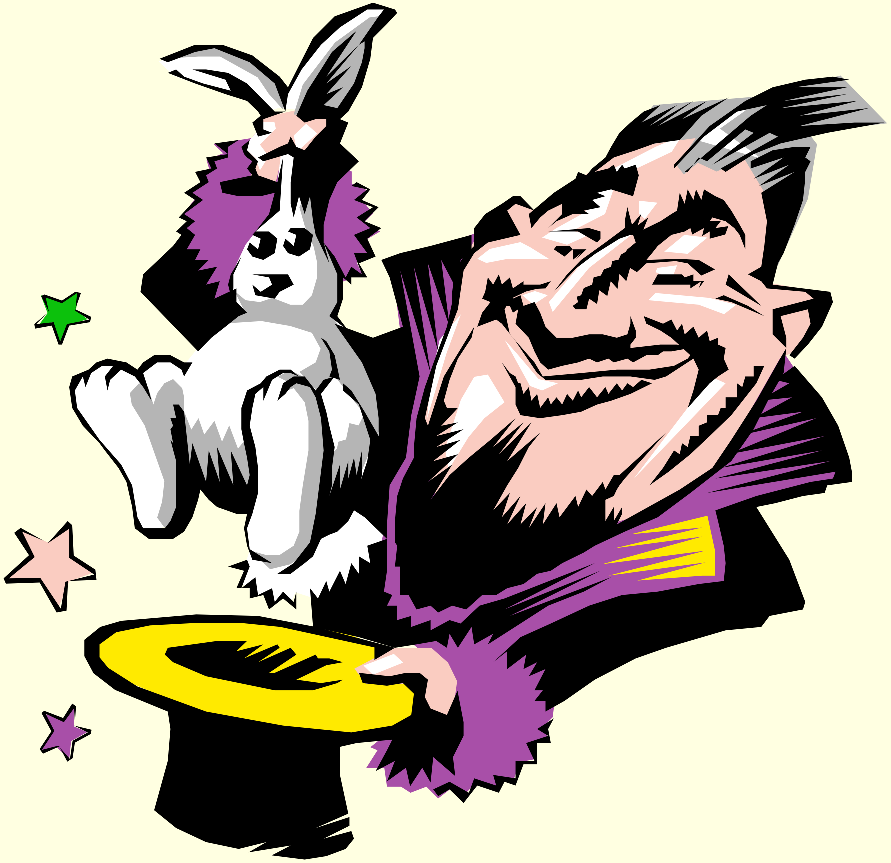
Нелинейные модели

- Нелинейная регрессия
- Преобразования переменных
- Экономическая интерпретация регрессионной модели



Нелинейные модели

Модели линейные по переменным и параметрам:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

Модели линейные по параметрам и нелинейные по переменным:

$$Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + u$$

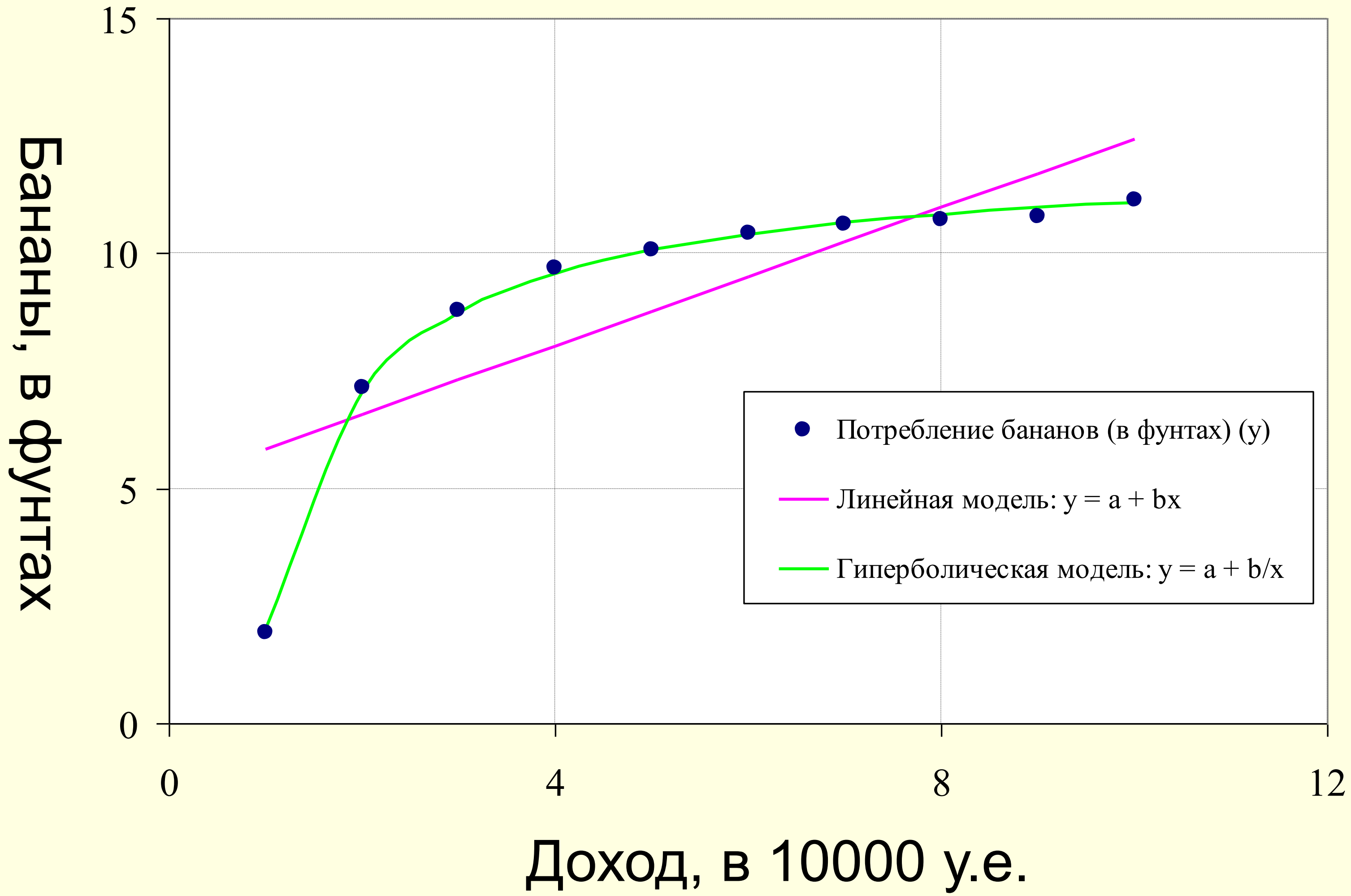
$$Z_2 = X_2^2, \quad Z_3 = \sqrt{X_3}, \quad Z_4 = \log X_4$$

$$Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + u$$

Модели нелинейные по параметрам:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_2 \beta_3 X_4 + u$$

