Статистические свойства оценок

Эконометрика. Лекция 2

Статистические свойства оценок коэффициентов

- стандартные предпосылки для модели линейной регрессии
- доверительные интервалы для коэффициентов
- гипотезы о коэффициентах

Условное математическое ожидание

r — одна случайная величина

s — одна случайная величина

$E(s|r)$ — 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: $Cov(s, g(r)) = Cov(\tilde{s}, g(r))$

На практике

Теорема:

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

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

Задача [у доски]

s, r	$r = 1$	$r = 2$
$s = 0$	0.25	0.2
$s = 10$	0.25	0.3

Найдите: $E(s|r)$, $E(s^2|r)$

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.

Свойства условного ожидания

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

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

Задача [у доски]

s, r	$r=1$	$r=2$
$s=0$	0.25	0.2
$s=10$	0.25	0.3

Найдите: $Var(s|r)$

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.

Свойства условной дисперсии

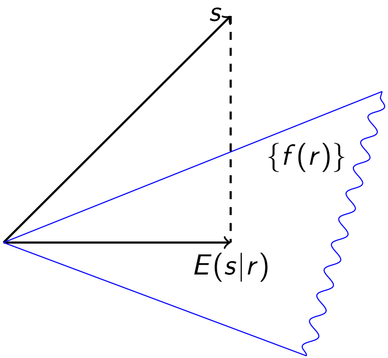
$$Var(as + b|r) = a^2 Var(s|r)$$

$$Var(s + h(r)|r) = Var(s|r)$$

$$Var(h(r)s|r) = h^2(r) Var(s|r)$$

$$Var(s) = Var(E(s|r)) + E(Var(s|r))$$

Геометрическая интерпретация [у доски]



Мораль геометрической интерпретации:

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

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

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 матрица

Ковариационная матрица вектора ε:

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

Запись предпосылок с помощью ковариационной матрицы

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

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 $Var(\hat{\beta}_j|X)$, $Cov(\hat{\beta}_j, \hat{\beta}_l|X)$

Пример вычислений в парной регрессии [у доски]

В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

Предположим, что: $Var(\varepsilon_i|X) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j|X) = 0$

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 всё аналогично просто считалось! Настоящая безусловная дисперсия оценок коэффициентов — гораздо сложнее, чем условная.

Теорема (без доказательства):

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

RSS_j — сумма квадратов остатков в регрессии j-ой объясняющей переменной на остальные объясняющие переменные (включая константу)

ЛИНАЛ. Ковариационная матрица оценок коэффициентов

Средствами линейной алгебры можно доказать, что:

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

ЛИНАЛ. Предварительная информация к доказательству:

Свойство: $Var(Ay) = A \cdot Var(y) \cdot A'$

Это матричный аналог свойства $Var(a \cdot y_1) = a^2 \cdot Var(y_1)$.

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

Поэтому:

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

ЛИНАЛ. доказательство формулы [у доски]

Если оценки МНК существуют и единственны, $Var(\varepsilon|X) = \sigma^2 I_{n\times n}$

то ковариационная матрица равна:

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

Как оценить σ^2?

Константа σ^2 неизвестна.

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

Замечательная в смыслах:

- $E(\hat{\sigma}^2) = \sigma^2$, в среднем оценивает верно
- $\hat{\sigma}^2 \to \sigma^2$ по вероятности с ростом n

Оценка ковариационной матрицы

Идея: заменим во всех формулах σ^2 на $\hat{\sigma}^2$:

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

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

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

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

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

- 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 ε_i

- Асимптотические:

верны на больших выборках даже без предположения о нормальности ε_i

- При нормальности:

верны при нормальности ε_i даже на малых выборках

БСХС — предположение о связи y и регрессоров

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 β: $n > k$

БСХС — предположения на ε_i:

Если:

4. Строгая экзогенность: $E(\varepsilon_i|$ все регрессоры $) = 0$
   - В матричном виде: $E(\varepsilon_i|X) = 0$
5. Условная гомоскедастичность: $E(\varepsilon_i^2|$ все регрессоры $) = \sigma^2$
   - В матричном виде: $E(\varepsilon_i^2|X) = \sigma^2$
6. $Cov(\varepsilon_i, \varepsilon_j|X) = 0$ при $i \neq j$

БСХС — предположения на регрессоры

Если:

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: $rank(X) = k$ или $det(X'X) \neq 0$ или $(X'X)^{-1}$ существует

БСХС — базовые свойства (т. Гаусса-Маркова)

То:

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

То:

- Оценки эффективны среди линейных и несмещенных

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

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

БСХС — базовые свойства

То:

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

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

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

БСХС — асимптотические свойства

То при $n \to \infty$:

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

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

БСХС — при нормальности

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

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

Доверительные интервалы для коэффициентов

Возможно строить в двух подходах:

- Асимптотически: $t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \to N(0, 1)$
- 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 требующий нормальности ошибок ε_i

- проверяемая H_0 против H_a
- формула для вычисления статистики
- закон распределения статистики при верной H_0

Практическая последовательность действий

1. выбираем уровень значимости α,
   $\alpha = P(H_0$ отвергнута $|H_0$ верна $)$
2. находим наблюдаемое значение статистики S_{obs}
3. находим критическое значение статистики S_{cr}
4. сравниваем критическое и наблюдаемое S_{obs} и S_{cr}

(можно сравнить P-значение и уровень значимости α)

5. вывод: "H_0 отвергается" или "H_0 не отвергается"

Проверка гипотез и построение доверительных интервалов [у доски]

```
summary(model)
            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 *

Residual standard error: 11.07
```

- проверьте гипотезу $\beta_a = 0$
- постройте доверительный интервал для β_a
- постройте доверительный интервал для σ^2

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 статистическом пакете [у доски]

	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	*

Особые моменты при проверки гипотез

- Плохое устоявшееся название гипотез
- Смысл формулировки “H_0 не отвергается”
- Значимость и существенность — разные вещи
- Проблема множественных сравнений

Плохое устоявшееся название

Проверка значимости — на самом деле проверка незначимости:

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

Смысл “H_0 не отвергается”

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

Вполне возможно, что данные не противоречат H_a (!)

Значимость и существенность

- Коэффициент может быть значимым и совершенно несущественным

На огромных выборках как правило все коэффиценты значимы

- Коэффициент может быть существенным, но незначимым

Значимость — статистическое отвержение гипотезы о точном равенстве

Существенность — насколько данное отличие от нуля важно в прикладном смысле

Стандартизированные коэффициенты

Существенность — можно придать разный математический смысл

Например:

- стандартизировать переменные:

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

- 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. Плохой метод.

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

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

Переформулировка модели

Хотим проверить гипотезу $\beta_2 = \beta_3$ или $\beta_2 - \beta_3 = 0$

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

Пример проверки гипотезы о связи коэффициентов [у доски]

```
summary(model)
            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 *
vcov(model)
             (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
```

Проверьте гипотезу $\beta_a = \beta_c$ (два способа)

Мораль лекции 2:

В этой лекции мы научились:

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

В следующей:

- более сложные гипотезы
- прогнозирование