PROАГРО
ЛЕКТОРИЙ

Корреляционно-регрессионный анализ данных. Множественная регрессия

Кагирова Мария Вячеславовна,
кандидат экономических наук, доцент кафедры статистики и кибернетики РГАУ-МСХА им. К.А. Тимирязева

Множественная регрессия

Вопросы

1. Понятие множественной регрессии и отбор факторов в модель множественной регрессии.
2. Интерпретация параметров множественной регрессии.
3. Оценка качества модели множественной регрессии. Предпосылки МНК.

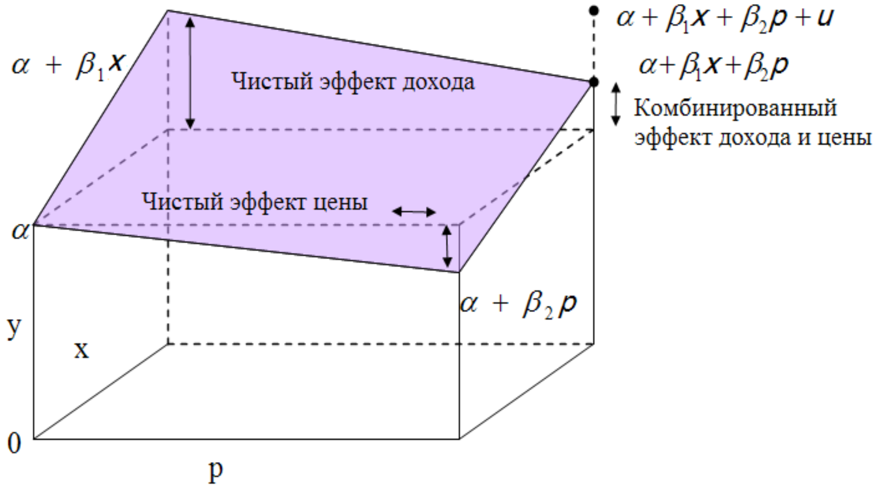
Понятие множественной регрессии

$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$

Множественный регрессионный анализ является развитием парного регрессионного анализа применительно к случаям, когда зависимая переменная гипотетически связана с более чем одной независимой переменной

Основные проблемы множественной регрессии:

- при оценке влияния данной независимой переменной на зависимую переменную придется решать проблему разграничения ее воздействия и воздействий других независимых переменных;
- какие из факторов следует включить в уравнение регрессии, а какие - исключить из него.



Понятие множественной регрессии

Правила отбора факторов:

- включение факторов должно опираться на понимание природы взаимосвязи экономических переменных
- факторы должны быть количественно измеримы
- каждый из факторов не может быть частью другого
- на каждый включенный в уравнение фактор должно приходиться не менее 6-7 наблюдений
- каждый дополнительно включенный в уравнение регрессии фактор должен увеличивать множественный коэффициент детерминации
- факторы, включенные в модель, должны быть независимы друг от друга, то есть они не должны быть интеркоррелированы ($r_{x1x2} \geq 0,7$)

Интерпретация параметров множественной регрессии

$y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + u$

а – условное начало

b_i – коэффициент чистой регрессии измеряет среднее изменение *y* при изменении фактора x_i на единицу, но при условии, что действие других факторов, включенных в уравнение регрессии, учтено и зафиксировано на среднем уровне.

Интерпретация параметров множественной регрессии

$$Э_1 = в_1 \cdot \frac{\bar{x}_1}{\bar{y}}; Э_2 = в_2 \frac{\bar{x}_2}{\bar{y}}$$

$$\beta_1 = в_1 \cdot \frac{\sigma_{x_1}}{\sigma_y}; \beta_2 = в_2 \cdot \frac{\sigma_{x_2}}{\sigma_y}$$

Стандартизованные коэффициенты:
Коэффициенты эластичности (Э) - показывают, на сколько процентов изменится признак-результат, если признак-фактор изменится на один процент.
Бетта-коэффициенты (β) - показывают, на сколько средних квадратических отклонений изменится y, если xi изменится на одно среднее квадратическое отклонение

Интерпретация параметров множественной регрессии

Функция зависимости потребления от дохода в текущем и предыдущем периодах:

$П_t = a + b_1 D_t + b_2 D_{t-1}$

b_1 - краткосрочная предельная склонностью к потреблению, показывает, на сколько увеличится потребление товара при увеличении доходов текущего периода на единицу.

