# Тема. Парный регрессионный анализ

## Рекомендуемая литература:

1. **Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование:** Учеб.пособие – М.: Вузовский учебник, 2007, 2011.

2. **Эконометрика:** Учебник / Под ред. И.И.Елисеевой. - 2-е изд.; перераб. и доп. - М.: Финансы и статистика, 2005-2008.

3. **Кремер Н.Ш., Путко Б.А. Эконометрика:** Учебник для вузов / Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ – ДАНА, 2003-2008.

4. **Эконометрика. Методические указания** по изучению курса и выполнению контрольной работы. Для самостоятельной работы студентов 3 курса, обучающихся по направлению 521600 «Экономика» (бакалавр). – М.: ВЗФЭИ.

5. **Эконометрика. Компьютерный практикум** для студентов третьего курса, обучающихся по специальностям 080105.65 «Финансы и кредит», 080109.65 «Бухгалтерский учет, анализ и аудит». – М.: ВЗФЭИ, 2011.

# Вопросы:

1. Общие сведения о регрессионном анализе
2. Реализация основных этапов построения и анализа парной линейной регрессии
   - **2.1.** Оценка параметров. Определение вида модели
   - **2.2.** Проверка качества модели
   - **2.3.** Оценка статистической значимости уравнения и параметров
   - **2.4.** Экономический прогноз
3. Парная нелинейная регрессия
4. Причины ложных результатов регрессионного анализа

# 1. Общие сведения о регрессионном анализе

**Регрессионный анализ** предназначен для исследования количественных взаимосвязей переменных и представления их в виде **регрессионной модели**.

**Виды регрессий:**

1) по числу переменных:
   - парная,
   - множественная,
   - частная;
2) по виду связи переменных:
   - линейная,
   - нелинейная;
3) по направлению связей:
   - положительная,
   - отрицательная.

**Задачи регрессионного анализа:**

1. Установление формы связи, построение модели.
2. Оценка качества моделей.
3. Распределение факторов по степени влияния на показатель.
4. Построение прогноза.

# Общий вид регрессионной модели:

$$Y = f(X_1, X_2, ..., X_m) \qquad (1)$$

## парная линейная модель регрессии

$$Y_i = \alpha + \beta \cdot X_i + e_i \qquad (2)$$

$$\boxed{Y_i = Y_{ip} + e_i}$$

$$\boxed{e_i = Y_i - Y_{ip}}$$

$$Y_i = -5{,}79 + 36{,}84 \cdot X_i + e_i$$

# Основные этапы построения и анализа модели (2)

1. Оценка параметров. Определение вида модели.

2. Проверка качества модели.

3. Оценка статистической значимости уравнения и параметров.

4. Экономический прогноз.

# 2. Реализация основных этапов построения и анализа парной линейной регрессии

2.1. Оценка параметров. Определение вида модели

2.2. Проверка качества модели

2.3. Оценка статистической значимости уравнения и параметров

2.4. Экономический прогноз

# 2.1. Оценка параметров. Определение вида модели

**Мозговой штурм:**

- *Для чего применяется метод наименьших квадратов?*
- *Какая идея лежит в основе подбора параметров теоретической кривой?*
- *Что такое система нормальных уравнений?*
- *Как она выглядит при оценке параметров линейной модели?*

