

# Эконометрика. Лекция 2

3 декабря 2014 г.

# Статистические свойства оценок коэффициентов

- сформулируем стандартные предпосылки
- строить доверительные интервалы для коэффициентов
- проверять гипотезы о коэффициентах

# Условное математическое ожидание

$r$  — одна случайная величина

$s$  — одна случайная величина

## Условное математическое ожидание. Неформально

$E(s|r)$  — это такая функция от случайной величины  $r$ , которая наиболее похожа на случайную величину  $s$

## Условное математическое ожидание. Формально

$E(s|r)$  — это случайная величина  $\tilde{s}$ :

- ① представимая в виде  $\tilde{s} = f(r)$
- ②  $E(\tilde{s}) = E(s)$
- ③  $Cov(s - \tilde{s}, g(r)) = 0$  для любой  $g(r)$ .

Или:  $Cov(s, g(r)) = Cov(\tilde{s}, g(r))$

Теорема: Если величина  $r$  дискретна и принимает значения  $a$ ,  $b$  или  $c$ , то

$$E(s|r) = \begin{cases} E(s|r = a), & \text{если } r = a \\ E(s|r = b), & \text{если } r = b \\ E(s|r = c), & \text{если } r = c \end{cases}$$

# Пример расчета чудо-доска

# Если величины непрерывны и есть совместная функция плотности

Теорема: Если пара величин  $x, y$  имеет функцию плотности  $f(r, s)$ , то

$$E(s|r) = \int_{-\infty}^{\infty} s \cdot f(s|r) dx$$

где  $f(s|r) = f(r, s)/f(r)$  — условная функция плотности



# Свойства условного ожидания

Пусть  $a, b$  — константы,  $s, r$  — случайные величины.

Идея: свойства  $E(s|r)$  аналогичны свойствам  $E(s)$ , если считать  $r$  и любую функцию  $h(r)$  константой.

# Список свойств

- $E(E(s|r)) = E(s)$
- $E(as + b|r) = aE(s|r) + b$
- $E(h(r)|r) = h(r)$
- $E(h(r)s|r) = h(r)E(s|r)$

# Условная дисперсия и ковариация

Обычная дисперсия:  $Var(s) = E(s^2) - (E(s))^2$

Условная дисперсия.  $Var(s|r) = E(s^2|r) - (E(s|r))^2$

Обычная ковариация:  $Cov(s_1, s_2) = E(s_1 s_2) - E(s_1)E(s_2)$

Условная ковариация:  $Cov(s_1, s_2|r) = E(s_1 s_2|r) - E(s_1|r)E(s_2|r)$

## чудо-доска. пример расчета условной дисперсии

# Свойства условной дисперсии

Пусть  $a, b$  — константы,  $s, r$  — случайные величины.

Идея: свойства  $Var(s|r)$  аналогичны свойствам  $Var(s)$ , если считать  $r$  и любую функцию  $h(r)$  константой.

$$\text{Var}(as + b|r) = a^2 \text{Var}(s|r)$$

$$\text{Var}(s + h(r)|r) = \text{Var}(s|r)$$

$$\text{Var}(h(r)s|r) = h^2(r) \text{Var}(s|r)$$

$$\text{Var}(s) = \text{Var}(E(s|r)) + E(\text{Var}(s|r))$$

# Геометрическая интерпретация. Чудо-доска.

# Мораль геометрической интерпретации:

Если считать, что  $Cov(r, s)$  — скалярное произведение, то

- квадрат длины случайной величины  $r$ ,  $Var(r)$
- косинус угла между случайными величинами,  $Corr(s, r)$

Верны “школьные” теоремы: теорема Пифагора, Фалеса, etc



# Предпосылки на ошибки

- $E(\varepsilon_i|X) = 0$
- $E(\varepsilon_i^2|X) = \sigma^2$  или  $Var(\varepsilon_i|X) = \sigma^2$
- $E(\varepsilon_i\varepsilon_j|X) = 0$  или  $Cov(\varepsilon_i, \varepsilon_j|X) = 0$

## ковариационная матриц

Ковариационная матрица вектора  $\varepsilon$ :

$$Var(\varepsilon) = \begin{pmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1, \varepsilon_2) & Cov(\varepsilon_1, \varepsilon_3) & \dots \\ Cov(\varepsilon_2, \varepsilon_1) & Var(\varepsilon_2) & Cov(\varepsilon_2, \varepsilon_3) & \dots \\ Cov(\varepsilon_3, \varepsilon_1) & Cov(\varepsilon_3, \varepsilon_2) & Var(\varepsilon_3) & \dots \\ \vdots & & & \end{pmatrix}$$