$b = b_1 + b_2$. Коэффициент b рассматривается как долгосрочная склонность к потреблению

Интерпретация параметров множественной регрессии

Модель Кобба-Дугласа – производственная функция

$$y = a x_1^{b_1} x_2^{b_2}$$

b1, b2 являются коэффициентами эластичности, показывают, на сколько процентов изменяется в среднем результат с изменением соответствующего фактора на 1 процент при неизменности других факторов.

B = b1+b2 обобщенная характеристика эластичности производства (показывает, на сколько процентов в среднем увеличиваются объемы производства при увеличении всех факторов на 1%)

Оценка качества модели множественной регрессии

$$R^2 = \frac{\sigma^2_{воспр.}}{\sigma^2_{общ.}}$$

$$d^2{}_i = в_i \cdot \frac{Cov(x_i, y)}{\sigma^2{}_y} \qquad d^2{}_i = r_{y.x_i} \cdot \beta_i$$

$$\sum_{i=1}^{n} d_i^2 = R^2$$

Наиболее общим показателем тесноты связи всех входящих в уравнение регрессии факторов с результативным признаком является коэффициент множественной детерминации R2

Вклад каждого из факторов в формирование вариации результативного признака позволяют оценить коэффициенты раздельной детерминации $d^2{}_i$

Оценка качества модели множественной регрессии

$$R_{y.x_1x_2...x_p} = \sqrt{1 - \frac{\sigma_{ocm}^{2}}{\sigma^2_y}}$$

$$\bar{R}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1}$$

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции R

Для более точной характеристики качества модели (обеспеченного включением только существенных факторов) рассчитывают скорректированный коэффициент детерминации $\bar{R}^2$

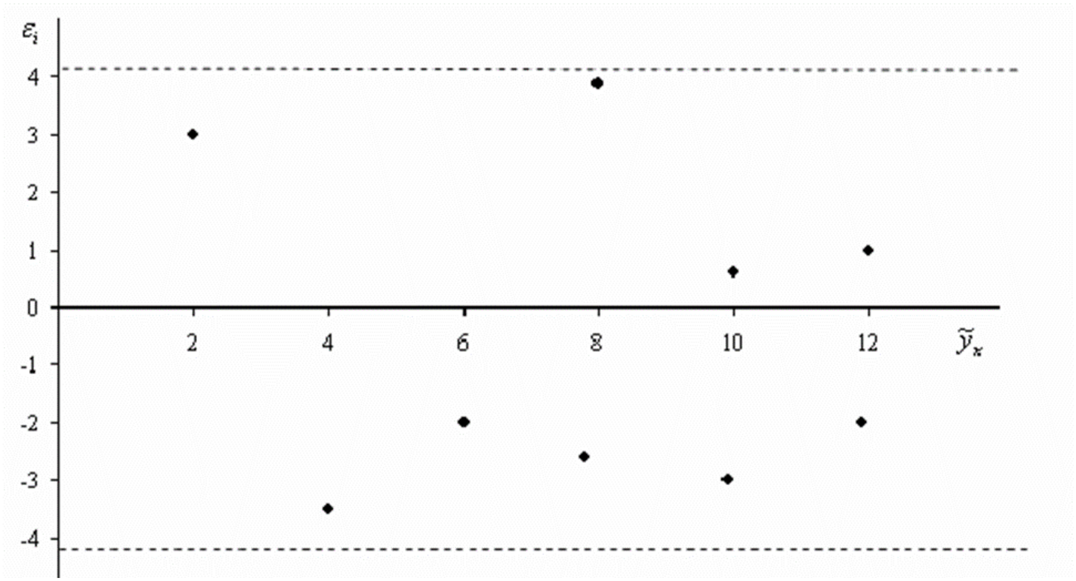
Оценка качества модели множественной регрессии