Пример нелинейной зависимости



Этапы построения модели

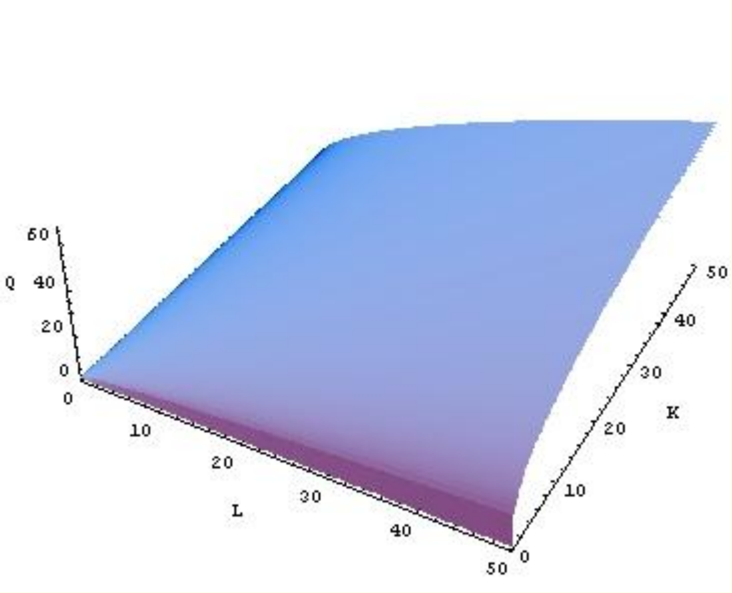
1. Выбор теоретических предпосылок
2. Формализация предпосылок
3. Построение математической модели
4. Анализ построенной модели

Производственная функция Кобба-Дугласа

Многие экономические процессы не являются линейными по сути. Их моделирование линейными уравнениями не даст положительного результата.

<ins>Пример</ins>. Производственная функция Кобба – Дугласа

$$Y = AK^{\alpha}L^{\beta}$$



Y – объем выпуска;

A – техн коэфф;

K, L – затраты капитала и труда;

α, β – параметры модели.

Анализ экономического роста

Анализ теоретических предпосылок: прирост пропорционален накопленному потенциалу

Формализация предпосылок:

$$dY = \beta Y \Rightarrow \frac{dY}{Y} = \beta \Rightarrow \ln Y = \alpha + \beta t$$

Интерпретация и анализ: коэффициент регрессии β – годовой темп роста, возможно сопоставление с реальными данными

Классы нелинейных регрессий

Различают ***два класса*** нелинейных регрессий:

1. Регрессии, нелинейные относительно переменных, но линейные по оцениваемым параметрам.

2. Регрессии, нелинейные по оцениваемым параметрам.

Регрессии, нелинейные относительно объясняющих переменных, всегда сводятся к линейным моделям.

Правила выбора формы зависимости

1. ***Теоретическое обоснование***
2. Ф***ормальная оценка качества модели***
3. ***Дополнительная проверка качества модели с использованием нескольких содержательных критериев***
4. ***Ответ на вопросы:***

- каковы признаки качественной модели;
- какие ошибки спецификации встречаются и каковы их последствия;
- как обнаружить ошибку спецификации;
- каким образом можно исправить ошибку спецификации и перейти к более качественной модели.

Линейная форма

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Интерпретация коэффициента регрессии

β – предельный эффект независимого фактора

$$\beta = Y'_X = \frac{dY}{dX} = \frac{\Delta Y}{\Delta X}$$

Линейная форма

Для полученных оценок *a*, *b* уравнения регрессии:

$$Y = a + bX$$

$$b = \frac{\Delta Y}{\Delta X} \qquad \Delta X = 1 \qquad \Rightarrow \qquad b = \Delta Y$$

$$a = \overline{Y} - b\overline{X} \qquad \Rightarrow \qquad a = \overline{Y}(\overline{X} = 0)$$

Линейная форма

Коэффициент регрессии *b* показывает прирост зависимой переменной при изменении объясняющей переменной на единицу.

Коэффициент регрессии *b* – угловой коэффициент линии регрессии

Коэффициент регрессии *a* – среднее значение зависимой переменной при нулевом значении объясняющей переменной

Линейная зависимость от времени

$$Y_i = a + bt_i$$

Интерпретация коэффициента регрессии от времени – ежегодный (ежемесячный и т.д.) прирост зависимой переменной

$$b = \frac{\Delta Y}{\Delta t}$$

Моделирование эластичности

Независимо от вида математической связи между *Y* и *X* ***эластичность*** равна:

$$L = \frac{dY / Y}{dX / X} = \frac{dY / dX}{Y / X} \approx \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}$$

Эластичность ***у*** по ***х*** рассчитывается как относительное изменение ***у*** на единицу относительного изменения ***х***

Пример расчета эластичности

Кривая Энгеля:

$$Y = \alpha X^{\beta}$$

где ***Y*** – спрос на товар, ***X*** – доход.

Иллюстрирует зависимость между объёмом потребления товаров или услуг и доходом потребителя при неизменных ценах и предпочтениях.

$$\text{Эластичность} = \frac{dY/dX}{Y/X} = \frac{\alpha\beta X^{\beta-1}}{\alpha X^{\beta-1}} = \beta,$$

Пример: для модели $Y = 0{,}01X^{0,3}$ эластичность спроса по доходу равна 0,3.