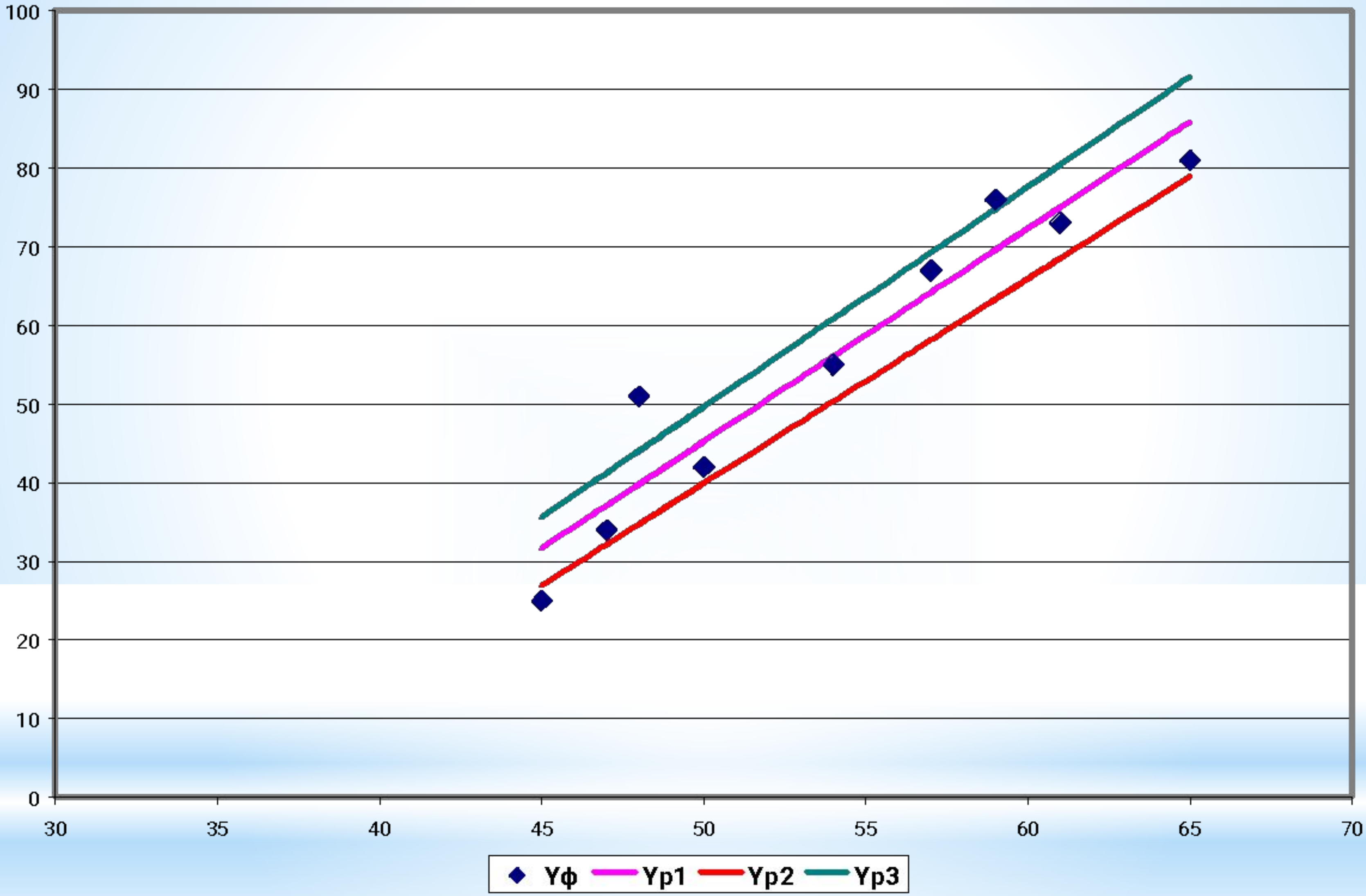



**Рис. 1. Расположение линии регрессии относительно фактических значений исследуемого показателя**

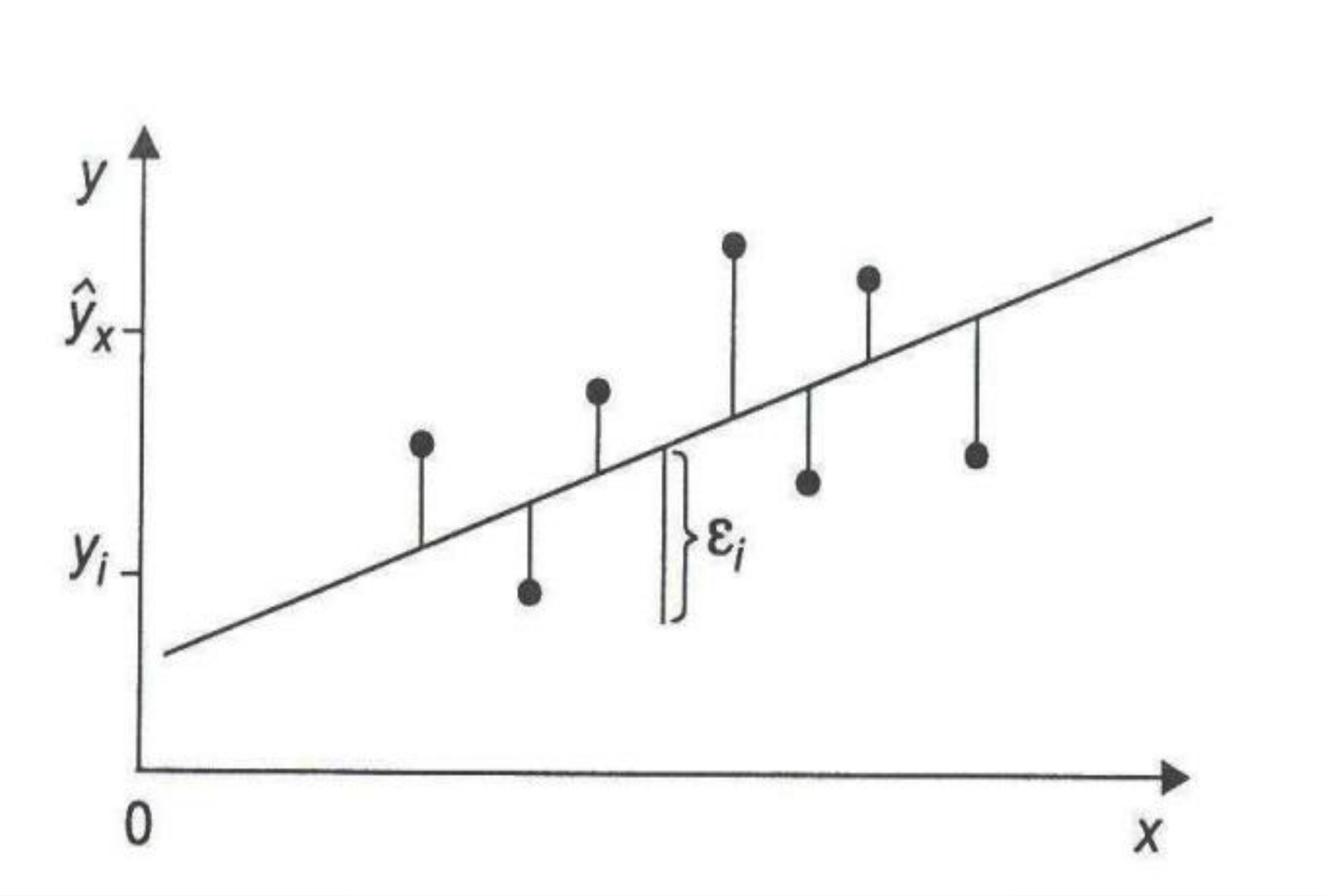



**Рис. 2. Линия регрессии с минимальными отклонениями от фактических данных**

# Метод наименьших квадратов (МНК)

$$Q(\alpha,\beta)=\sum_{i=1}^{n}(Y_{i\phi}-Y_{ip})^2=\sum_{i=1}^{n}(Y_i-\alpha-\beta\cdot X_i)^2\to\min$$

$$\frac{\partial Q(\alpha,\beta)}{\partial\alpha}=2\cdot\sum_{i=1}^{n}(Y_i-\alpha-\beta\cdot X_i)\cdot(-1)=0$$

$$\frac{\partial Q(\alpha,\beta)}{\partial\beta}=2\cdot\sum_{i=1}^{n}(Y_i-\alpha-\beta\cdot X_i)\cdot(-X_i)=0$$

$$\begin{cases}\beta\sum\limits_{i=1}^{n}X_i+\alpha n=\sum\limits_{i=1}^{n}Y_i.\\ \beta\sum\limits_{i=1}^{n}X_i^2+\alpha\sum\limits_{i=1}^{n}X_i=\sum\limits_{i=1}^{n}X_iY_i\end{cases}$$

$$\beta=\frac{\sum\limits_{i=1}^{n}X_i\cdot\sum\limits_{i=1}^{n}Y_i-n\cdot\sum\limits_{i=1}^{n}X_i\cdot Y_i}{\left(\sum\limits_{i=1}^{n}X_i\right)^2-n\cdot\sum\limits_{i=1}^{n}X_i^2}$$

$$\alpha=\frac{\sum\limits_{i=1}^{n}Y_i-\beta\cdot\sum\limits_{i=1}^{n}X_i}{n}=\overline{Y}-\beta\cdot\overline{X}.$$

# Уравнение в отклонениях

$$\overline{Y} = \alpha + \beta \cdot \overline{X} \qquad \left(\overline{X}, \overline{Y}\right)$$

$$\left(X_i - \overline{X}\right) \qquad \left(Y_i - \overline{Y}\right)$$

$$\beta = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y}) \cdot (X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}$$

$$\alpha = \overline{Y} - \beta \cdot \overline{X}.$$

# Условие идентифицируемости

$$\sum_{i-1}^{n}(X_i-\overline{X})^2 \neq 0.$$

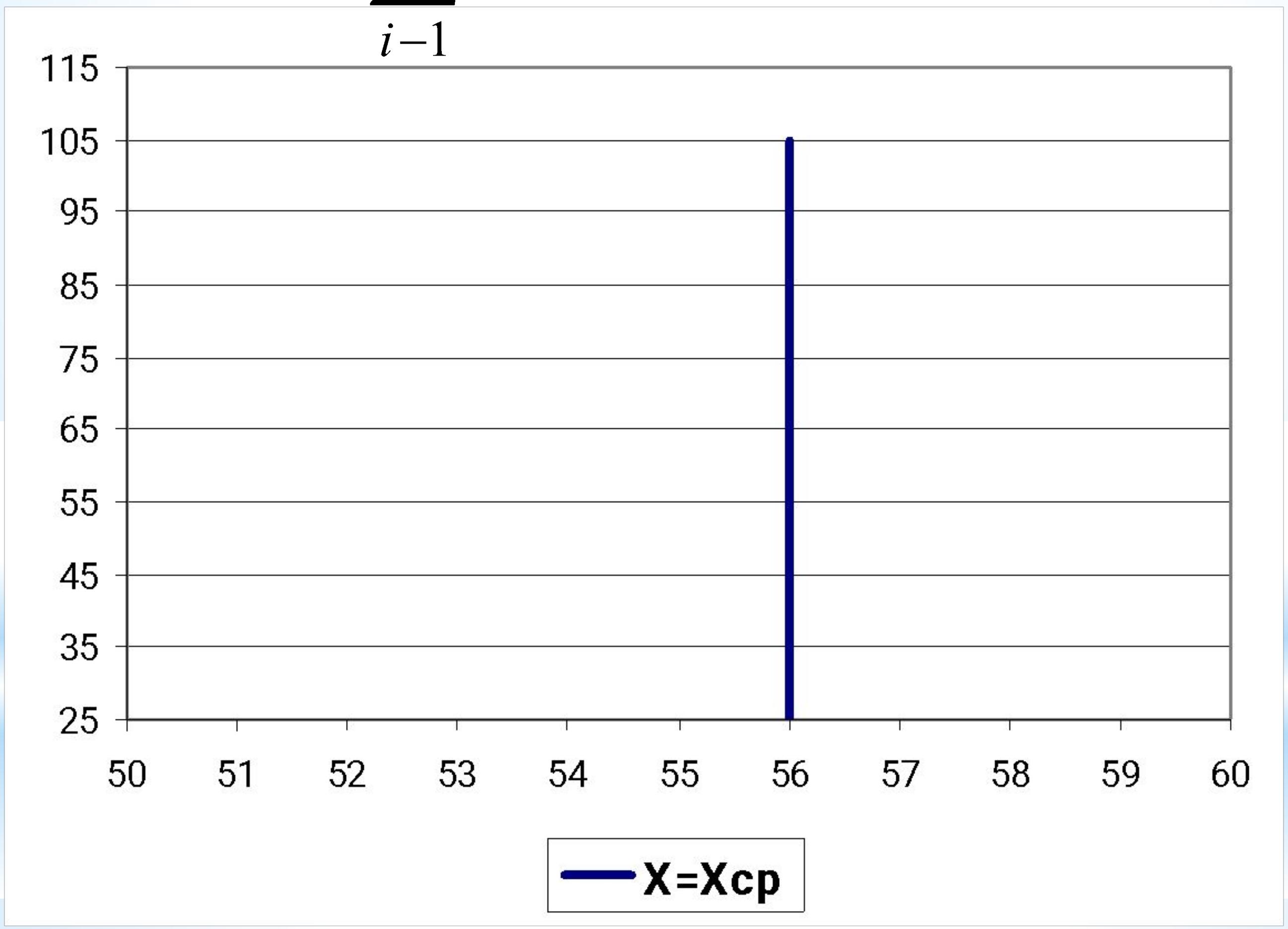



**Рис. 3. Условие идентифицируемости не выполняется**

# Пример

Пусть зависимая пере-менная Y – квартальная прибыль девяти компаний одной отрасли (в млн. руб.), а фактор X – объем продаж товара этих компаний за квартал (в тыс. шт.).

| № п. п. | Y | X |
|---|---|---|
| 1 | 25 | 45 |
| 2 | 34 | 47 |
| 3 | 42 | 50 |
| 4 | 51 | 48 |
| 5 | 55 | 54 |
| 6 | 67 | 57 |
| 7 | 73 | 61 |
| 8 | 76 | 59 |
| 9 | 81 | 65 |

1. Найти параметры уравнения линейной регрессии, дать экономическую интерпретацию коэффициента регрессии