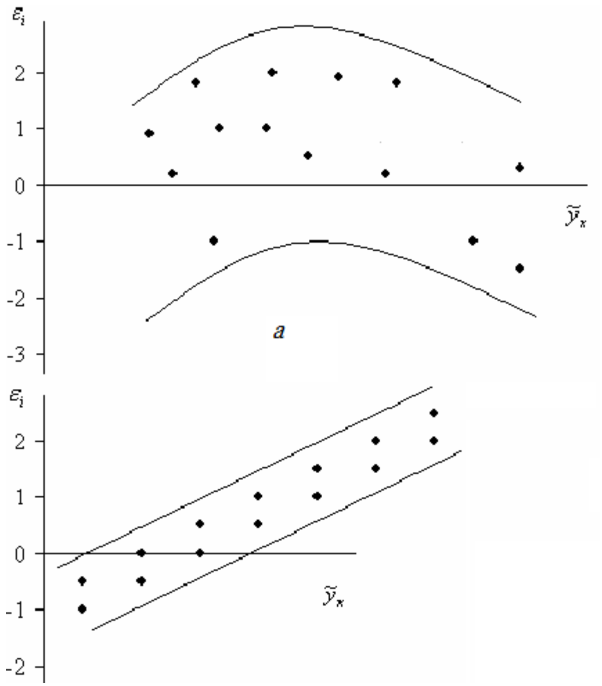
$$r_{yx2\cdot x1} = \sqrt{\frac{s_{yx1}^{2} - s_{yx1x2}^{2}}{s_{yx1}^{2}}} \qquad r_{yx1\cdot x2} = \sqrt{\frac{s_{yx2}^{2} - s_{yx1x2}^{2}}{s_{yx2}^{2}}}$$

Алгоритм построения модели методом исключения переменных:

- определяется уравнение регрессии с полным набором факторов;
- рассчитывается матрица частных коэффициентов корреляции;
- отбирается фактор с наименьшей и несущественной по критерию t-Стьюдента величиной показателя частной корреляции, он исключается из модели;
- строится новое уравнение регрессии и процедура повторяется до тех пор, пока не окажется, что все оставшиеся факторы существенно отличаются от нуля. Если исключен несущественный фактор, то множественные коэффициенты детерминации на двух смежных шагах построения модели почти не отличаются друг от друга.