## запись предпосылок с помощью ковариационной матрицы

$$\text{Var}(\varepsilon|X) = \begin{pmatrix} \sigma^2 & 0 & 0 & \dots \\ 0 & \sigma^2 & 0 & \dots \\ 0 & 0 & \sigma^2 & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \sigma^2 \cdot I_{n \times n}$$

# Дисперсия и ковариация оценок коэффициентов

Предпосылки:

- $Var(\varepsilon|X) = \sigma^2 \cdot I_{n \times n}$
- $Var(\varepsilon_i|X) = \sigma^2$
- $Cov(\varepsilon_i, \varepsilon_j|X) = 0$
- $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$

Позволяют посчитать  $Var(\hat{\beta}_j)$ ,  $Cov(\hat{\beta}_j, \hat{\beta}_l)$

# Пример вычислений в парной регрессии (в регрессии на константу)

чудо-доска.

- Найдите  $Var(\hat{\beta}_2|X)$ ,  $Cov(\hat{\beta}_1, \hat{\beta}_2|X)$ ,  $Var(\hat{\beta}_1|X)$

## Итого в парной регрессии:

- $Var(\hat{\beta}_2|X) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$
- $Cov(\hat{\beta}_1, \hat{\beta}_2|X) = \frac{-\bar{x}\sigma^2}{\sum (x_i - \bar{x})^2}$
- $Var(\hat{\beta}_1|X) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$

## Вопрос:

- Зачем придумали эту условную дисперсию, если все свойства аналогичны обычной дисперсии?
- А вот как раз и придумали, чтобы аналогичны всё было :)

## Теорема (без доказательства):

$$\text{Var}(\hat{\beta}_j|X) = \sigma^2 / \text{RSS}_j$$

$\text{RSS}_j$  — сумма квадратов остатков в регрессии  $j$ -ой объясняющей переменной на остальные объясняющие переменные



# ЛИНАЛ: прелюдия к доказательству

ЛИНАЛ:  $Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

Свойство:  $Var(Ay) = AVar(y)A'$

Напомним, что  $(AB)' = B'A'$  и  $(A^{-1})' = (A')^{-1}$  поэтому:

- $(X'X)' = X'X'' = X'X$
- $((X'X)^{-1})' = (X'X)^{-1}$

## ЛИНАЛ: чудо-доска

Если оценки МНК существуют и единственны,  $\text{Var}(\varepsilon|X) = \sigma^2 I_{n \times n}$   
то  $\text{Var}(\hat{\beta}|X) = \sigma^2 (X'X)^{-1}$

Доказательство

# Как оценить $\sigma^2$ ?

Константа  $\sigma^2$  неизвестна.

Случайная величина  $\hat{\sigma}^2 = \frac{RSS}{n-k}$  — замечательная оценка для  $\sigma^2$ .

# Оценка ковариационной матрицы

- $Var(\hat{\beta}_j|X) = \sigma^2 \cdot f(X)$
- $\widehat{Var}(\hat{\beta}_j|X) = \hat{\sigma}^2 \cdot f(X)$

а именно:  $\widehat{Var}(\hat{\beta}_j|X) = \hat{\sigma}^2 / RSS_j$

- $se(\hat{\beta}_j) = \sqrt{(\hat{\beta}_j|X)}$

Например, в модели  $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ :  $se(\hat{\beta}_2) = \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}$

## оценка ковариационной матрицы

$$\widehat{Var}(\hat{\beta}|X) = \begin{pmatrix} \widehat{Var}(\hat{\beta}_1|X) & \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2|X) & \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3|X) & \dots \\ \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_1|X) & \widehat{Var}(\hat{\beta}_2|X) & \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3|X) & \dots \\ \widehat{Cov}(\hat{\beta}_3, \hat{\beta}_1|X) & \widehat{Cov}(\hat{\beta}_3, \hat{\beta}_2|X) & \widehat{Var}(\hat{\beta}_3|X) & \dots \\ \vdots & & & \end{pmatrix}$$

