
3. Значения случайной компоненты независимы в двух любых наблюдениях
4. Случайная компонента распределена нормально с параметрами 0 и 1 для

При выполнении условий гарантированно, что полученные оценки уравнений регрессии будут несмещёнными, состоятельными и эффективными.

Ѳ

1. Понятие несмещённости M [] = Ѳ
2. Если дисперсия является минимальной из всех возможных оценок параметра. Для проверки положений используются остатки, которые вычисляются: *e*i – оценка случайной компоненты .

**Множественная линейная регрессия.**

Модель описывает линейную зависимость между результирующей переменной «y» и несколькими переменными «m».

Для нахождения коэффициентов используется метод наименьших квадратов.

n – объём выборки

Оценки, получаемые таким образом – несмещённые, эффективные и самостоятельные, если выполняются условия Гаусса-Маркова. Для оценки качества полученной модели используется множественный коэффициент корреляции, который показывает наличие линейной зависимости между результирующей и факторными переменными. Либо отсутствие таких зависимостей «R» равное или близкое к 0. Индекс детерминации показывает долю общих отклонений «y» от её среднего значения, которое объясняется уравнением регрессии:

Множественный коэффициент корреляции всегда положительный

Проверка статистической значимости осуществляется с помощью критерия Фишера(F).

MS – сумма квадратов отклонений на одну степень свободы

Квантиль p(X=xа0) =a0

Оценка статистич значимости коэф а0, а1,… и тд производ так же как и для простой лин рег., расч значение t критерия Стьюдента сравнивается с критическим, которое вычисляется:

t1-α/2, df < |t1| H0 aj=0 то отвергаем

df2=n-m-1

PF, P0< α статистич значимы

Вычисл доверительные интервалы т.е. такие интервалы, которые с вероятностью 1- α будет находиться истинное значение коэффициента.

df=n-m-1

mj – станд ошибка коэф aj

Доверит интервал для результирующей

dj=(xj-

a=)

Условие Гаусса-Маркова

Как и для простой лин. регрессии необх условие Гаусса-Маркова для того, чтобы полученные к оценки были не смещ, эффект., состоятельными

Условия:

1. M(εi)=0, ∀i
2. D(εi)=σ2, ∀i
3. cov (εi, εj)=0, ∀i,j i≠j
4. cov (xj, εi)=0, ∀I, j=1,…,m

5 N(0;1)

Исп множ коэф корреляции и индекс детерминации(R2)

Для проверки статист значимости коррел и ин. Детерм исп F-критерий Фишера.

Fp

MSрег- сумма квадратов отклонений приход на 1 степень свободы

MSрег= , df1=m MSост= , df2=n-m-1

Спецификация регрессионной модели

В процессе исследования и построения экономич моделей необх. правильно специфицировать(выбрать) факторные переменные, которые необх. вкл. В модель.

Если чисто факторных переменных меньше, чем это необходимо для правильного описания экон. процесса,, то оценки, полученные с помощью метода наим. Квадратов оказываются смещенными и полученные значения t статистики явл завышенными, а значения показат. статист-ки не значимыми. Если вкл. перем. Сверх необх числа, то оценки будут не смещенными, однако они могут быть статистически не значимыми.

В общих случаях становится трудно интерпретировать эк смысл модели.

Для правильной спецификации модели необх проводить доп исследования, направленные на вкл только тех факторных перем, которые существенны для данной модели.

Вкл:

1. Исправленное значение коэф. детерминации 2

При введении в модель доп переменной, коэф детерминации автоматически увеличивается

m m+1

R2m < R2m+1

SSmост > SSm+1ост

2 скорректир индекс детерминации

R2

m m-1

R2m > R2m-1

SSmост< SSm-1ост

Другим эл. Для корректности спецификации служит стандартизированная форма уравнения

множественной регрессии - каноническая форма уравнения мн. лин. регрессии

ty=β1tx+…+ βmtm -стандартная

ty-стандартизированное значение результирующей переменной

tx1…txm – стандартиз значение факторной переменной

-нормализация

: переменные, кот выражаются в натур ед приводятся к безразмерному виду

β-коэф

В отличие от коэф модели чистой регрессии β-коэф позволяют упорядочить факторные перем по степени их влияния на результирующий показатель.

Чем больше по абсолютной величине значение β показателя тем сильнее фактор влияет на результирующую переменную. Значение β коэф можно найти, решив систему лин уравнений след вида:

ry1x=β1+β2rx1x2+…+βmrx1xm

…………………………………………

коэффициент парной корреляции значений факторных переменных xj и xi.

По значениям парной корреляции можно построить матрицу. Анализируя матрицу парной корреляции, можно определить значение наиболее сильно связанные друг с другом. Эти значения могут иск. из модели

x1 x2 xm

x1 1 rx1x2  rx1xm

x2 rx2x1 1 rx2xm

x rxm x1 rxm x2 1

Взаимосвязь факторных переменных называется – коллинеарностью – высокая связь.

Мульти – коллинеарность – связь не только между парами, но и с другими переменными. Приводит к тому, что существует риск того, что обратная матрица (XTX)-1 становится вырожденной. Полученные оценки будут смещенными. Нужно определить мульти – к-ть. Нужно определить мульти – к-ть.

Метод выявления мульти – к-ти.

17

= def

= def

Если близко к 0, то коллинеарность

Если близко к 1, то отсутствует мульти – коллинеарность.