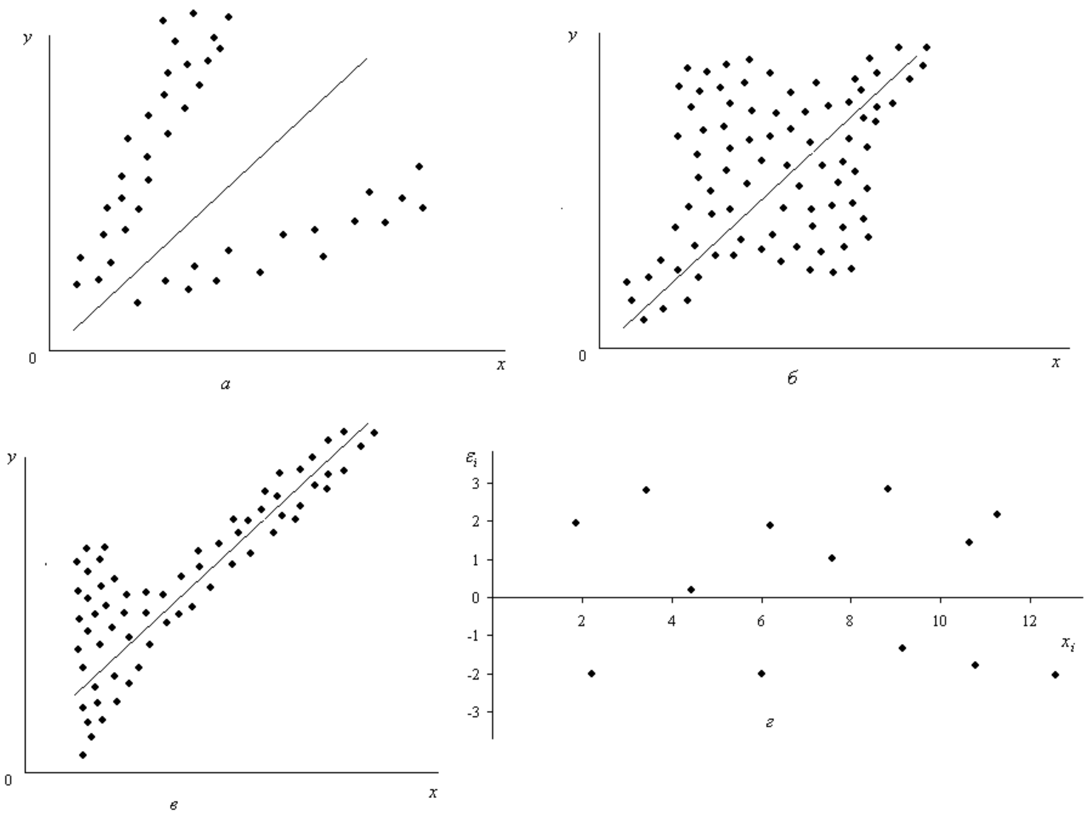
Предпосылки МНК





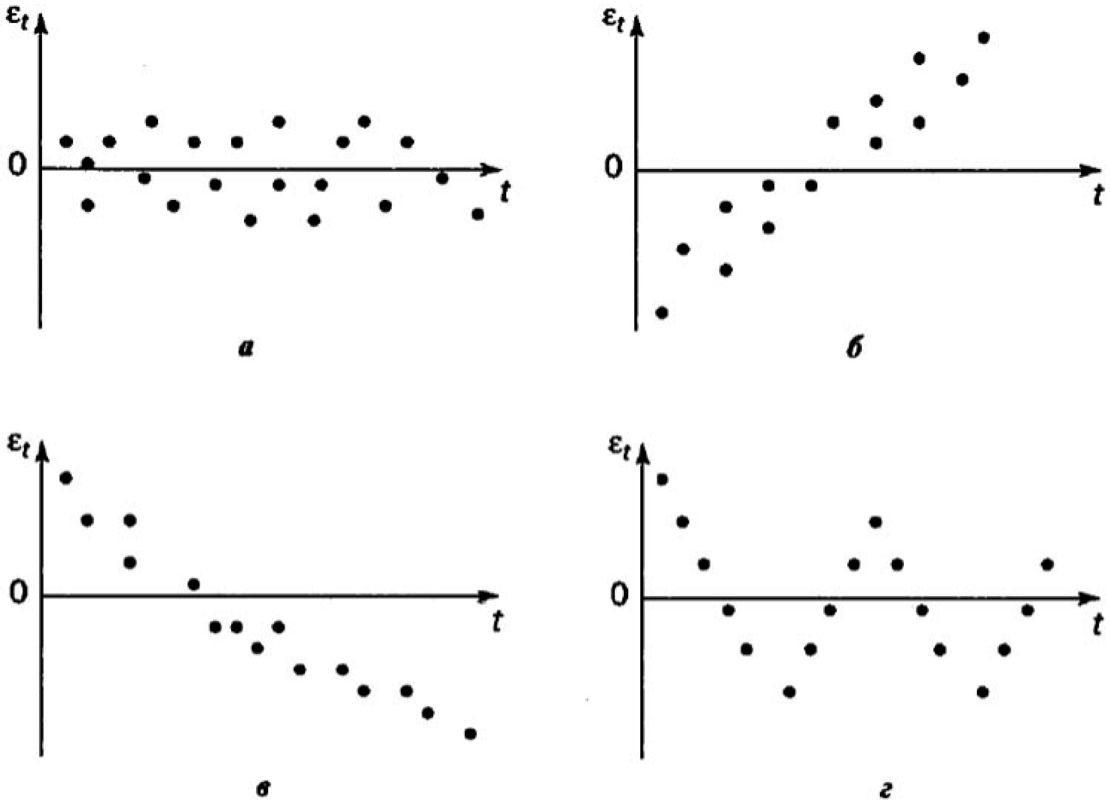
Случайный характер остатков

Предпосылки МНК



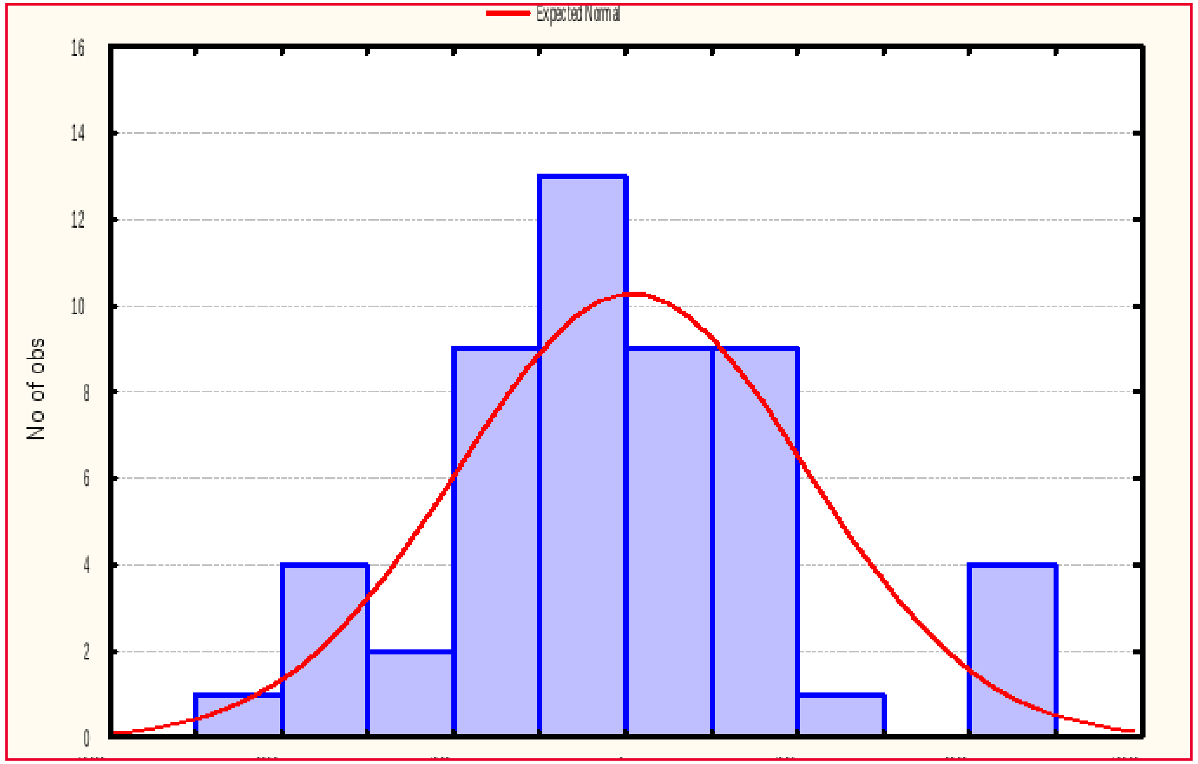
Гомоскедастичность – дисперсия остатков одинакова для всех значений фактора

Предпосылки МНК



Отсутствие автокорреляции остатков (то есть остатки распределены независимо друг от друга)

Предпосылки МНК



Остатки подчиняются нормальному закону распределения

PROАГРО
ЛЕКТОРИЙ

Спасибо за внимание!