

1. Гомоскедастичность

Для проверки гипотез мы предполагали условную гомоскедастичность ошибок:

$$E(\varepsilon_i^2|X) = \sigma^2$$

Что произойдет если эта предпосылка будет нарушена?

2. Гомоскедастичность:

Условная гомоскедастичность $E(\varepsilon_i^2|X) = \sigma^2$

Условная гетероскедастичность $E(\varepsilon_i^2|X) \neq const$

Безусловная гомоскедастичность $E(\varepsilon_i^2) = \sigma^2$

Безусловная гетероскедастичность $E(\varepsilon_i^2) \neq \sigma^2$

тут вставка чудо-доска

3. Когда логично ожидать гетероскедастичность?

* безусловной в случайной выборке не бывает

* условная присутствует почти всегда

* наличие «размера» объекта

4. В остальном всё ок

Все остальные предпосылки классической модели со стохастическими регрессорами для случайной выборки выполнены.

(тут пачка предпосылок) — файл `predposilki.pdf`

5. Мы используем прежние формулы:

Для оценок коэффициентов: $\hat{\beta} = (X'X)^{-1}X'y$

Для оценки ковариационной матрицы оценок коэффициентов, $\widehat{Var}(\hat{\beta}|X)$
 $\frac{RSS}{n-k}(X'X)^{-1}$

В частности, $\widehat{Var}(\hat{\beta}_j|X) = \frac{\hat{\sigma}^2}{RSS_j}$ и $se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j|X)}$

6. Три группы свойств:

- конечная выборка без предположения о нормальности ε
- конечная выборка с предположением о нормальности ε
- асимптотические свойства (без предположения о нормальности ε)

Что происходит в каждом случае?

7. Конечная выборка без предположения о нормальности ε

* Линейность по y

* Условная несмещенность, $E(\hat{\beta}|X) = \beta$

* (—) Оценки неэффективны

(—) - свойство потеряно при условной гетероскедастичности

8. Конечная выборка с предположением о нормальности ε

$$* \left(- \right) \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} | X \sim t_{n-k}$$

$$* \left(- \right) \frac{RSS}{\sigma^2} | X \sim \chi_{n-k}^2$$

$$* \left(- \right) \frac{(RSS_R - RSS_{UR})/r}{RSS_{UR}/(n-k)} \sim F_{r, n-k}$$

9. Асимптотические свойства:

$$* \hat{\beta} \rightarrow \beta$$

$$* \frac{RSS}{n-k} \rightarrow \sigma^2$$

$$* \left(- \right) \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \rightarrow N(0, 1)$$

$$* \left(- \right) \frac{RSS_R - RSS_{UR}}{RSS_{UR}/(n-k)} \rightarrow \chi_r^2$$

10. Мораль:

- * Сами $\hat{\beta}$ можно интерпретировать и использовать
- * Стандартные ошибки $se(\hat{\beta}_j)$ несостоятельны
- * Не можем строить доверительные интервалы для β_j и проверять гипотезы

11. Что делать?

- * Исправить стандартные ошибки!
- * Другая формула для оценки $\widehat{Var}_{HC}(\hat{\beta}|X)$
- * Следовательно, другие $se_{HC}(\hat{\beta}_j)$

12. Робастная (устойчивая) к гетероскедастичности оценка ковариационной матрицы

- * Вместо $\widehat{Var}(\hat{\beta}|X) = \frac{RSS}{n-k}(X'X)^{-1}$

использовать $\widehat{Var}_{HC}(\hat{\beta}|X) = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}$

- * Уайт, 1980, HC0:

$$\hat{\Omega} = diag(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2)$$

* Современный вариант, НСЗ:

$$\hat{\Omega} = diag \left(\frac{\hat{\varepsilon}_1^2}{(1-h_{11})^2}, \dots, \frac{\varepsilon_n^2}{(1-h_{nn})^2} \right)$$

Суть корректировки:

Мы меняем $se(\hat{\beta}_j)$ на $se_{HC}(\hat{\beta}_j)$

Какие проблемы решены?

$$* \frac{\hat{\beta}_j - \beta_j}{se_{HC}(\hat{\beta}_j)} \rightarrow N(0, 1) \text{ (УРА!)}$$

13. Какие проблемы не решены?