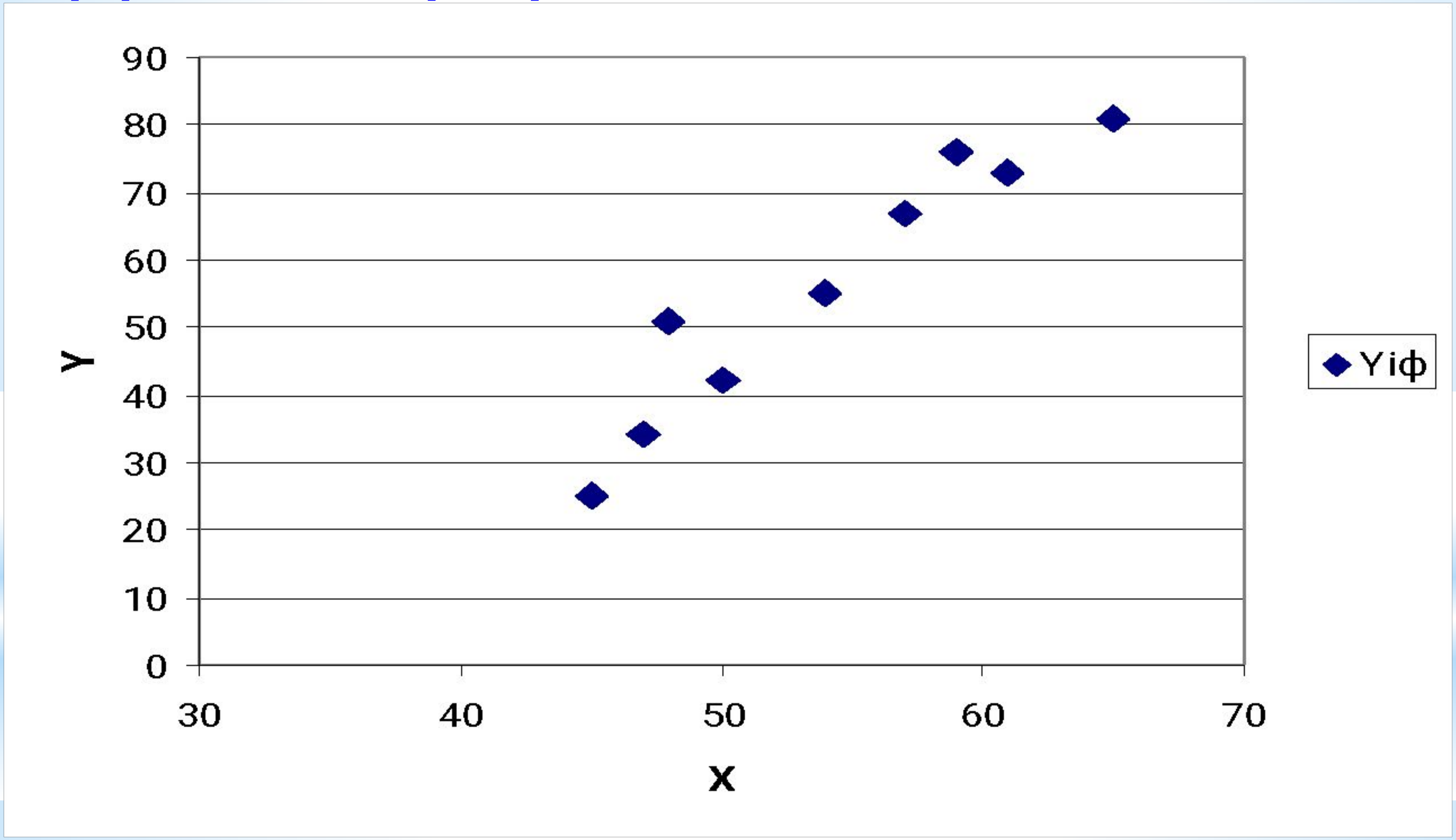



Рис. 4. Диаграмма рассеяния исходных данных

# Расчетная таблица

| № п.п. | Yi | X | X-Xср | (X-Xср)^2 | Y-Yср | (X-Xср)*(Y-Yср) | Yiр | $e_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 25 | 45 | -9 | 81 | -31 | 279 | 31.65 | -6.65 |
| 2 | 34 | 47 | -7 | 49 | -22 | 154 | 37.07 | -3.07 |
| 3 | 42 | 50 | -4 | 16 | -14 | 56 | 45.2 | -3.2 |
| 4 | 51 | 48 | -6 | 36 | -5 | 30 | 39.78 | 11.22 |
| 5 | 55 | 54 | 0 | 0 | -1 | 0 | 56.04 | -1.04 |
| 6 | 67 | 57 | 3 | 9 | 11 | 33 | 64.17 | 2.83 |
| 7 | 73 | 61 | 7 | 49 | 17 | 119 | 75.01 | -2.01 |
| 8 | 76 | 59 | 5 | 25 | 20 | 100 | 69.59 | 6.41 |
| 9 | 81 | 65 | 11 | 121 | 25 | 275 | 85.85 | -4.85 |
| **Сумма** | **504** | **486** | | **386** | | **1046** | | **-0.36** |
| **Среднее** | **56** | **54** | | | | | | |

$$\beta = \frac{1046}{386} = 2,71$$

$$\alpha = 56 - 2,71 * 54 = -90,3$$

# Уравнение регрессии

$$Yi = -90.3 + 2.71 * Xi + ei$$

$$Yip = -90.3 + 2.71 * Xi$$

$$e_i = Y_{iф} - Y_{ip}$$

**Экономический смысл коэффициента регрессии**:

при изменении объема продаж компании (X) на 1 тысячу штук прибыль (Y) будет меняться в ту же сторону на 2,71 млн. руб.

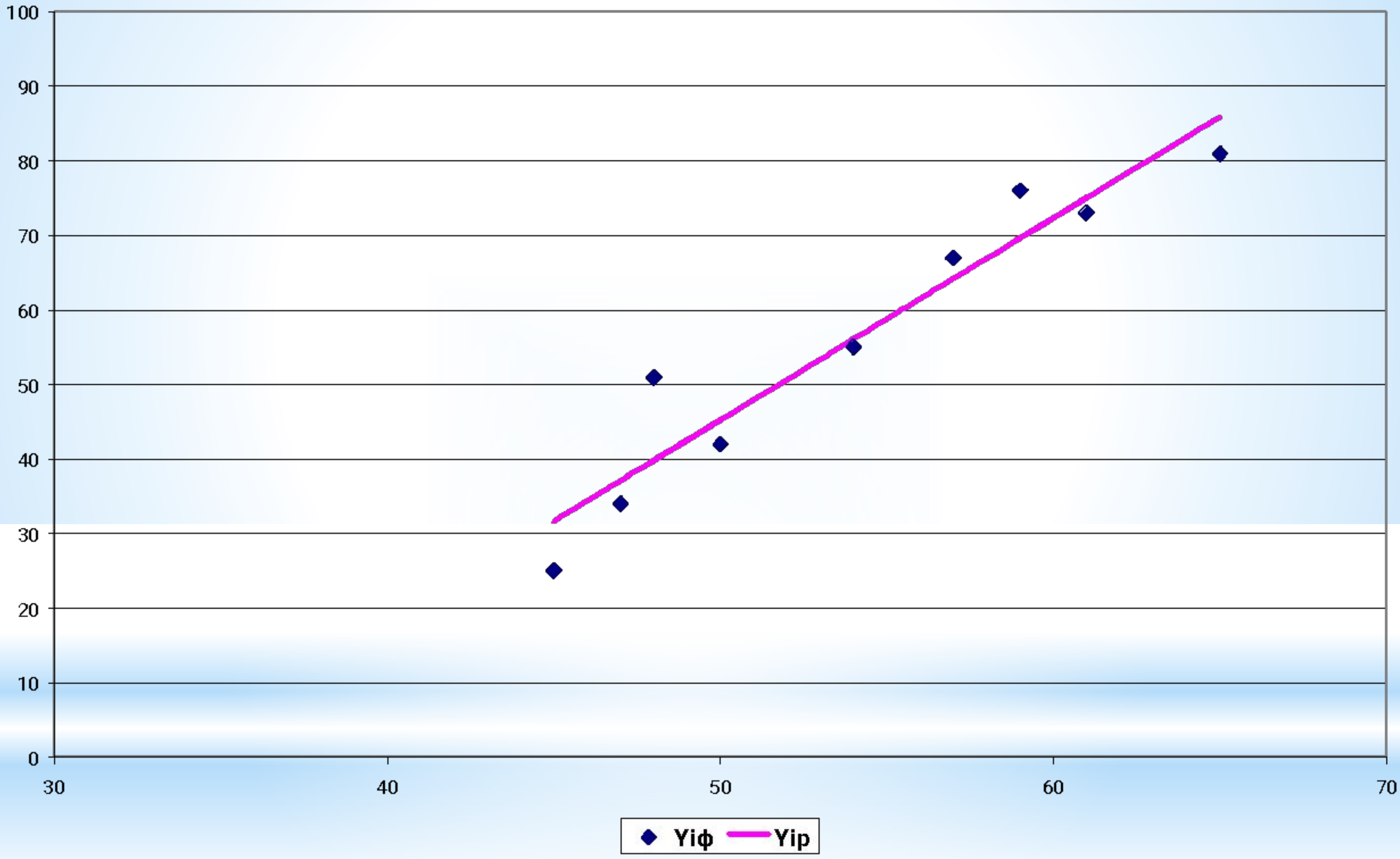



**Рис. 5. Результаты приближения фактических значений прибыли линией регрессии**

# Результаты работы с инструментом Регрессия

| | *Коэффициенты* |
|---|---|
| **Y** | **-90.3** |
| **X** | **2.71** |

| *Наблюдение* | *Предсказанное Y* | *Остатки* |
|---|---|---|
| 1 | 31.61 | -6.61 |
| 2 | 37.03 | -3.03 |
| 3 | 45.16 | -3.16 |
| 4 | 39.74 | 11.26 |
| 5 | 56.00 | -1.00 |
| 6 | 64.13 | 2.87 |
| 7 | 74.97 | -1.97 |
| 8 | 69.55 | 6.45 |
| 9 | 85.81 | -4.81 |