- ЛИНАЛ:  $\widehat{Var}(\hat{\beta}|X) = \hat{\sigma}^2 \cdot (X'X)^{-1}$
- В R: `vcov(model)`

# БСХС - Большой Список Хороших Свойств

- Базовые:

верны даже на малых выборках без предположения о нормальности  $\varepsilon_i$

- Асимптотические:

верны на больших выборках даже без предположения о нормальности  $\varepsilon_i$

- При нормальности:

верны при нормальности  $\varepsilon_i$  даже на малых выборках

# БСХС — предпосылки

Если:

- 1 Истинная зависимость имеет вид  $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ 
  - В матричном виде:  $y = X\beta + \varepsilon$
- 2 С помощью МНК оценивается регрессия  $y$  на константу,  $x_i$ ,  $z_i$ 
  - В матричном виде:  $\hat{\beta} = (X'X)^{-1}X'y$
- 3 Наблюдений больше, чем оцениваемых коэффициентов  $\beta$ :  $n > k$

## БСХС — предположения на $\varepsilon_i$ :

- ❶ Строгая экзогенность:  $E(\varepsilon_i | \text{все регрессоры}) = 0$ 
  - В матричном виде:  $E(\varepsilon_i | X) = 0$
- ❷ Условная гомоскедастичность:  $E(\varepsilon_i^2 | \text{все регрессоры}) = \sigma^2$ 
  - В матричном виде:  $E(\varepsilon_i^2 | X) = \sigma^2$
- ❸  $Cov(\varepsilon_i, \varepsilon_j | X) = 0$  при  $i \neq j$



# БСХС — предпосылки на регрессоры

- 7 векторы отдельных наблюдений  $(x_i, z_i, y_i)$  — независимы и одинаково распределены
- 8 с вероятностью 1 среди регрессоров нет линейно зависимых
- Синонимы в матричном виде:  $\text{rank}(X) = k$  или  $\det(X'X) \neq 0$  или  $(X'X)^{-1}$  существует

## БСХС — базовые свойства (т. Гаусса-Маркова)

- Оценки  $\hat{\beta}_j$  линейны по  $y_i$ :  $\hat{\beta}_j = c_1 y_1 + \dots + c_n y_n$
- Оценки несмещены:  $E(\hat{\beta}_j|X) = \beta_j$ , и в частности  $E(\hat{\beta}_j) = \beta_j$

# БСХС — базовые свойства (т. Гаусса-Маркова)

- Оценки эффективны среди линейных и несмещенных

Для любой линейной по  $y_i$  и несмещенной альтернативной оценки  $\hat{\beta}^{alt}$ :

$$Var(\hat{\beta}_j^{alt}|X) \geq Var(\hat{\beta}_j|X) \text{ и } Var(\hat{\beta}_j^{alt}) \geq Var(\hat{\beta}_j)$$

- Ковариационная матрица:  $\text{Var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

Дисперсии:  $\text{Var}(\hat{\beta}_j|X) = \sigma^2 / \text{RSS}_j$

- $\text{Cov}(\hat{\beta}_j, \hat{\varepsilon}_i|X) = 0$
- $E(\hat{\sigma}^2|X) = \sigma^2$ , и  $E(\hat{\sigma}^2) = \sigma^2$

# БСХС — асимптотические свойства

При  $n \rightarrow \infty$ :

- $\hat{\beta}_j \rightarrow \beta_j$  по вероятности
- $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \rightarrow N(0, 1)$  по распределению
- $\hat{\sigma}^2 \rightarrow \sigma^2$  по вероятности

$$\hat{\sigma}^2 = \frac{RSS}{n-k}$$

# БСХС — при нормальности

Если дополнительно известно, что  $\varepsilon_i \sim N(0, \sigma^2)$ :

- Оценки эффективны среди несмещенных
- $\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} | X \sim t_{n-k}, \frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \sim t_{n-k}$
- $RSS/\sigma^2 | X \sim \chi_{n-k}^2, RSS/\sigma^2 \sim \chi_{n-k}^2$

# Доверительные интервалы для коэффициентов

Возможно строить в двух подходах:

- Асимптотически:  $t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \rightarrow N(0, 1)$
- При нормальности:  $t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k}$

Примерный 95%-ый интервал:

$$[\hat{\beta}_j - 2se(\hat{\beta}_j); \hat{\beta}_j + 2se(\hat{\beta}_j)]$$

## Описание любого теста:

- предпосылки теста (например, асимптотический или точный)
- проверяемая  $H_0$  против  $H_a$
- формула для вычисления статистики
- закон распределения статистики при верной  $H_0$