Или: изменение дохода (*X*) на 1% вызывает изменение спроса (*Y*) на 0,3%

Эластичность – переменная величина

Эластичность не всегда бывает постоянной для различных значений *X* и *Y*

Для линейной модели

$$L = \frac{dY / dX}{Y / X} = \frac{\beta}{Y / X} = \beta \frac{X}{Y}$$

Средний коэффициент эластичности

Средний коэффициент эластичности $\bar{L}$ показывает, на сколько процентов в среднем по совокупности изменится результат *Y* от своей средней величины при изменении фактора *X* на 1% от своего среднего значения

$$\bar{L} = f'(\bar{X})\frac{\bar{X}}{\bar{Y}}$$

Логарифмическая форма

$$Y_i = \alpha' X_i^{\beta} \varepsilon_i'$$

Прологарифмировав обе части уравнения, получим

$$\ln Y_i = \alpha + \beta \ln X_i + \varepsilon_i$$

Логарифмическая форма

$$\ln Y_i = \alpha + \beta \ln X_i + \varepsilon_i$$

Интерпретация коэффициента регрессии β – эластичность зависимой переменной по объясняющей переменной

$$\frac{dY}{Y} = \beta \frac{dX}{X} \quad \Rightarrow \quad \beta = \frac{dY / Y}{dX / X}$$

Коэффициент при объясняющей переменной показывает,

на сколько процентов возрастает Y при возрастании X на 1%.

Логарифмическую форму используют, где есть основание предполагать постоянство эластичности

Логарифмическая форма

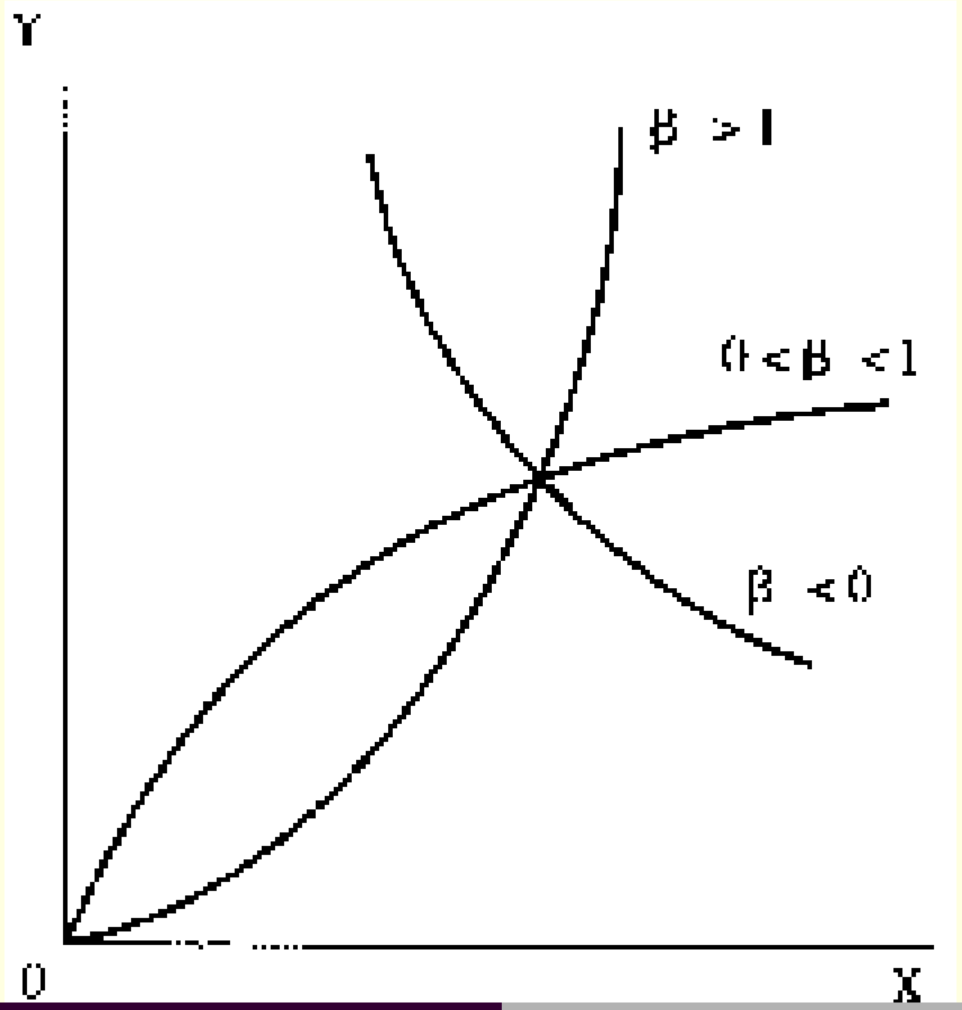
$$\ln Y_i = \alpha + \beta \ln X_i + \varepsilon_i$$

Вычисление наклона (скорости роста)

$$\frac{dY}{dX} = \beta \frac{Y}{X}$$

Наклон постоянно меняется с изменением номера наблюдения

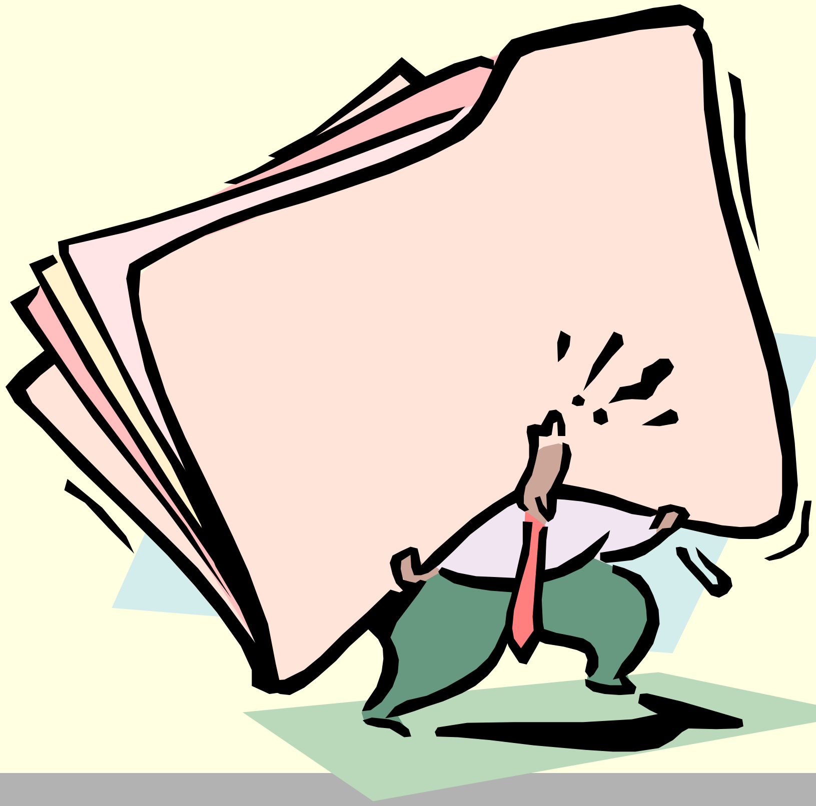
Графики логарифмической формы зависимости