# Анализ вариации зависимой переменной в уравнении регрессии

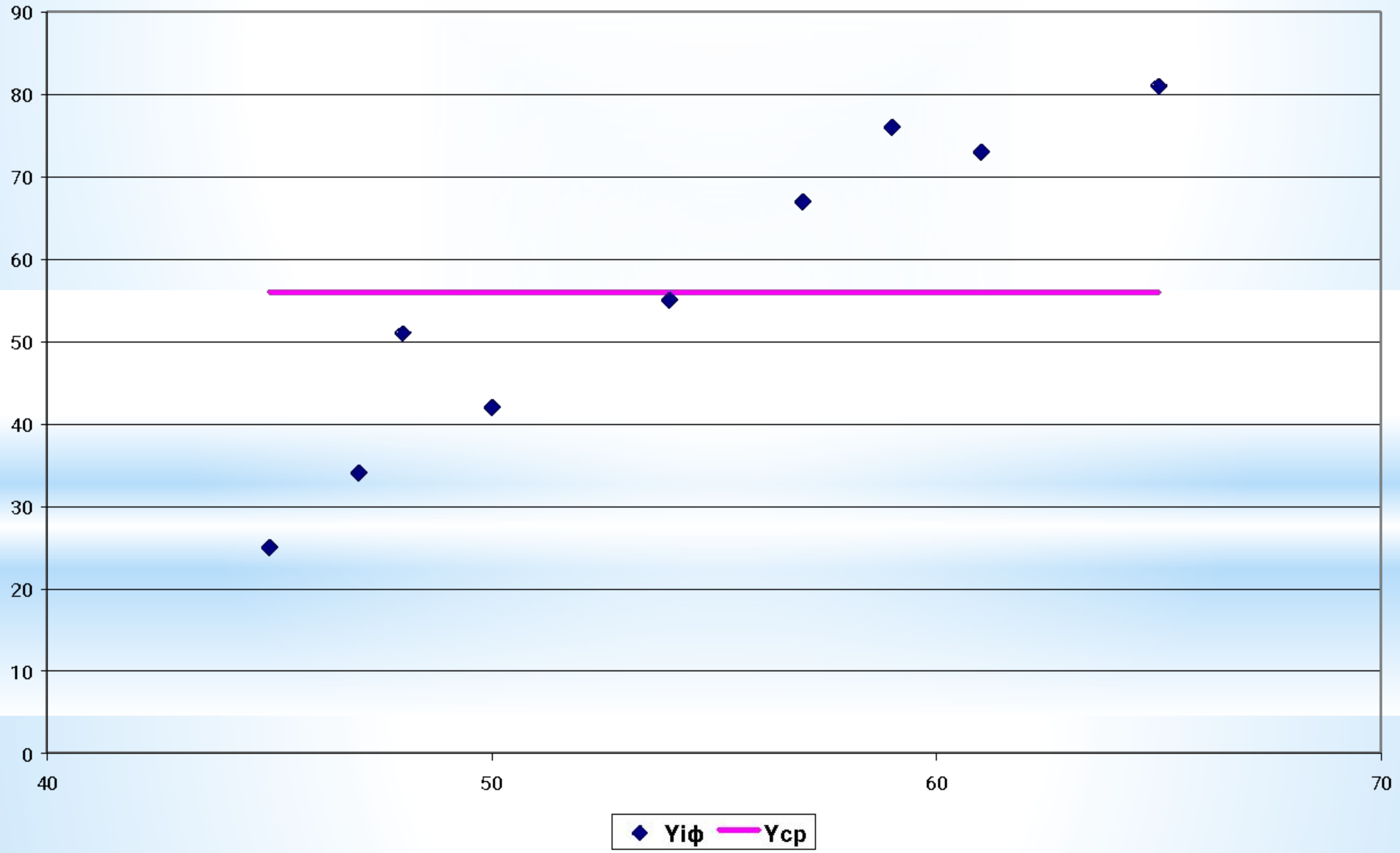
$$Y_i - \overline{Y} = (Y_i - Y_{ip}) + \left(Y_{ip} - \overline{Y}\right)$$

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}\left(Y_i - Y_{ip}\right)^2 + \sum_{i=1}^{n}\left(Y_{ip} - \overline{Y}\right)^2 + 2\cdot\sum_{i=1}^{n}\left(Y_i - Y_{ip}\right)\cdot\left(Y_{ip} - \overline{Y}\right)$$

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}\left(Y_i - Y_{ip}\right)^2 + \sum_{i=1}^{n}\left(Y_{ip} - \overline{Y}\right)^2$$

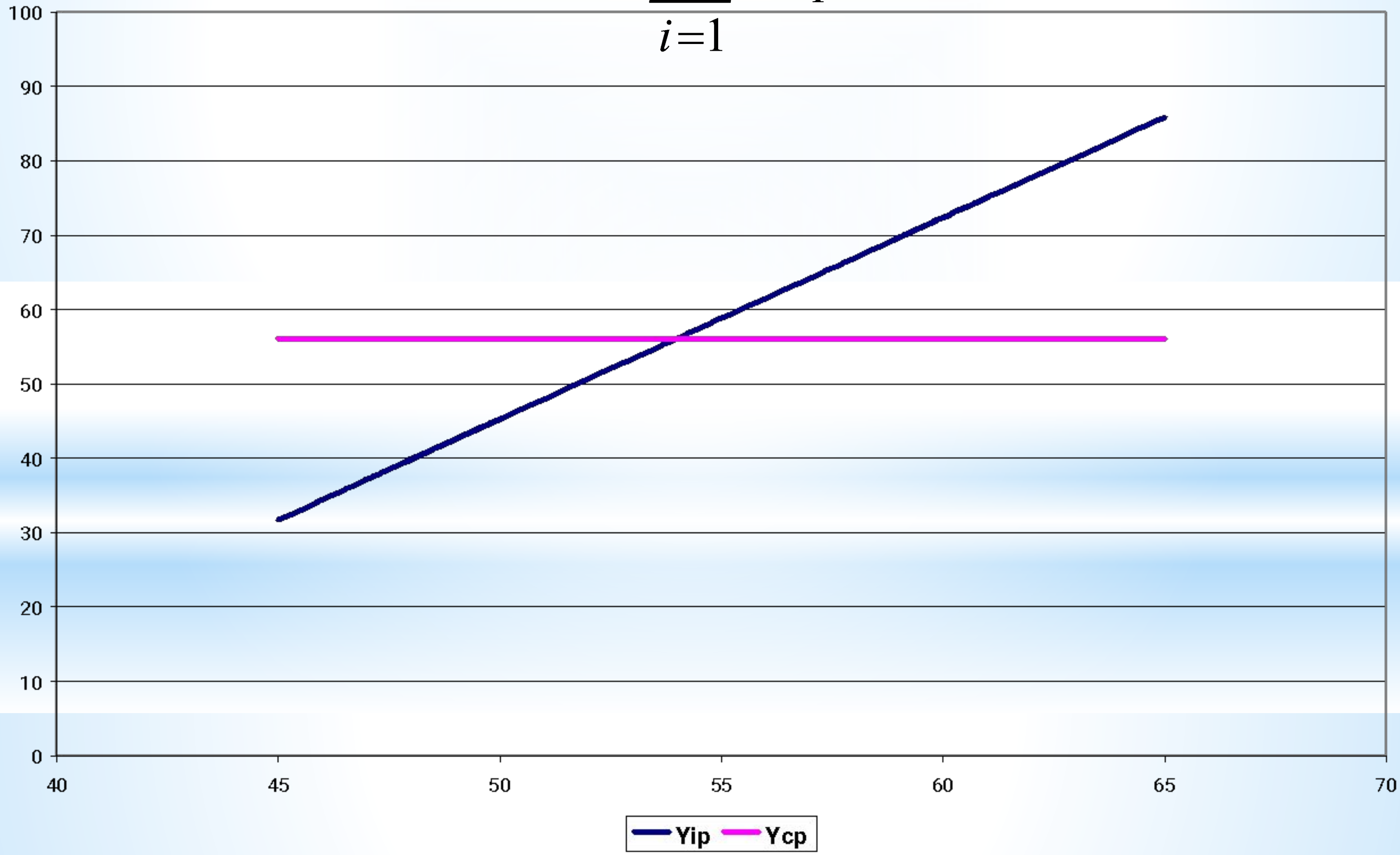
# Общая сумма квадратов отклонений

$$TSS = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$



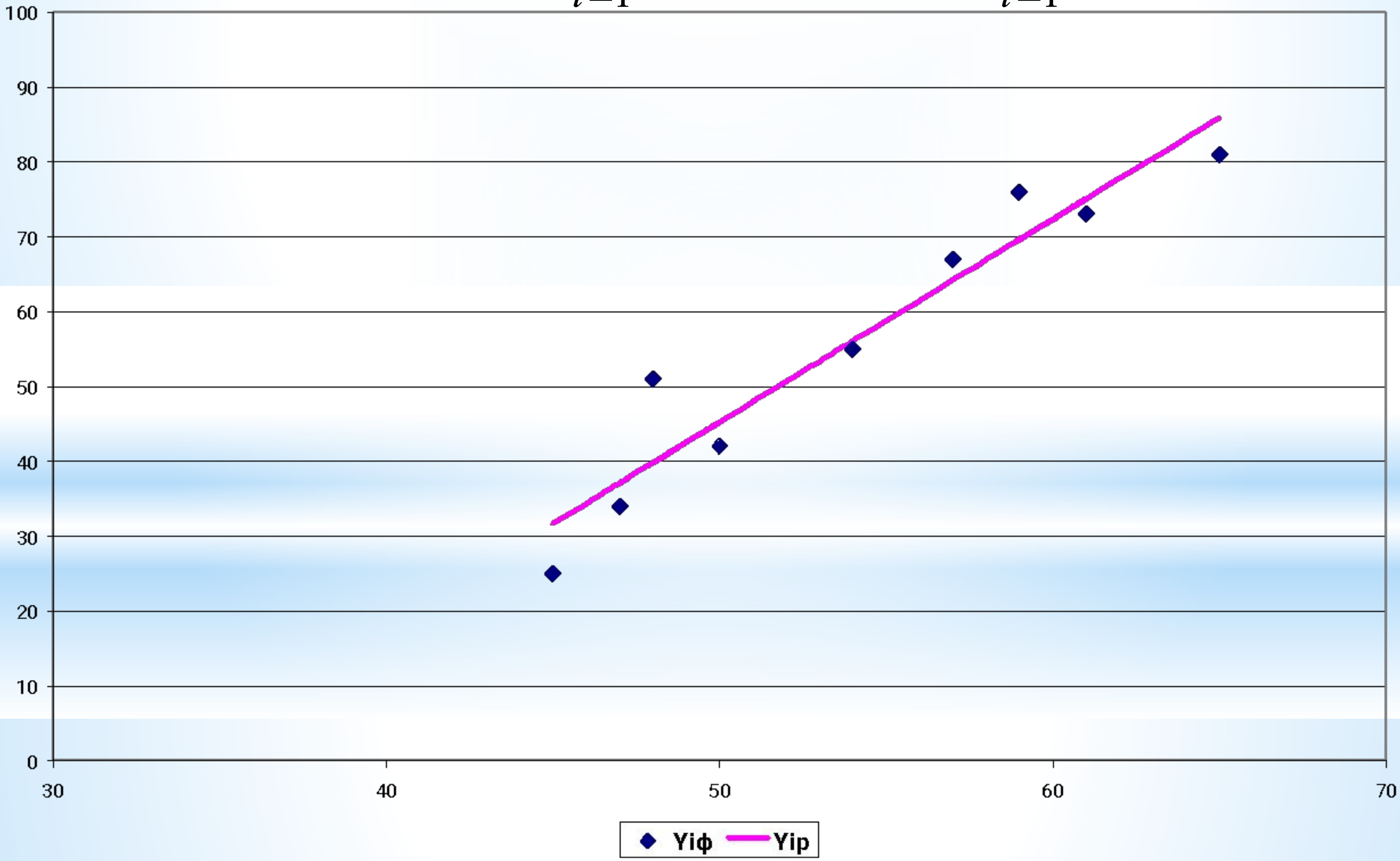
# Сумма квадратов отклонений, объясненная регрессией

