Условия Гаусса-Маркова

Для того чтобы полученные по МНК оценки обладали некоторым полезными статистическими свойствами необходимо выполнение ряда предпосылок относительно оцениваемой модели, называемыми ***условиями Гаусса-Маркова***.

Условия Гаусса-Маркова

$$y_i = ax_i + b + \varepsilon_i \qquad i = \overline{1,n}$$

1. $M\varepsilon_i = 0 \quad \forall i = \overline{1,n}$

На самом деле это требование несущественно, если в модель включена константа

Условия Гаусса-Маркова

$$y_i = ax_i + b + \varepsilon_i \qquad i = \overline{1,n}$$

2. $D\varepsilon_i = \sigma^2 \ \forall i = \overline{1,n}$ условие гомоскедастичности (постоянства дисперсии)

Условия Гаусса-Маркова

Иллюстрация гомоскедастичности

Условия Гаусса-Маркова

Иллюстрация гетероскедастичности

Условия Гаусса-Маркова

$$y_i = ax_i + b + \varepsilon_i \qquad i = \overline{1,n}$$

3. $\operatorname{cov}\left(\varepsilon_i, \varepsilon_j\right) = 0 \quad \forall\, i \neq j$

автокорреляция отсутствует

Условия Гаусса-Маркова

автокорреляции отсутствует

$$\operatorname{cov}\left(\varepsilon_i, \varepsilon_j\right) = 0$$

автокорреляции присутствует

$$\operatorname{cov}\left(\varepsilon_i, \varepsilon_j\right) \neq 0$$

Условия Гаусса-Маркова

Если выполнены все 3 условия, то модель $y_i = ax_i + b + \varepsilon_i$ **,**

$i = \overline{1, n}$ **называется** **<u>классической линейной моделью</u>**

<u>парной регрессии</u>

Условия Гаусса-Маркова

Если к 3-м условиям добавляют четвертое

4) Нормальность ошибок: $\varepsilon_i \sim N\left(0, \sigma^2\right)$

То модель $y_i = ax_i + b + \varepsilon_i, \quad i = \overline{1,n}$ **называется**

<u>классической нормальной линейной моделью</u>

<u>парной регрессии</u>

Условия Гаусса-Маркова

Предположение о нормальности основано на <u>центральной предельной теореме</u>.

$$\sigma_2^2 > \sigma_1^2$$

ТЕОРЕМА ГАУССА-МАРКОВА

В КЛАССИЧЕСКОЙ ЛИНЕЙНОЙ МОДЕЛИ ПАРНОЙ РЕГРЕССИИ (выполнены 3 условия Гаусса-Маркова) ОЦЕНКИ НАИМЕНЬШИХ КВАДРАТОВ

$$\hat{a} = \frac{\operatorname{cov}(x, y)}{s_x^2} \qquad \hat{b} = \overline{y} - \hat{a}\,\overline{x}$$

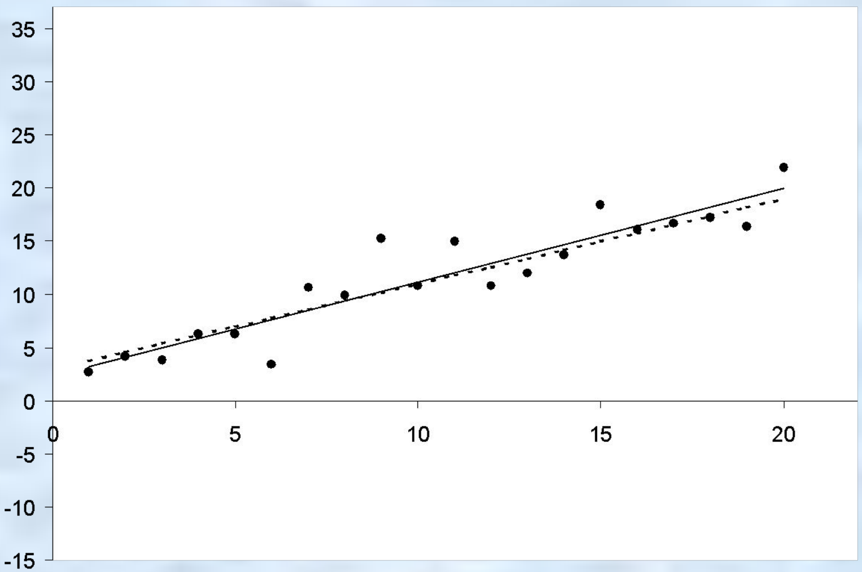
ЯВЛЯЮТСЯ НАИЛУЧШИМИ (имеют наибольшую точность).

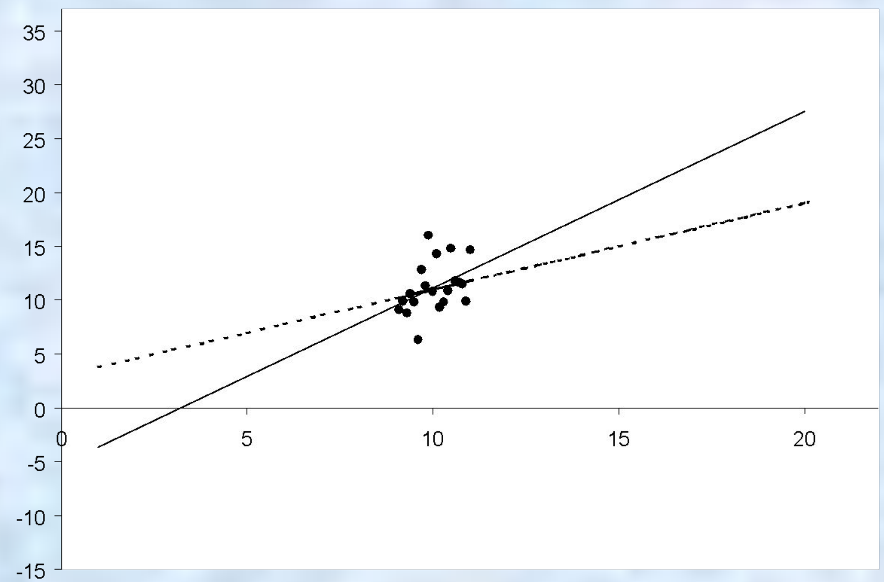
Если модель является нормальной (выполнены 4 условий Гаусса-Маркова), то ОНК имеют нормальное распределение

Нормальность позволяет проверять гипотезы и строить доверительные интервалы для прогноза.

Оценки тем точнее, чем больше наблюдений n
и чем разнообразнее выборка по значениям регрессоров

Оценки тем точнее, чем разнообразнее выборка по значениям регрессоров





s_x^2 на левом рисунке больше, оценка линии регрессии точнее