Полулогарифмические формы

1.Линейно-логарифмическая форма
(логарифм при объясняющей переменной)

2.Логарифмически-линейная форма
(логарифм при зависимой переменной)



Линейно-логарифмическая форма

$$Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 X_2 + \varepsilon$$

Интерпретация коэффициента регрессии β:

$$dY = \beta \frac{dX}{X} \quad \Rightarrow \quad \beta = \frac{dY}{dX / X} \quad \Rightarrow \quad \frac{\beta}{100} = \frac{dY}{100 \cdot (dX / X)}$$

Коэффициент при объясняющей переменной показывает, на сколько единиц возрастает *Y* при возрастании *X* на 1%

Если *X* увеличится на 1%, то прирост *Y* составит β /100 единиц (в которых измеряется *Y*). При интерпретации коэффициент следует делить на 100.

Линейно-логарифмическая форма

$$Y_i = \alpha + \beta \ln X_i + \varepsilon_i$$

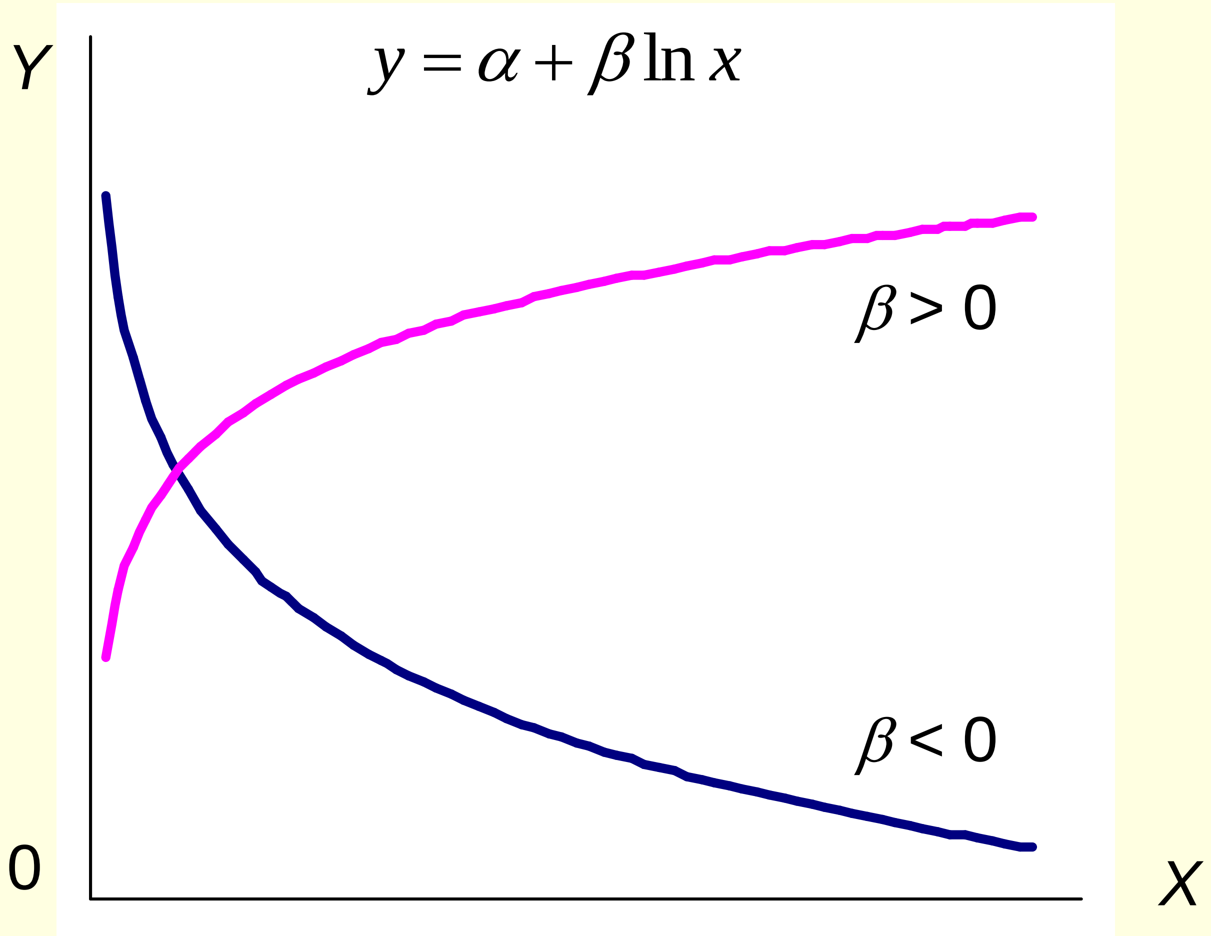
Эластичность убывает с ростом *Y*:

$$dY = \beta \frac{dX}{X} \quad \Rightarrow \quad L = \frac{dY}{dX} \cdot \frac{X}{Y} = \beta \frac{1}{X} \cdot \frac{X}{Y} = \beta \frac{1}{Y}$$

Это указывает на класс зависимостей, где следует применять линейно-логарифмическую форму регрессии

Логарифм при *X* снижает влияние роста *X* (*степень влияния X снижается с ростом X*). Моделирование эффектов насыщения на уровне скорости роста: «возрастание с убывающей скоростью»

Графики линейно-логарифмической формы зависимости



Логарифмически-линейная форма

$$\ln Y_i = \alpha + \beta X_i + \varepsilon_i$$

Интерпретация коэффициента регрессии β:

$$\frac{dY}{Y} = \beta dX \quad \Rightarrow \quad \beta = \frac{dY}{YdX} \quad \stackrel{dX=1}{\Rightarrow} \quad \beta \cdot 100\% = \frac{dY}{Y} \cdot 100\%$$