$$\mathbf{RSS} = \sum_{i=1}^{n}\left(Y_{ip} - \overline{Y}\right)^2$$



# Остаточная сумма квадратов отклонений

$$\textbf{ESS} = \sum_{i=1}^{n}\left(Y_i - Y_{ip}\right)^2 = \sum_{i=1}^{n} e_i^2$$



# Средний квадрат отклонений или дисперсия на одну степень свободы

$$n - 1 = 1 + (n - 2)$$

$$S^2_{общ} = \frac{1}{n-1} \cdot \sum_i (Y_i - \overline{Y})^2$$

$$S^2_{объясн(факторн)} = \frac{1}{1} \cdot \sum_i (Y_{ip} - \overline{Y})^2$$

$$S^2_e = \frac{1}{n-2} \cdot \sum_i (Y_i - Y_{ip})^2 = \frac{1}{n-2} \cdot \sum_i e_i^2$$

# Дисперсионный анализ

| | *df* | *SS* | *MS* |
|---|---|---|---|
| Регрессия | 1 | 2834.5 | 2834.50 |
| Остаток | 7 | 267.50 | 38.21 |
| Итого | 8 | 3102.0 | |

# 2.2. Проверка качества модели

**Свойства оценок МНК**

**Несмещенность**

**Состоятельность**

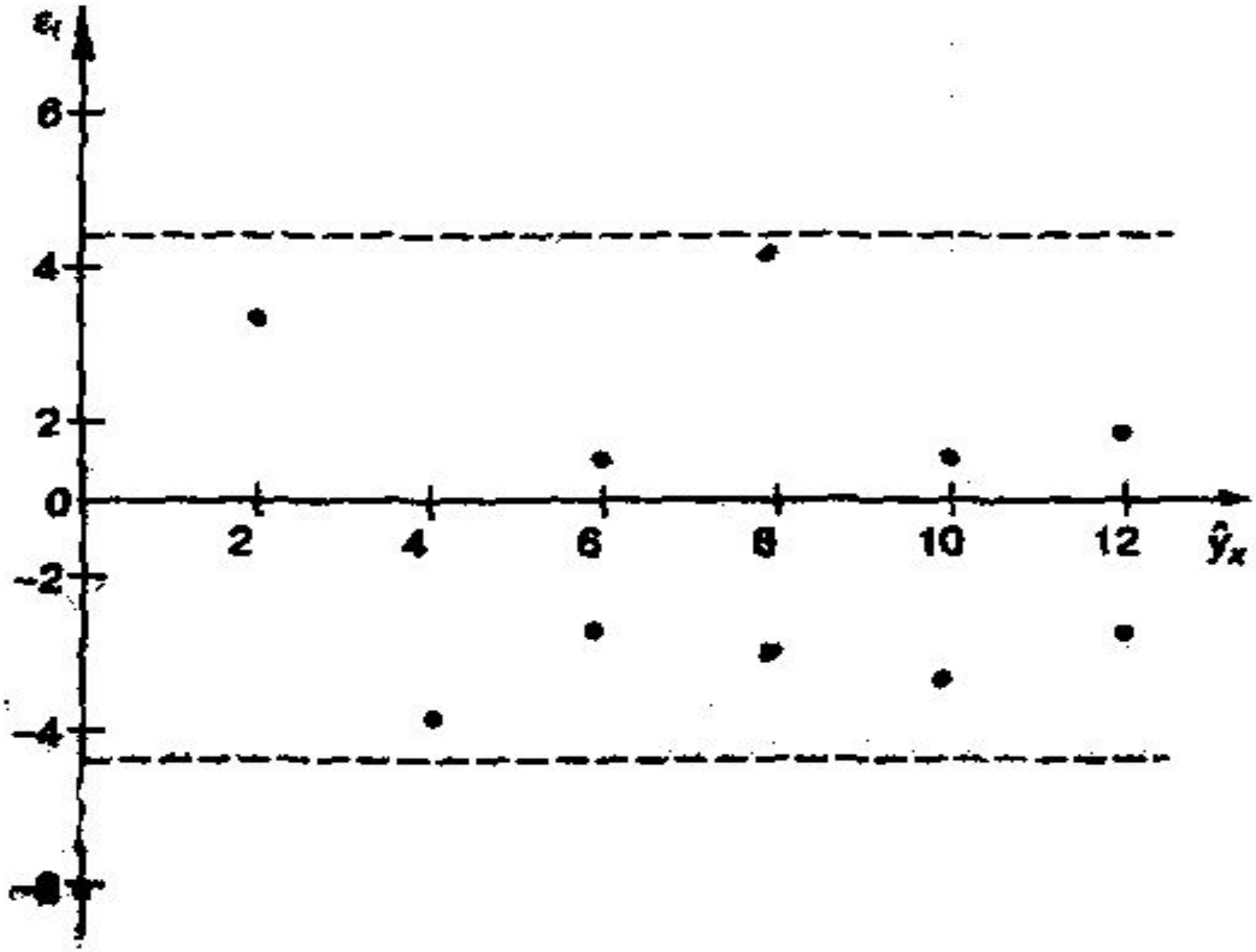
**Эффективность**

**Теорема Гаусса-Маркова**

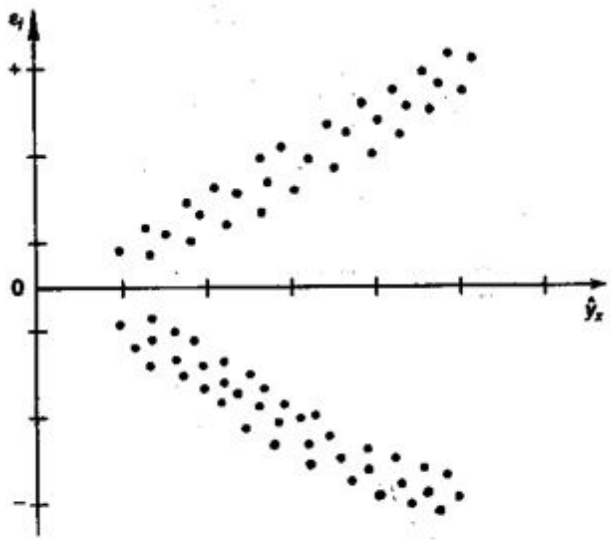
**Пять предпосылок МНК:**

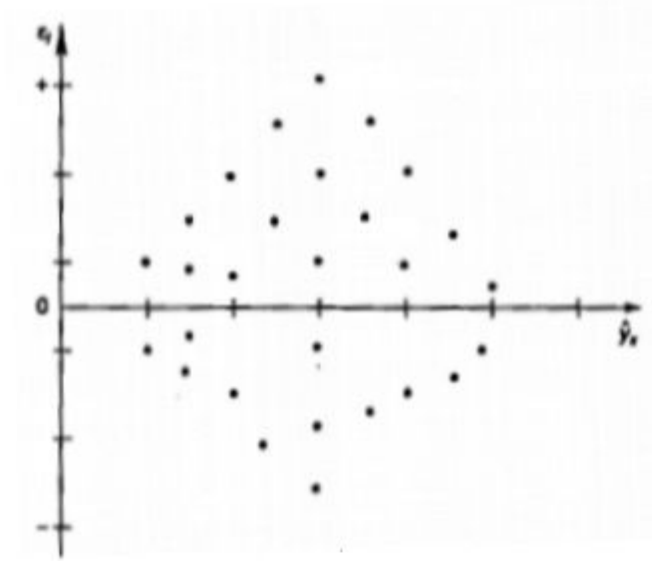
1) **случайный характер** остатков;
2) **независимость** остатков или отсутствие их автокорреляции;
3) остатки подчиняются **нормальному распределению**;
4) **нулевая средняя** величина остатков (или их математическое ожидание), не зависящая от уровней фактора X;
5) ***гомоскедастичность*** остатков

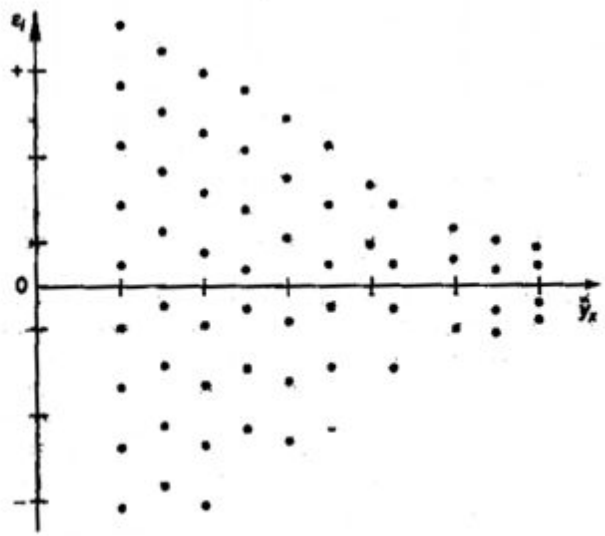
# Гомоскедастичность



# Гетероскедастичность







# Метод Гольдфельда – Квандта

1. Упорядочение n наблюдений по мере возрастания переменной X.
2. Исключение из рассмотрения C центральных наблюдений; при этом (n-C):2>p, где p – число оцениваемых параметров $C \approx \frac{1}{4}n$
3. Разделение совокупности из (n-C) наблюдений на две группы (соответственно с малыми и большими значениями фактора X) и определение по каждой из групп уравнений регрессии.
4. Определение остаточной суммы квадратов для первой (S1) и второй (S2) групп и нахождение их отношения: R=S1:S2 (в числителе должна стоять большая величина).
5. F-критерий Фишера с (n-C-2p):2 степенями свободы.