# Последовательность действий

- 1 выбираем уровень значимости  $\alpha$ ,  
 $\alpha = P(H_0 \text{ отвергнута} \mid H_0 \text{ верна})$
- 2 находим наблюдаемое значение некоторой статистики
- 3 находим критическое значение статистики (можно посчитать Р-значение)
- 4 сравниваем критическое и наблюдаемое (можно сравнить Р-значение и  $\alpha$ )
- 5 вывод: “ $H_0$  отвергается” или “ $H_0$  не отвергается”

# Чудо-доска

Дано:

- уравнение регрессии
- ковариационная матрица
- оценка  $\hat{\sigma}^2$

```
(Intercept) 59.86392 3.98754 15.013 <2e-16 ** Agriculture 0.10953  
0.07848 1.396 0.1698  
Catholic 0.11496 0.04274 2.690 0.0101
```

```
(Intercept) Agriculture Catholic
```

```
(Intercept) 15.900471817 -0.256680712 -0.006998292 Agriculture  
-0.256680712 0.006159437 -0.001345371 Catholic -0.006998292  
-0.001345371 0.001826622
```

Residual standard error: 11.07

Надо:

стандартные ошибки часто выписывают под коэффициентами

$$\widehat{Fertility}_i = \underset{(3.98)}{59.8} + \underset{(0.078)}{0.109} Agriculture_i + \underset{(0.042)}{0.115} Catholic_i$$

## стандартная табличка в любом пакете + чудо-доска

Вывели на экран, рядом рассказали на примере одного коэффициента с графиком!

Estimate Std. Error t value Pr(>|t|)

(Intercept) 59.86392 3.98754 15.013 <2e-16 \*\*\*

Agriculture 0.10953 0.07848 1.396 0.1698

Catholic 0.11496 0.04274 2.690 0.0101 \*

Проверка значимости — на самом деле проверка незначимости:

- “Мы проверили значимость коэффициента при доходе”
- Мы проверили  $H_0 : \beta_{inc} = 0$ .

# $H_0$ не отвергается

- недостаточно данных чтобы отвергнуть  $H_0$
- имеющиеся данные не противоречат  $H_0$

(много еще чему не противоречат)

# Значимость и существенность

- Коэффициент может быть значимым и совершенно несущественным

На огромных выборках — все коэффициенты значимы

- Коэффициент может быть существенным но не значимым

# Стандартизированные коэффициенты

Существенность — можно придать разный математический смысл

Например:

- стандартизировать переменные:

$$y_i^{st} := \frac{y_i - \bar{y}}{sd(y)}, \quad x_i^{st} := \frac{x_i - \bar{x}}{sd(x)}, \quad z_i^{st} := \frac{z_i - \bar{z}}{sd(z)}$$

- переоценить модель:

$$y_i^{st} = \beta_1^{st} + \beta_2^{st} x_i^{st} + \beta_3^{st} z_i^{st} + \varepsilon_i^{st}$$



# Проблема множественных сравнений

- Исследователь хочет проверить гипотезу о том, что  $\beta_{42} = 0$ . Ok.
- Исследователь хочет выяснить какие регрессоры из 100 значимы. Нужна поправка.

# Проверка гипотезы об одном ограничении

Хотим проверить гипотезу о  $\beta_2 - \beta_3$ .

Статистика  $t = \frac{\hat{\beta}_2 - \hat{\beta}_3 - (\beta_2 - \beta_3)}{se(\hat{\beta}_2 - \hat{\beta}_3)}$  распределена

- асимптотически  $N(0, 1)$
- при нормальности  $t_{n-k}$

# Переформулировка модели

Хотим проверить гипотезу о  $\beta_2 - \beta_3$ .

Всегда можно переформулировать модель так, что  $\beta_2 - \beta_3$  станет новым коэффициентом  $\beta'_2 = \beta_2 - \beta_3$ .

# Пример у чудо-доски

- способ через ковариационную матрицу
- способ через переформулировку модели

# Мораль:

В этой лекции мы научились:

- строить доверительные интервалы
- проверять гипотезы об отдельном коэффициенте
- сформулировали все стандартные предпосылки

В следующей:

- более сложные гипотезы
- прогнозирование