Коэффициент при объясняющей переменной показывает, на сколько процентов возрастает *Y* при возрастании *X* на одну единицу

При интерпретации коэффициент следует умножать на 100

Логарифмически-линейная форма

$$\ln Y_i = \alpha + \beta X_i + \varepsilon_i$$

Эластичность растет с ростом *Y*:

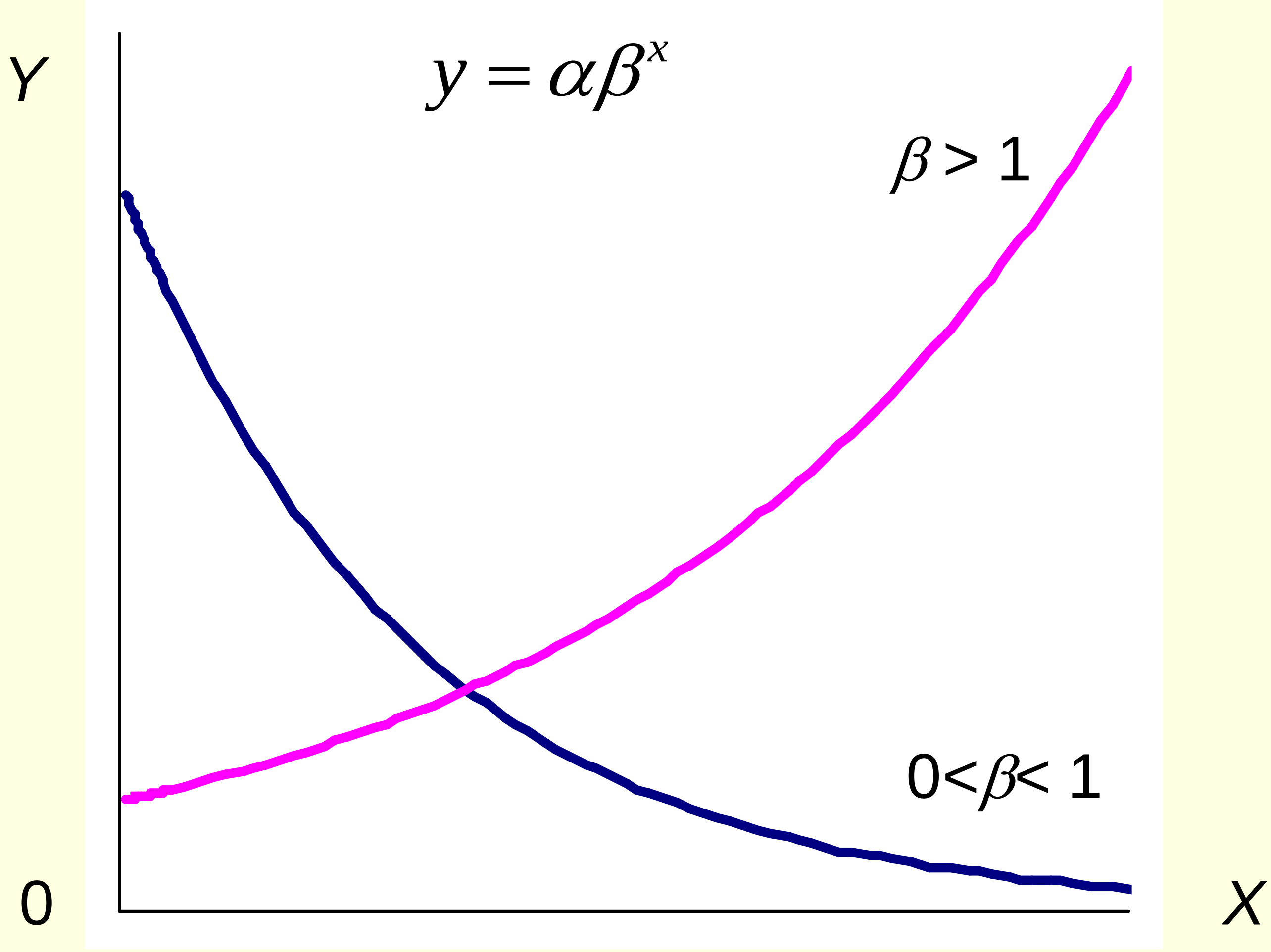
$$\beta = \frac{dY}{YdX} \quad \Rightarrow \quad L = \frac{dY}{dX} \cdot \frac{X}{Y} = \beta \frac{YX}{Y} = \beta X$$

Это указывает на класс зависимостей, где следует применять линейно-логарифмическую форму регрессии

Моделирование эффектов насыщения на уровне скорости роста: «возрастание с возрастающей скоростью»

Примеры: кривые Энгеля для товаров роскоши, моделирование оплаты труда (процентная надбавка за стаж и опыт)

Графики логарифмически-линейной формы зависимости



Логарифмически-линейная форма от времени

$$\ln Y_i = \alpha + \beta t_i + \varepsilon_i$$

Вид уравнения: $Y_i = e^{\alpha} e^{\beta t_i} e^{\varepsilon_i}$

Интерпретация: $\dfrac{dY}{Y} = \beta t$

Коэффициент при переменной времени выражает темп прироста. Он показывает на сколько процентов (если умножить его на 100) возрастает *Y* ежегодно

Эту функциональную форму удобно использовать для моделирования процессов экономического роста

Обратные зависимости

$$Y_i = \alpha + \beta \frac{1}{X_i} + \varepsilon_i$$

Вычисление эластичности

$$L = \frac{dY/Y}{dX/X} = \beta \cdot \left(-\frac{1}{XY} \right)$$

С ростом X зависимая переменная приближается к некоторому числу (моделирование эффекта насыщения)

<u>Пример</u>: Моделирование потребления товаров первой необходимости (быстрое достижение насыщения)