# Пример (продолжение)

| Y | X |
|---|---|
| 25 | 45 |
| 34 | 47 |
| 51 | 48 |
| 42 | 50 |
| 55 | 54 |
| 67 | 57 |
| 76 | 59 |
| 73 | 61 |
| 81 | 65 |

**Для первой совокупности**:

**Дисперсионный анализ**

| | *df* | *SS* | *MS* |
|---|---|---|---|
| Регрессия | 1 | 200.0769 | 200.0769 |
| Остаток | 2 | **169.9231** | 84.96154 |
| Итого | 3 | 370 | |

**Для второй совокупности**:

**Дисперсионный анализ**

| | *df* | *SS* | *MS* |
|---|---|---|---|
| Регрессия | 1 | 78.75 | 78.75 |
| Остаток | 2 | **24** | 12 |
| Итого | 3 | 102.75 | |

R= 169,9/24 = 7.08

Число степеней свободы: (9-1-2*2):2=2

Fтаб(0,05;2;2)=19

$$F_{табл} > R$$

# Характеристики качества

**индекс корреляции**

$$R = \sqrt{\frac{RSS}{TSS}} = \sqrt{1 - \frac{ESS}{TSS}} = \sqrt{1 - \frac{\sum_i e_i^2}{\sum_i (Y_i - \overline{Y})^2}}$$

$$R = \left| r_{X,Y} \right|$$

**коэффициент детерминации**

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum_i e_i^2}{\sum_i (Y_i - \overline{Y})^2}$$

**средняя относительная ошибка аппроксимации**

$$\overline{e}_{отн.} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| e_i \right|}{y_i} \cdot 100\%$$

**стандартная ошибка модели (среднеквадратическое (отклонение)**

$$S_e = \sqrt{S_e^2}$$

# Пример (продолжение)

| № пп | Y | Y-Ycp | e | e^2 | eотн | (Y-Ycp)^2 | X^2 |
|---|---|---|---|---|---|---|---|
| 1 | 25 | -31 | -6.65 | 44.22 | 26.6 | 961 | 2025 |
| 2 | 34 | -22 | -3.07 | 9.42 | 9.03 | 484 | 2209 |
| 3 | 42 | -14 | -3.2 | 10.24 | 7.62 | 196 | 2500 |
| 4 | 51 | -5 | 11.22 | 125.89 | 22 | 25 | 2304 |
| 5 | 55 | -1 | -1.04 | 1.08 | 1.89 | 1 | 2916 |
| 6 | 67 | 11 | 2.83 | 8.01 | 4.22 | 121 | 3249 |
| 7 | 73 | 17 | -2.01 | 4.04 | 2.75 | 289 | 3721 |
| 8 | 76 | 20 | 6.41 | 41.09 | 8.43 | 400 | 3481 |
| 9 | 81 | 25 | -4.85 | 23.52 | 5.99 | 625 | 4225 |
| **Сумма** | **504** | | **-0.36** | **267.52** | **88.54** | **3102** | **26630** |
| **Среднее** | **56** | | | | **9.84** | | |

$$R^2 = 1 - \frac{267.5}{3102} = 0.914 \to 1$$

$$\boxed{\overline{e}_{отн} = 9,8\,\%}$$

$$1 - R^2 = 1 - 0.914 = 0.086$$

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0.956 |
| **R-квадрат** | **0.914** |
| Нормированный R-квадрат | 0.901 |
| **Стандартная ошибка** | **6.182** |
| Наблюдения | 9.000 |

$$S_Y = 19.7$$

$$S_e < S_Y$$

# 2.3. Проверка статистической значимости уравнения регрессии и его параметров

а) **проверка статистической значимости уравнения:**

**F-критерий Фишера**:

$$F = \frac{S^2_{факторн}}{S^2_e} \qquad F = \frac{R^2 / k}{(1 - R^2)/(n - k - 1)}$$

$$F_{табл} < F \qquad F_{табл}(\alpha, k_1, k_2)$$

# Пример (продолжение)

$$F = \frac{0.914/1}{(1-0.914)/(9-1-1)} = 74.4$$

$$F_{табл}(0.05,\ 1,\ 9-1-1) = 5.59$$

$$\boxed{F_{табл} << F}$$

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Регрессия | 1 | 2834.50 | 2834.50 | **74.2** |
| Остаток | 7 | 267.50 | 38.21 | |
| Итого | 8 | 3102.00 | | |

# а) проверка статистической значимости параметров уравнения:

**t-критерий Стьюдента**

$$S_\beta = \sqrt{\frac{S_e^2}{\sum_i (X_i - \overline{X})^2}}$$

$$\boxed{S_e^2 = \frac{\sum_i e_i^2}{n-k-1}}$$

$$S_\alpha = \sqrt{\frac{S_e^2 \cdot \sum_i X_i^2}{n \cdot \sum_i (X_i - \overline{X})^2}}$$

$$t_\alpha = \frac{\alpha}{S_\alpha} \qquad t_\beta = \frac{\beta}{S_\beta}$$

$$t_{табл} < |t_\alpha| \qquad \boxed{t_{табл}(\alpha,\ \mathrm{n-k-1})} \qquad t_{табл} < |t_\beta|$$

**интервальные оценки параметров**

$$\beta \in \left(\beta \pm t_{табл} \cdot S_\beta\right) \qquad \alpha \in \left(\alpha \pm t_{табл} \cdot S_\alpha\right)$$

# Пример (продолжение)

$$S_e^2 = \frac{267.5}{9-1-1} = 38.2$$

$$S_\beta = \sqrt{\frac{38.2}{386}} = 0.315$$

$$S_\alpha = \sqrt{\frac{38.2 \cdot 26630}{9 \cdot 386}} = 17.1$$

$$t_\beta = \frac{2.71}{0.315} = 8.74$$

$$t_\alpha = \frac{-90.3}{17.1} = -5.28$$

$$t_{табл}(0{,}05;9-1-1) = 2.36$$

$$t_{табл} < t_\beta$$

$$t_{табл} < |t_\alpha|$$

$$\alpha \in (-90.3 \pm 2.36 \cdot 17.1) = (-130.8;-49.86)$$

$$\beta \in (2.71 \pm 2.36 \cdot 0.315) = (1.97;3.45)$$

| | *Коэффи-циенты* | ***Стан-дартная ошибка*** | ***t-статис тика*** | *P-Значе ние* | ***Ниж-ние 95%*** | ***Верх-ние 95%*** |
|---|---|---|---|---|---|---|
| Y | -90.33 | **17.12** | **-5.28** | 0.00 | **-130.80** | **-49.86** |
| X | 2.71 | **0.31** | **8.61** | 0.00 | **1.97** | **3.45** |

# 2.4. Экономический прогноз

**Три основных этапа:**

1) точечный прогноз фактора X;
2) точечный прогноз показателя Y;
3) интервальный прогноз показателя Y

1) $$X_t = f(t)$$

$$X_{n+k} = f(n+k)$$

$$X_{n+k} = X_n + k \cdot САП$$

$$САП = \frac{X_n - X_1}{n-1}$$

$$X_{пр} = 0{,}8 \cdot \overline{X}$$

2) $$Y_{пр} = \alpha + \beta \cdot X_{пр} \qquad Y_{n+k} = \alpha + \beta \cdot X_{n+k}$$

