Анализ остатков

Грауэр Л.В.

Основные предположения

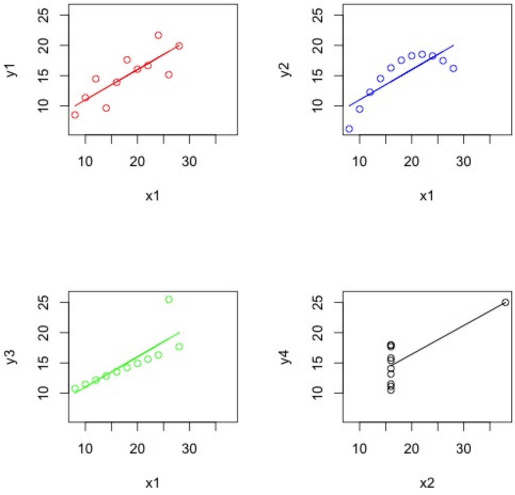
об ошибках наблюдений ε_i

- статистически независимы
- имеют нулевые мат.ожидания
- имеют одинаковые дисперсии
- нормально распределены
- отсутствие выбросов

о спецификации модели

- выбрана верно

$$\hat{y}(x) = 6 + 0.5x$$



y1
10 15 20 25
10 20 30
x1
y2
10 15 20 25
10 20 30
x1
y3
10 15 20 25
10 20 30
x1
y4
10 15 20 25
10 20 30
x2

Последствия отклонений от предположений

- $\hat{\beta}_0, \ldots, \hat{\beta}_k$ смещенные
- оценки дисперсий оценок $\hat{\beta}_0, \ldots, \hat{\beta}_k$ смещенные
- доверительные интервалы для $\beta_0, \ldots, \beta_k$ не соответствуют заявленным уровням значимости
- ошибочные статистические выводы о значимости модели или отдельных парамтеров
- прогнозы смещенные

Виды остатков

$e_i = y_i - \hat{y}_i$

$D(e_i) = \sigma^2(1 - h_{ii})$, $i = 1, \ldots, n$, где $h_{ii} = [A(A^T A)^{-1} A^T]_{ii}$

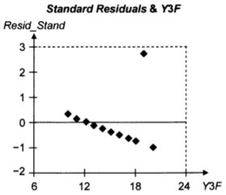
$$\frac{e_i}{\sqrt{D(e_i)}} = \frac{e_i}{\sigma\sqrt{1 - h_{ii}}}, \quad i = 1, \ldots, n.$$

стандартизованные остатки:

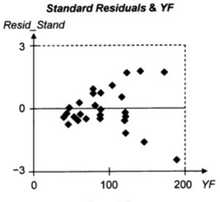
$$d_i = \frac{e_i}{S\sqrt{1 - h_{ii}}}, \quad i = 1, \ldots, n, \quad S^2 = \frac{RSS}{n - k - 1}.$$

Графические методы анализа

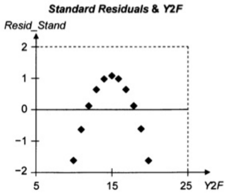
- График зависимости отстатков d_i от оцененных значений $\hat{y}_i = A\hat{\beta}$ позволяет выявить:



наличие выбросов

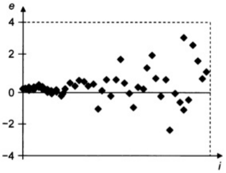


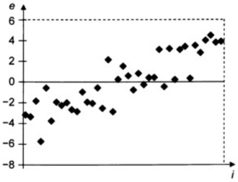
неоднородность дисперсий, гетероскедастичность

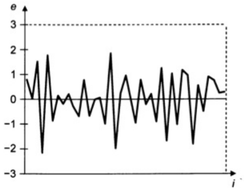


неправильная спецификация модели

- График зависимости d_i от значений объясняющих переменных x_{ij}
- График зависимости остатков от номера наблюдений







- Графические методы проверки предположения о нормальности распределения случайных составляющих (диаграмма "кантиль-квантиль")

Критерии для проверки предположений

Нормальность

- критерий Шапиро-Уилка
- критерий Андерсона-Дарлинга

$E\varepsilon_i = 0$

- параметрический или непараметрический критерий о равенстве мат.ожидания 0

Гомоскедастичность

- Критерий Голдфелда-Квандта

Независимость

- критерий Дарбина-Ватсона

Провека функциональной формы модели

- Критерий Рэмси

Критерий Дарбина-Ватсона

H_0 : отсутствует автокорреляция

H_1: имеет место автокорреляция

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

- $D > D_1(\alpha)$ или $D > 4 - D_1(\alpha) \quad \Rightarrow \quad H_1$
- $D_2(\alpha) < D < 4 - D_2(\alpha) \quad \Rightarrow \quad H_0$
- $D_2(\alpha) > D > D_1(\alpha) \quad \Rightarrow$ нет решения
- $4 - D_1(\alpha) > D > 4 - D_2(\alpha) \quad \Rightarrow$ нет решения

Критерий Голдфелда-Квандта

$$H_0: \quad D(\varepsilon_1) = \ldots = D(\varepsilon_n) = \sigma^2$$

$$H_1: \quad D(\varepsilon_l) \neq D(\varepsilon_p)$$

- Упорядочим данные по возрастанию дисперсий ошибок
- Исключим r средних наблюдений
- Построим 2 модели: по первым $(n-r)/2$ наблюдениям и по последним $(n-r)/2$ наблюдениям
- Вычислим остаточные суммы квадратов RSS_1 и RSS_2

$$F = \frac{RSS_2}{RSS_1} \sim F\left(\frac{n-r}{2} - k - 1, \frac{n-r}{2} - k - 1\right)$$

$$V_k = \{F > F_{1-\alpha}((n-r)/2 - k - 1, (n-r)/2 - k - 1)\}$$

Критерий Рэмси, RESET

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \alpha_2 \hat{y}_i^2 + \ldots + \alpha_m \hat{y}_i^m + \eta_i,$$

где $\hat{y}_i$ — предсказанные значения в соответствии с исходной моделью.

$$H_0 : \alpha_2 = \ldots = \alpha_m = 0$$

$$H_1 : \exists \alpha_p \neq 0$$

Проверяется с помощью F-теста