Сводка результатов для альтернативных функциональных форм в парной регрессии

Функциональная форма	Уравнение (без учета других факторов)	Наклон $\frac{\Delta Y}{\Delta X}$	Эластичность $\frac{\Delta Y}{\Delta X}\cdot\frac{X}{Y}$
Линейная	$Y_i = \alpha + \beta X_i + u$	β	$\beta\left(\frac{X_i}{Y_i}\right)$
Двойная логарифмическая	$\ln Y_i = \alpha + \beta \ln X_i + u$	$\beta\left(\frac{Y_i}{X_i}\right)$	β
Линейно-логарифмическая (lnX)	$Y_i = \alpha + \beta \ln X_i + u$	$\beta\left(\frac{1}{X_i}\right)$	$\beta\left(\frac{1}{Y_i}\right)$
Логарифмически-линейная (lnY)	$\ln Y_i = \alpha + \beta X_i + u$	βY_i	βX_i
Обратная	$Y_i = \alpha + \beta\frac{1}{X_i} + u$	$-\beta\left(\frac{1}{X_i^2}\right)$	$-\beta\left(\frac{1}{X_i Y_i}\right)$

Преобразование случайного отклонения

МНК применяется к преобразованным (линеаризованным) уравнениям. Поэтому необходимо особое внимание уделять рассмотрению свойств случайных отклонений – выполнимости предпосылок теоремы Гаусса-Маркова.

Пример.

$$Y = \alpha X^{\beta} + \varepsilon \quad \overset{\ln(\cdot)}{\Rightarrow} \quad \ln Y = \ln(\alpha X^{\beta} + \varepsilon)$$

Логарифмирование нелинейной модели с аддитивным случайным членом не приводит к линеаризации соотношения относительно параметров.

Признаки качественной модели

1. ***Простота модели*** (из примерно одинаково отражающих реальность моделей, выбирается та, которая содержит меньше объясняющих переменных.
2. ***Единственность*** (для любых данных коэффициенты модели должны вычисляться однозначно).
3. ***Максимальное соответствие*** (модель тем лучше, чем больше скорректированный коэффициент детерминации).
4. ***Согласованность с теорией*** (уравнение регрессии должно соответствовать теоретическим предпосылкам).
5. ***Прогнозные качества*** (прогнозы, полученные на основе модели, должны подтверждаться реальностью).

Сравнение различных моделей

1. Содержательный анализ

2. Формальный анализ:

- *Метод Зарембки*
- *Преобразование Бокса-Кокса*

Метод Бокса-Кокса

Идея метода. Переменная $\dfrac{Y^{\lambda}-1}{\lambda}$:

при λ=1 превращается в линейную функцию $\dfrac{Y-1}{1}$

при $\lambda \to 0$ переходит в логарифм $\dfrac{Y^{\lambda}-1}{\lambda} \to \ln Y$

Плавно изменяя λ, можно постепенно перейти от линейной регрессии к логарифмической, все время сравнивая качество

Сравнение различных моделей парной регрессии методом Бокса-Кокса

1. Преобразуют зависимую переменную по методу Зарембки:

$$Y_i^* = Y_i / \sqrt[n]{Y_1 Y_2 \ldots Y_n} = Y_i / e^{\frac{1}{n}\sum_{i=1}^{n} \ln Y_i}$$

2. Рассчитывают новые переменные (преобразование Бокса-Кокса) при значениях λ от 1 до 0:

$$Y_i^{(B-C)} = \frac{{Y_i^*}^{\lambda} - 1}{\lambda} \qquad X_i^{(B-C)} = \frac{X_i^{\lambda} - 1}{\lambda}$$

Сравнение различных моделей парной регрессии методом Бокса-Кокса

3. Рассчитывают уравнения регрессии для новых переменных при значениях λ от 1 до 0:

$$Y_i^{(B-C)} = \alpha + \beta X_i^{(B-C)} + \varepsilon_i$$

4. Определяют минимальное значение суммы квадратов остатков (SSR).

5. Выбирают одну из крайних регрессий, к которой ближе точка минимума.

Вопросы для самопроверки

- Какие вы знаете виды нелинейных моделей.
- Какие вы знаете нелинейные методы оценивания.
- Как определять эластичность.
- Что такое предельные эффекты переменных.
- Основные способы линеаризации моделей.
- Какие вы знаете типы производственных функций.
- Как выбрать между линейной и логарифмической моделями.
- Экономический смысл коэффициентов линейной модели.
- Экономический смысл коэффициентов логарифмической модели
- Экономический смысл коэффициентов полулогарифмической модели.

Случайные составляющие коэффициентов регрессии

- Модель :

$$y = \alpha + \beta x + u$$

- x – неслучайная экзогенная переменная
- Уравнение регрессии :

$$\hat{y} = a + bx$$

- Коэффициенты регрессии – случайные величины
- <u>Теорема</u>

$$b = \frac{Cov(x, y)}{Var(x)} = \beta + \frac{Cov(x, u)}{Var(x)}$$

2. Условия Гаусса-Маркова (предположения о случайном члене)

- 1. $E(u)=0$
- 2. $Var(u)$ постоянна для всех наблюдений
- 3. $Cov(u_i,u_j)=0 \qquad i\neq j$
- 4. $Cov(x_i,u_i)=0$

или $E(x_i,u_i)=0$

Дополнительно: u распределено по нормальному закону

Необходимо понимать

- Если условия не выполнимы, вы должны это сознавать
- Если корректирующие действия возможны, то аналитик должен быть в состоянии их выполнить
- Если ситуацию исправить невозможно, вы должны быть способны оценить, насколько серьезно это может влиять на результаты

Несмещенность коэффициентов регрессии

- 1. $$E(b)=\beta+E\left\{\frac{Cov(x,u)}{Var(x)}\right\}$$
  - И если x – неслучайная величина, то

$$E(b)=\beta$$

- 2. $$E(a)=\alpha$$

Точность коэффициентов регрессии

- Теоретические дисперсии оценок a и b

$$Var(a) = \frac{\sigma_u^2}{n}\left\{1 + \frac{\bar{x}^2}{Var(x)}\right\}$$

$$Var(b) = \frac{\sigma_u^2}{nVar(x)}$$

- Выводы:
  - a и b прямо пропорциональны дисперсии остаточного члена
  - чем больше число наблюдений, тем меньше дисперсии
  - чем больше дисперсия x, тем меньше дисперсии коэффициентов