3) $$U = S_e \cdot t_{табл} \cdot \sqrt{1 + \frac{1}{n} + (X_{пр} - \overline{X})^2 \div \sum_i (X_i - \overline{X})^2}$$

$$Y_{пр} - U \qquad Y_{пр} + U$$

# Пример (продолжение)

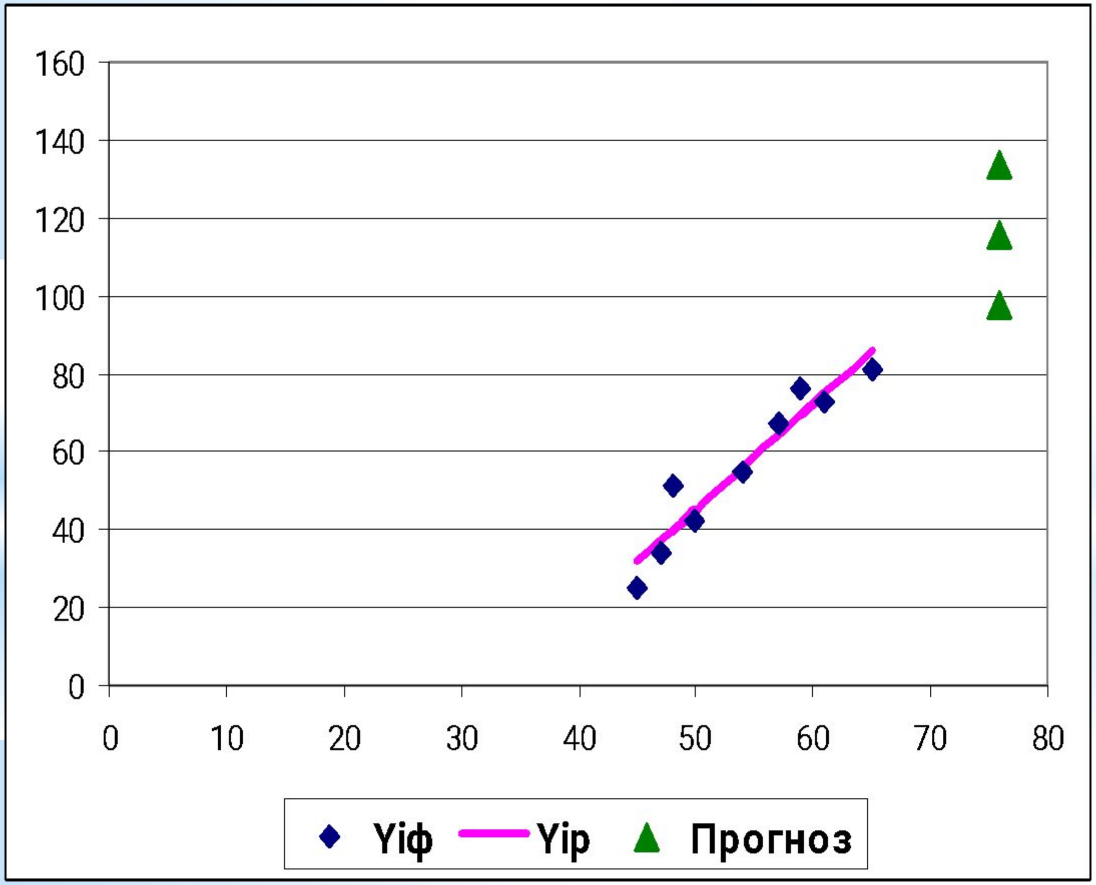
$$X_{пр} = 1.17 \cdot 65 = 76$$

$$Y_{пр} = -90{,}3 + 2{,}71 \cdot 76 = 115.66$$

$$t_{табл}(0{,}1;7) = 1.89$$

$$S_e = 6.18$$

$$U = 6.18 \cdot 1.89 \cdot \sqrt{1 + \frac{1}{9} + (76-54)^2 \div 386} = 17.97$$