(—) оценки $\hat{\beta}$ не меняются и остаются неэффективными даже при предположении о нормальности ε :

$$* \text{ (—) } \frac{\hat{\beta}_j - \beta_j}{se_{HC}(\hat{\beta}_j)} | X \sim t_{n-k}$$

$$* \text{ (—) } \frac{RSS}{\sigma^2} | X \sim \chi_{n-k}^2$$

$$* \text{ (—) } \frac{(RSS_R - RSS_{UR})/r}{RSS_{UR}/(n-k)} \sim F_{r, n-k}$$

14. С практической точки зрения:

* Новая формула для $\widehat{Var}_{HC}(\hat{\beta}|X)$, и, следовательно, для $se_{HC}(\hat{\beta}_j)$

* ковариационная матрица в R (по умолчанию HC3):

```
model <- lm(y~x, data=data)
```

```
vcovHC(model)
```

* С ней жизнь прекрасна!

$$\frac{\hat{\beta}_j - \beta_j}{se_{HC}(\hat{\beta}_j)} \rightarrow N(0, 1)$$

15. Когда следует использовать

* Как только есть случайная выборка и объекты могут быть разного «размера», использовать $se_{HC}(\hat{\beta}_j)$ для проверки гипотез

16. Обнаружение гетероскедастичности

- * Оцениваем интересующую нас модель с помощью МНК
- * Строим график квадратов (или модулей) остатков в зависимости от регрессора

Тут графики 1 и 2 (присланы как png файлы)

17. Формальные тесты на гетероскедастичность

- * Тест Уайта (White)
- * Асимптотический, не требуется нормальность остатков

18. Тест Уайта начало

- * Оценить основную регрессию, получить $\hat{\varepsilon}_i$
- * Оценить вспомогательную регрессию:

$$\hat{\varepsilon}_i^2 = \gamma_1 + \gamma_2 z_{i2} + \dots + \gamma_{i,k} z_{im} + u_i$$

z_{i2}, \dots, z_{im} — факторы, определяющие форму гетероскедастичности

Посчитать $LM = nR_{aux}^2$

19. Тест Уайта продолжение

При верной H_0 об условной гомоскедастичности

$$H_0: E(\varepsilon_i^2|X) = \sigma^2$$

$LM \sim \chi_{m-1}^2$, где m — число параметров во вспомогательной регрессии

По умолчанию во вспомогательной регрессии берут исходные регрессоры, их квадраты и попарные произведения

Здесь график 3 (прислан как фото рисунка от руки) Подписи на графике:

H_0 не отвергается H_0 отвергается χ_{cr}^2 $H_0 : E(\varepsilon_i^2|X) = const$

20. вставка чудо-доска тест Уайта

По 200 киоскам мороженого и исследователь оценил зависимость спроса (q) от цены (p), разнообразия ассортимента (a) и удаленности от метро (d).

Какой регрессор скорее всего влияет на условную дисперсию ошибок?

Исследователь провел классический тест Уайта и получил $R_{aux}^2 = 0.2$.

Как выглядит вспомогательная регрессия?

Имеет ли место условная гетероскедастичность?

21. Тест Голдфельда-Квандта (Goldfeld-Quandt)

- * Есть переменная, от которой может зависеть условная дисперсия ошибок
- * Требуется нормальность ошибок
- * Тест подходит для малых выборок

22. Процедура теста Голдфельда-Квандта

- * Сортируем наблюдения по предполагаемому убыванию условной дисперсии
- * Выкидываем часть наблюдений посередине (20%)
- * оцениваем исходную модель отдельно по первым и по последним наблюдениям
- * Считаем $F = \frac{RSS_1/(n_1-k)}{RSS_2/(n_2-k)}$

23. Тест Голдфельда-Квандта продолжение

* При верной H_0 об условной гомоскедастичности

$$H_0: E(\varepsilon_i^2|X) = \sigma^2$$

$$F \sim F_{n_1-k, n_2-k}$$

Здесь график 4 (прислан как фото рисунка от руки) Подписи на графике:

H_0 не отвергается H_0 отвергается F_{cr} $H_0 : E(\varepsilon_i^2|X) = const$

24. Вставка с чудо-доской

По 200 киоскам мороженого и исследователь оценил зависимость спроса (q) от цены (p), разнообразия ассортимента (a) и удаленности от метро (d).

Чтобы проверить наличие гетероскедастичности исследователь оценил эту модель отдельно по 80 самым удаленным от метро киоскам, получил, $RSS_2 = 120$. По 80 самым близки к метро киоскам, получил, $RSS_1 = 210$.

25. Эффективность оценок?

- * Да, надо смириться с тем, что оценки неэффективны
- * Мы довольны несмещенностью, состоятельностью и возможностью проверять гипотезы
- * Для получения эффективных оценок нужно точно понимать как устроена гетероскедастичность. Это большая редкость.

26. Вставка с чудо-доской

Задача про среднюю оценку по математике в классе Если бы мы знали как устроена гетероскедастичность...

27. Мораль

- * Мы рассмотрели ситуацию нарушения предпосылки условной гомоскедастичности
- * Почти всегда нарушена
- * Неприятность мелкая, мы используем робастные стандартные ошибки