Лекция 02.04.2015

Устранение мультиколлинеарности

1. Ридж – регрессия (гребневая)

2. Метод пошаговой регрессии

Использование частного F критерия Фишера и частичных критериев корреляции.

Частный F- критерий Фишера показывает степень влияния факторной переменной xj на результирующую переменную y при фиксированном значении всех остальных переменных.

SSост =

x1, x2, … , xm

SSрег(xj/c) = SSрег(x1, … , xm) – SSрег(x1, x2, xj-1, xj+1 … xm)

С помощью критерия Фишера проводится проверка следующей нулевой гипотезы:

H0: включение j-той факторной переменной не влияет на точность полученной модели

H1: включение j-той факторной переменной увеличивает точность регрессионной модели.

Для проверки нулевой гипотезы, расчетное значение частного F-критерия сравнивается с критическим значением. Критическое значение F- критерия есть квантиль распределения Фишера уровня α с числом степени свободы df1 и df2

df1 =1 df2 = n-m-1

n число наблюдений

m – число факторных переменных

– нулевую гипотезу отклоняем на уровне значимости α.

Считаем, что xj существенно влияет на точность регрессионной модели.

PF < x

Частный коэффициент корреляции

Показывает степень влияния факторной переменной xj на результирующую переменную y при фиксированном значении остальных переменных

Между частным F- критерием и частным коэффициентов корреляции существует связь.

Метод пошаговой регрессии

Это формальная процедура отбора факторных переменных для построения оптимальной регрессионной модели.

Шаг 1. Рассчитывается значение коэффициента корреляции ryxj

Fyxj > Fкр

Шаг 2. Последовательно выбирается факторная переменная и рассчитываются частные критерии Фишера для добавленной переменной.

Fyxj/c

c = xj2 – набор факторных переменных, которые проверяет на шаге 2.

Вкл. переменную отбираем из условия max значения частного F – критерия.

Следующие шаги заключаются аналогично шагу 2, доб. другие факторные переменные расч. частные F - критерии и отбираются факторные переменные с наибольшим значением частного F – критерия. Процедура продолжается до тех пор пока значение F – критерия не достигнет определенного уровняли или пока не вкл. все переменные.

Критерий Окайко

AIC =

Sост – остаточное стандартное отклонение

**Фиктивные переменные**

Используются для того, чтобы наряду с количественными факторными переменными включать в модель и качественные.

Когда объясняющие переменные носят качественный характер. Для того чтобы построить модель необходимо качественные перевести в количественный вид, это достигается путем введения фиктивных переменных. Фиктивные переменные, как правило, являются бинарными, то есть они принимают значение 0 и 1. Если одно значение признака принимает значение 0, то противоположное принимает 1.

**Пример:**

y – вес новорожденного

|  |  |  |
| --- | --- | --- |
| y | x | z |
|  | … | 0  0  1 |

y

(1)

(2)

x

Число фиктивных переменных всегда меньше градации качественного признака на 1. В противном случае определитель обратной матрицы становится равным 0. Можно вводить фиктивные переменные для того, чтобы изменять коэффициент при переменных.

**Гетероскедастичность. Выявление и устранение.**

Дисперсия случайной компоненты является случайной величиной:

Если это условие не выполняется, то есть дисперсия случайной компоненты изменяется от наблюдения к наблюдению, то это называется гетероскедастичность.

Наличие гетероскедастичности приводит к тому, что оценки, получаемые с помощью метода наименьших квадратов становятся неэффективными, то есть они не обладают минимальной дисперсией

Методы выявления гетероскедастичности:

1. Визуальный. Используется диаграмма рассеивания (график)

ε

ε

x

x

ε

ε

x

x

**Критерий Чоу**

Имеются 2 выборки значений

|  |  |  |  |
| --- | --- | --- | --- |
| .  . | .  . | .  . | .  . |

b является t-криткрием Стьюдента

Рассчитаем F-статистику

Расчетное значение сравниваем с критическим (квантиль распределения Фишера)

Если расчетное значение больше, то это две разные выборки.

**Гетероскедастичность**

***Методы:***

1. *Тест Спирмена*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| *…* | *…* | *…* | *…* | *…* |

Коэффициент ранговой корреляции Спирмена

Полученное значение коэффициента сравниваем с критическим значением

1. Тест Голфелда-Квандта

Наблюдаемые данные упорядочиваются по значению факторного признака и делятся на 3 части.

Первые значений данных и последние . По первым вычисляется коэффициент регрессии, и находятся .

*Рассчитанное значение F статистики сравнивается с критическим. Критическое значение – это квантиль распределения Фишера.*

1. *Тест Глейзера*

*Строится предположение, что стандартное отклонение случайной компоненты зависит от значения факторной переменной.*

– стандартное отклонение случайной компоненты в i-ом наблюдении

|  |  |
| --- | --- |
| j |  |
| -1  -0.5  0.5  1  1.5 |  |

Метод наименьших квадратов определяет пары соответствующие факторной переменной и значение остатков. Дальше строим несколько моделей.

Регрессионная модель с присутствием гетероскедостичности. Стандартные отклонения случайной компоненты в каждом наблюдении будут отличаться друг от друга. Для устранения гетероскедостичности поступают следующим образом:

Исходное уравнение делим на случайное отклонение в i-ом наблюдении

; - новая переменная

А значения оценок получаем используя тест Глейзера.