Оценки стандартных отклонений коэффициентов

- Термин «стандартная ошибка»

$$m_b = \sqrt{\frac{\sum\left(y-\widehat{y}_x\right)^2 / (n-2)}{\sum\left(x-\bar{x}\right)^2}} = \sqrt{\frac{S^2}{\sum\left(x-\bar{x}\right)^2}}$$

$$m_a = \sqrt{\frac{\sum\left(y-\widehat{y}_x\right)^2}{n-2} \cdot \frac{\sum x^2}{n \cdot \sum\left(x-\bar{x}\right)^2}} = \sqrt{S^2 \cdot \frac{\sum x^2}{n \cdot \sum\left(x-\bar{x}\right)^2}}$$

- Применение: 1) проверка существенности коэффициента регрессии
  2)

5. Оценка существенности коэффициента регрессии и свободного члена

- Фактическое значение t-критерия сравнивается с
- табличным значением при определенном уровне
- значимости α и числе степеней свободы (n-2)
- $t_b = \dfrac{b}{m_b}, \qquad t_a = \dfrac{b}{m_a}$ - фактические значения
- Теорема $t_b^2 = F$
- Доказательство:

$$t_b^2 = \frac{b^2}{m_b^2} = b^2 / \frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}$$

Продолжение доказательства

$$= \frac{b^2 \sum (x - \bar{x})^2}{\sum (y - \hat{y}_x)^2 / (n-2)} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\frac{\sum (y - \hat{y}_x)^2}{(n-2)}}$$

$$= \frac{D_{факт}}{D_{ост}} = F$$

Дисперсионный анализ результатов регрессии

Источники вариации	Число степеней свободы	Сумма квадратов отклонений	Дисперсия на одну степень свободы	F-отношение	
				фактическое	табличное при α=0,05
Общая	6	15000	--	--	--
Объясненная	1	14735	14735	278	6,61
Остаточная	5	265	53	1	--

$$m_b = \sqrt{\frac{53}{10,857}} = 2,21$$

$$t_b = \frac{36,84}{2,21} = 16,67 > t_{табл} = 2,57$$

Значимость коэффициента корреляции

- Стандартная ошибка коэффициента корреляции

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

- Фактическое значение t-критерия Стьюдента

$$t_r = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n - 2}$$

Теорема $\quad t_r^2 = t_b^2 = F$

- Доказательство:
  - Смотри выражения для t_r и F.

- **Вывод:** проверка гипотез о значимости коэффициента регрессии b и коэффициента корреляции r равносильна проверке гипотезы о существенности линейного уравнения в целом.

Доверительные интервалы параметров регрессии

- Доверительный интервал для коэффициента регрессии

$$(b - t_{табл} * m_b, b + t_{табл} * m_b)$$

- Доверительный интервал для свободного члена регрессии

$$(a - t_{табл} * m_a, a + t_{табл} * m_a)$$

- Доверительный интервал для коэффициента корреляции
- $$(r - t_{табл} * m_r, r + t_{табл} * m_r)$$

Интервалы прогноза по линейному уравнению регрессии

- Точечный прогноз y_p дополняется интервальной оценкой прогнозного значения

$$\widehat{y}_x - m_{\widehat{y}_x} \le y^* \le \widehat{y}_x + m_{\widehat{y}_x}$$

- Теорема

$$m_{\widehat{y}_x} = S \cdot \sqrt{\frac{1}{n} + \frac{\left(x_k - \bar{x}\right)^2}{\sum\left(x - \bar{x}\right)^2}}$$

Доказательство.

$$\hat{y}_x = \bar{y} - b \cdot \bar{x} + b \cdot x = \bar{y} + b \cdot (x - \bar{x})$$

- Тогда

$$m_{\hat{y}_x}^2 = m_{\bar{y}}^2 + m_b^2 (x - \bar{x})^2$$

- т.к. $Var(\bar{y}) = \dfrac{\sigma^2}{n}$,то $m_{\bar{y}}^2 = \dfrac{S^2}{n}$

- и $m_b^2 = \dfrac{S^2}{\sum (x - \bar{x})^2}$ $\Rightarrow$

$$m_{\hat{y}_x}^2 = \frac{S^2}{n} + \frac{S^2}{\sum (x - \bar{x})^2} \cdot (x_k - \bar{x})^2 = S^2 \cdot \left(\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2} \right)$$

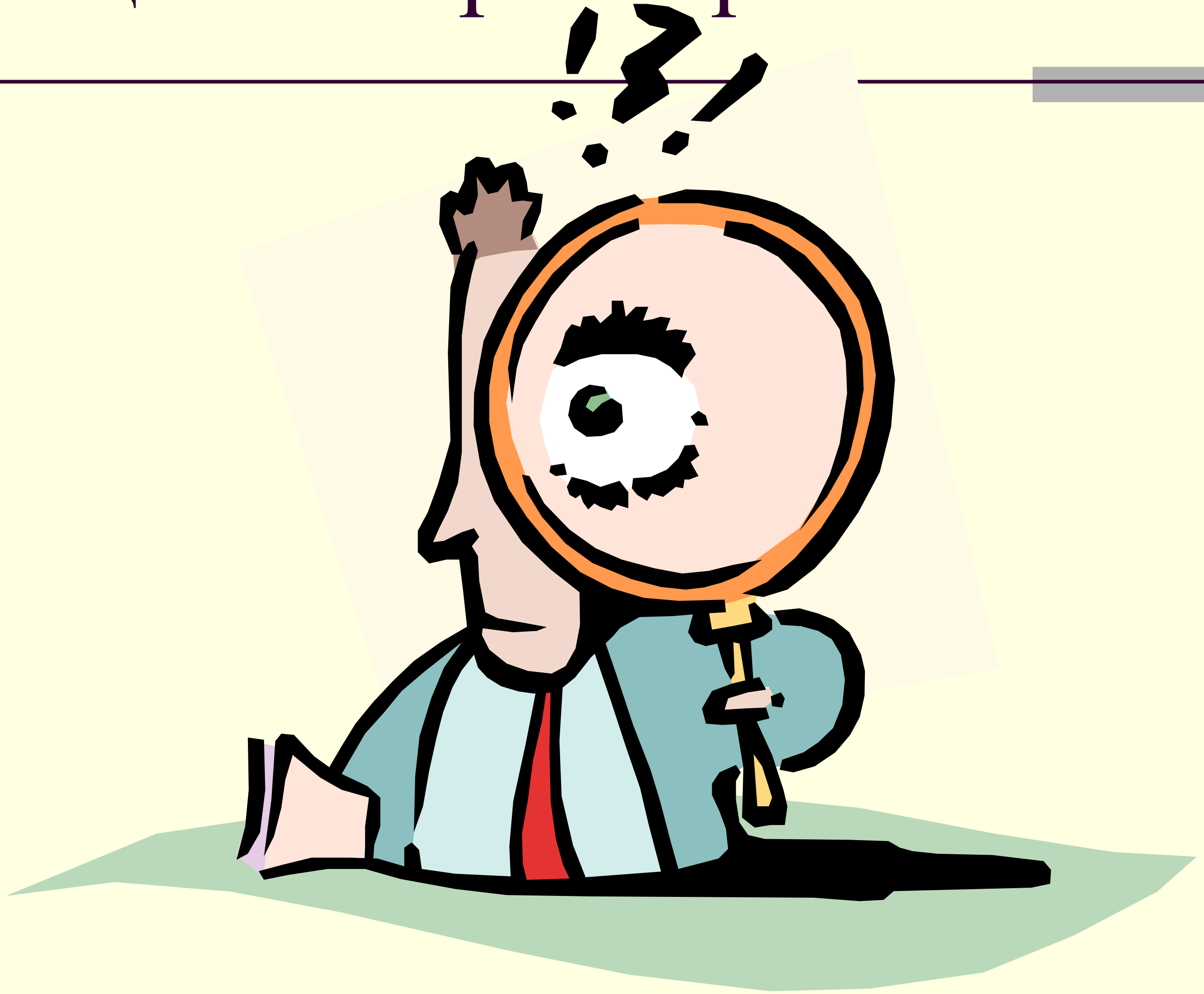
Доверительные границы для $\hat{y}_x$:

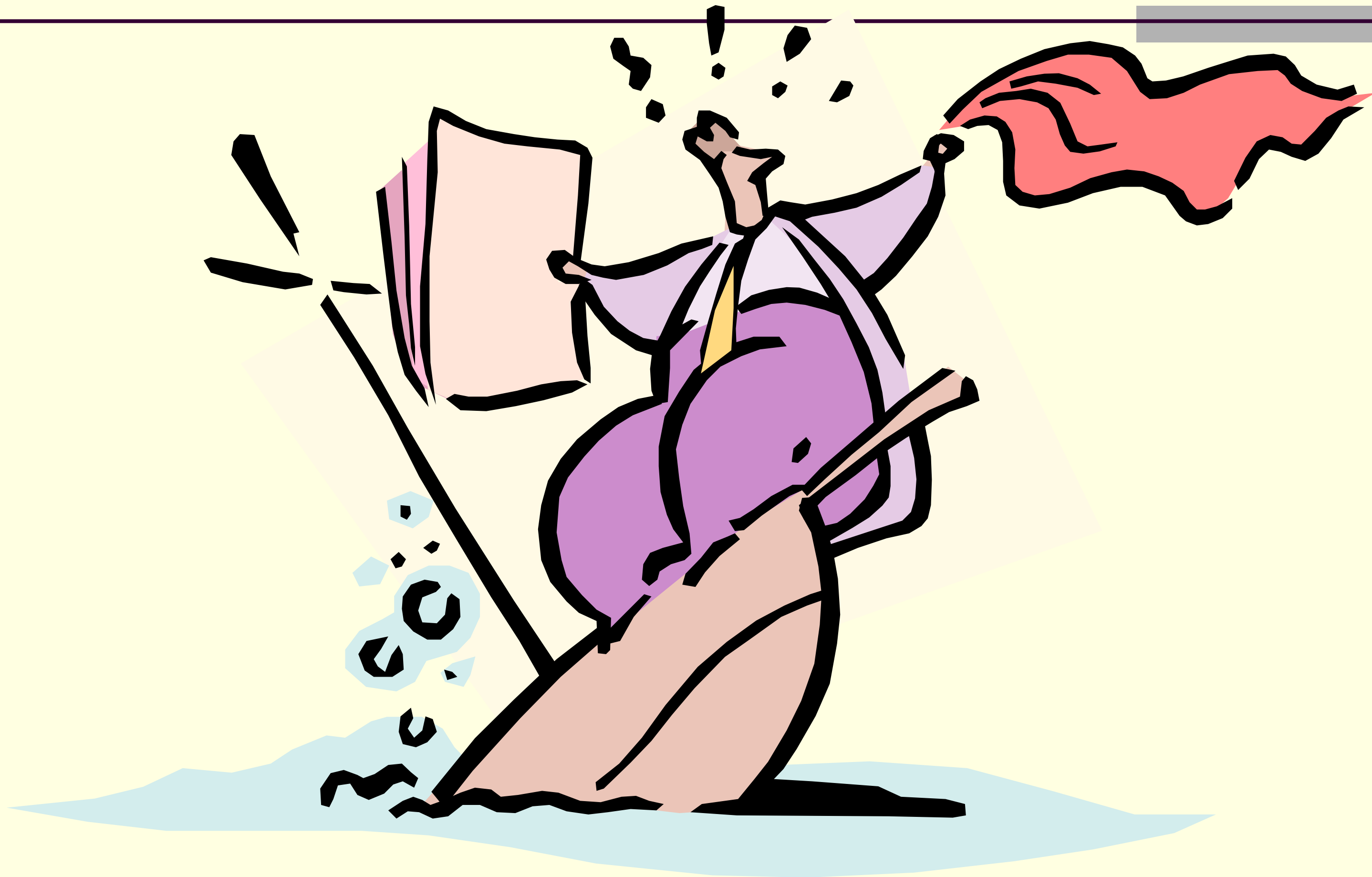
$$\left(\widehat{y}_{x_k} - t_\alpha \cdot m_{\widehat{y}_x}, \widehat{y}_{x_k} + t_\alpha \cdot m_{\widehat{y}_x}\right)$$

- Средняя ошибка прогнозируемого индивидуального значения y составит

$$m_{\widehat{y}(x_k)} = s \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{\sum\left(x - \bar{x}\right)^2}}$$

4. Оценка параметров





9. Прогноз срока службы ПВХ изоляции