**Нижняя граница интервала**:

115,66-17,97=97,69

**Верхняя граница интервала**:

115,66+17,97=133,63

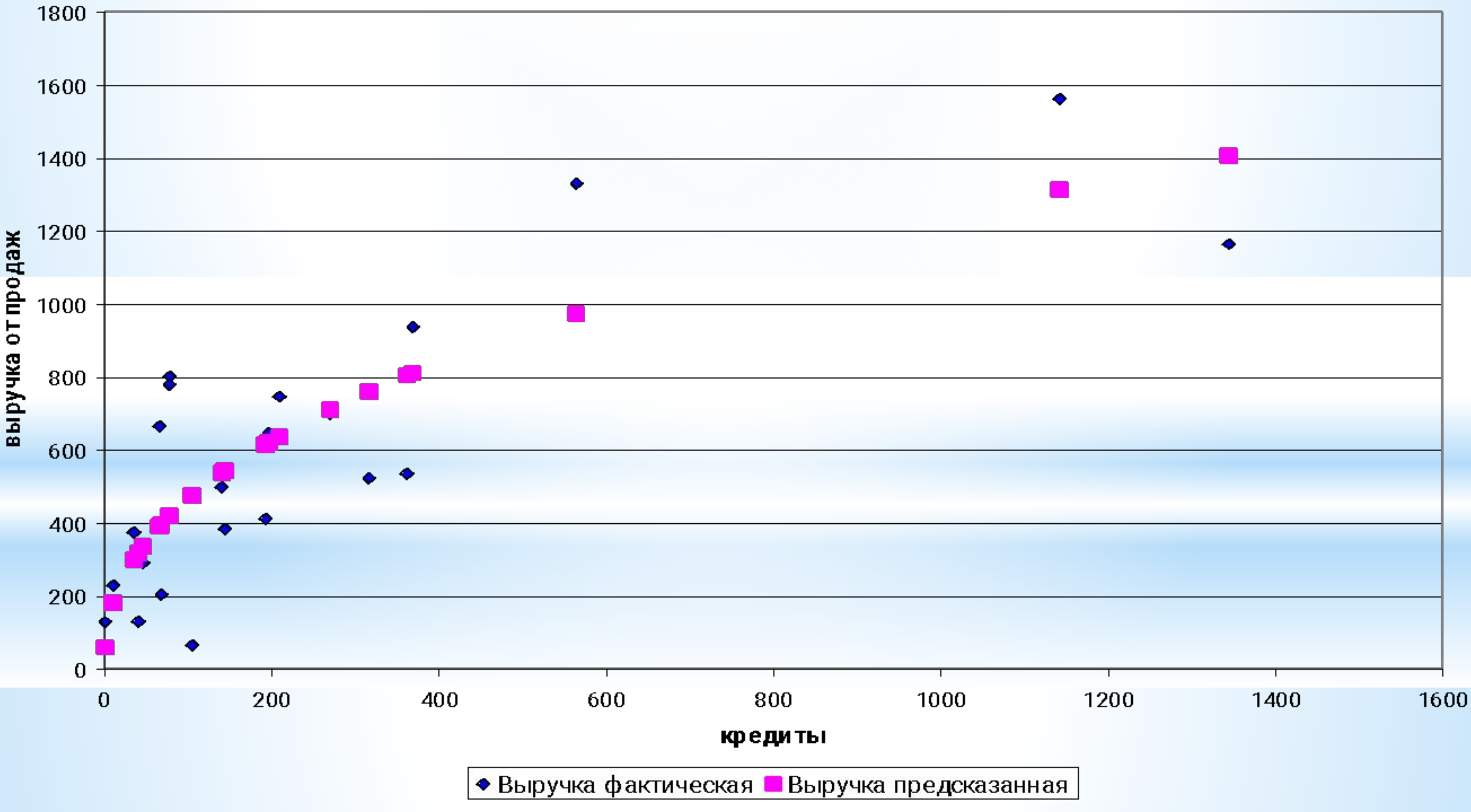
# 3. Парная нелинейная регрессия

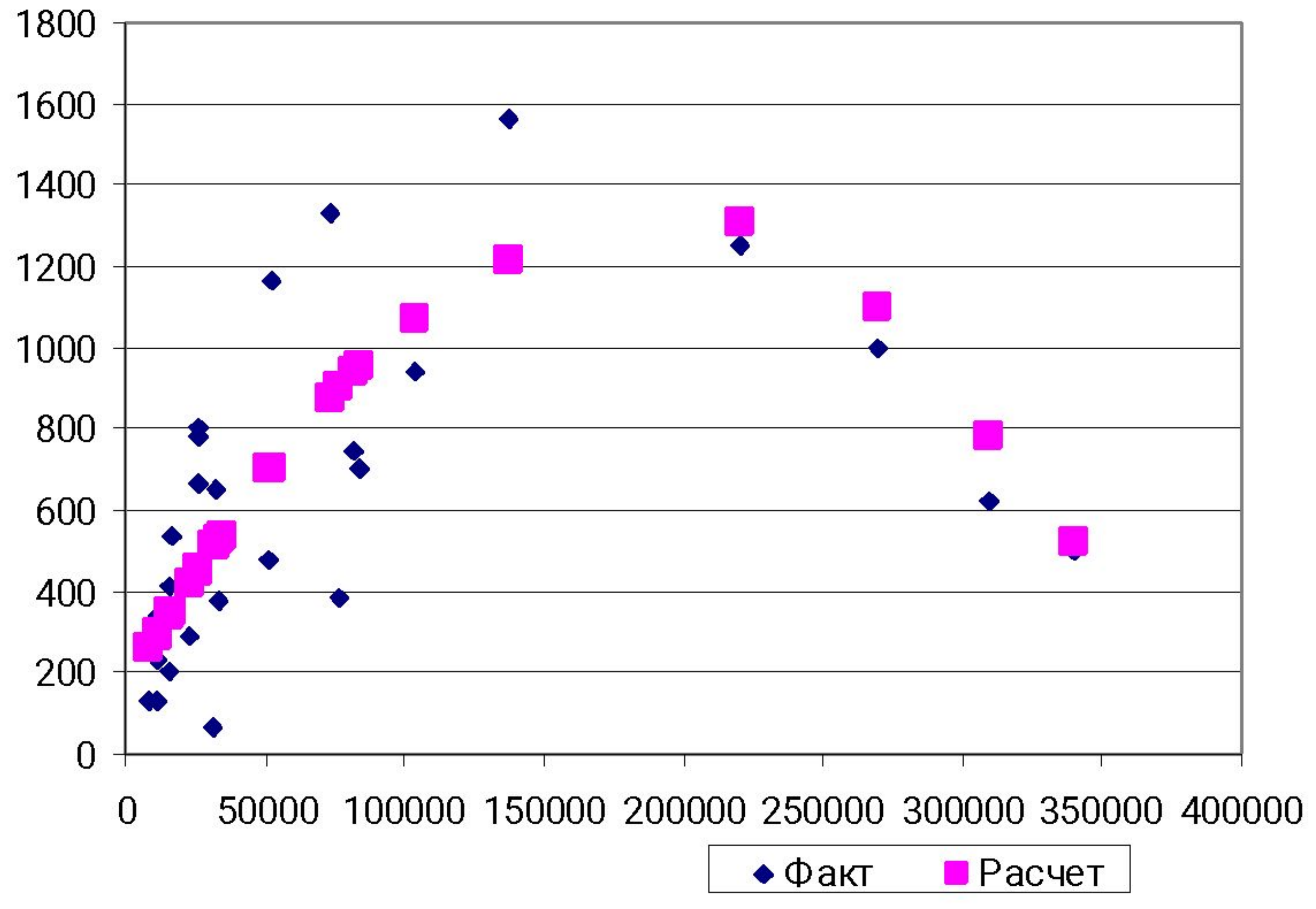
**Общий вид регрессионной модели**

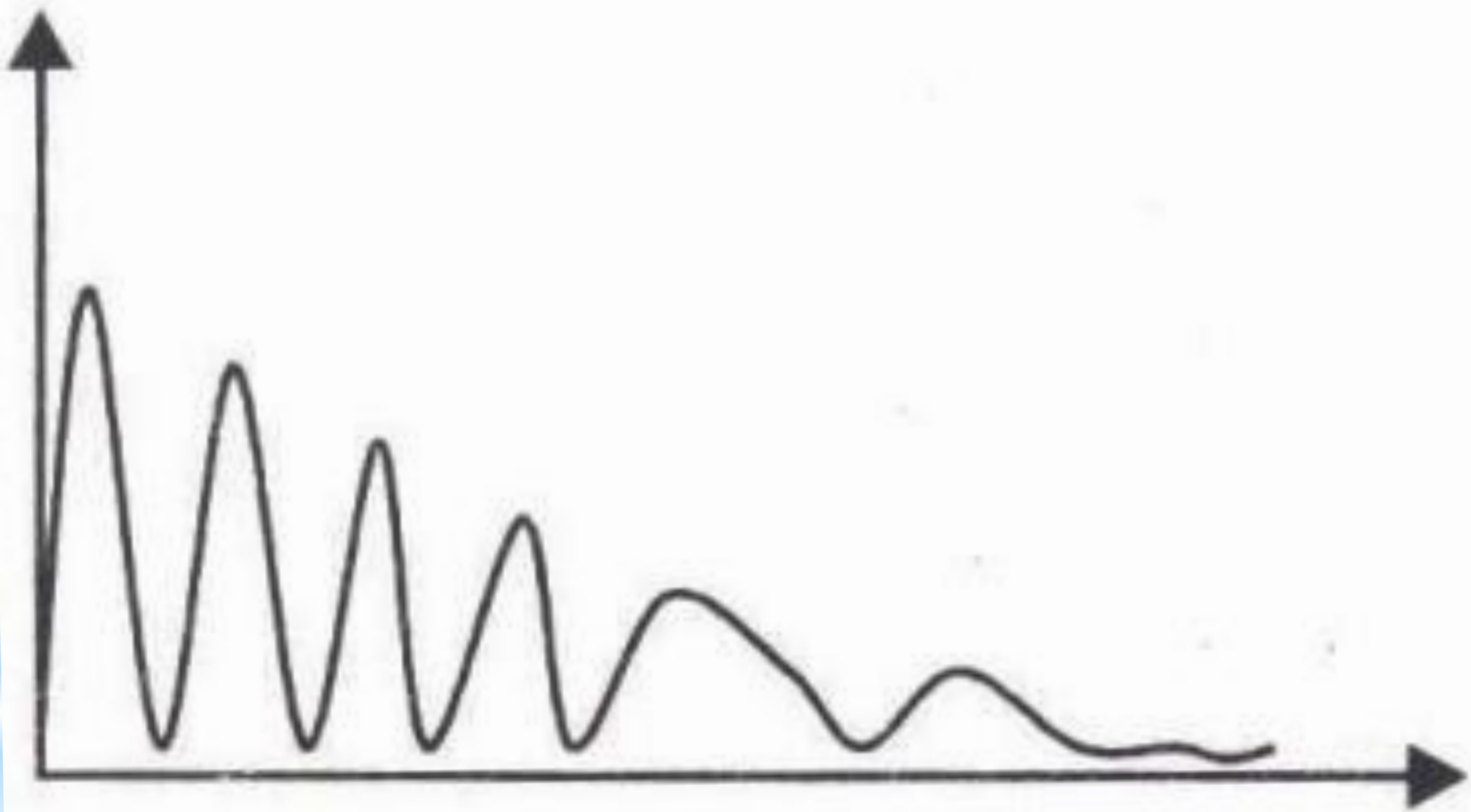
$$Y = f(X_1, X_2, ..., X_m)$$

**Парная нелинейная регрессия**

**Y=f(X)**




1800
1600
1400
1200
1000
800
600
400
200
0
0
50000
100000
150000
200000
250000
300000
350000
400000
Факт
Расчет



К *первому классу* относятся:

1) полиномы разных степеней

$$Y = a + b \cdot X + c \cdot X^2 + \ldots + e$$

2) равносторонняя гипербола

$$Y = a + \frac{b}{X} + e$$

Ко *второму классу* относятся:

1) степенная функция

$$Y = a \cdot X^b \cdot \varepsilon$$

2) показательная

$$Y = a \cdot b^X \cdot \varepsilon$$

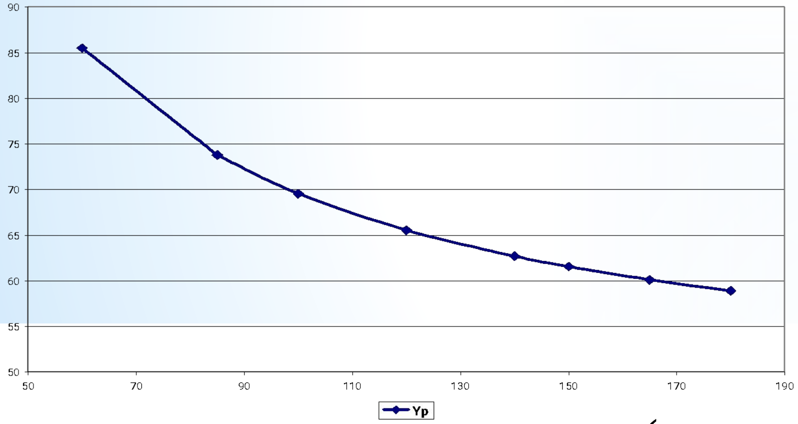
3) экспоненциальная

$$Y = e^{a+bX} \cdot \varepsilon$$

# Кривые Энгеля и Филипса

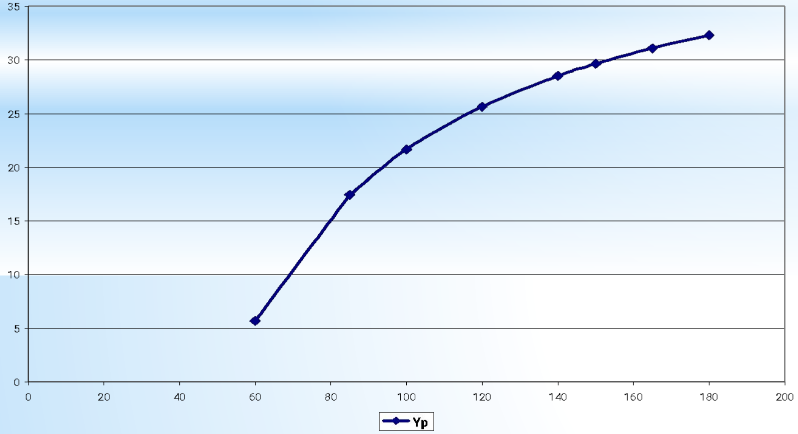
$$Y = a + \frac{b}{X} + e$$

## Кривая Филипса $(b > 0)$




показывает взаимное изменение уровней безработицы (*x*) и инфляции в экономике (процента прироста заработной платы) (*y*)

## Кривая Энгеля $(b < 0)$




показывает величину расходов на товары в зависимости от роста дохода

# Пример

| n | y | x |
|---|---|---|
| 1 | 2 | 50 |
| 2 | 4 | 60 |
| 3 | 11 | 85 |
| 4 | 17 | 85 |
| 5 | 18 | 100 |
| 6 | 28 | 120 |
| 7 | 34 | 140 |

***Требуется*:**

1. Построить степенную, показательную и гиперболическую модели нелинейной регрессии. Результаты моделирования отобразить на графике.

2. Сравнить качественные характеристики моделей, рассчитав коэффициенты детерминации и средние относительные ошибки аппроксимации.

# Степенная модель

$$y = a \cdot x^b \qquad \lg y = \lg a + b \cdot \lg x \qquad Y=A+bX$$

| n | y | x | lg y=Y | lg x=X | yp | ei | ei^2 | eiотн | y-ycp | (y-ycp)^2 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 50 | 0.301 | 1.699 | 2.464 | -0.464 | 0.215 | 23.200 | -14.286 | 204.082 |
| 2 | 4 | 60 | 0.602 | 1.778 | 4.097 | -0.097 | 0.009 | 2.427 | -12.286 | 150.939 |
| 3 | 11 | 85 | 1.041 | 1.929 | 10.82 | 0.177 | 0.031 | 1.606 | -5.286 | 27.939 |
| 4 | 17 | 85 | 1.230 | 1.929 | 10.82 | 6.177 | 38.15 | 36.333 | 0.714 | 0.510 |
| 5 | 18 | 100 | 1.255 | 2.000 | 17.03 | 0.970 | 0.941 | 5.389 | 1.714 | 2.939 |
| 6 | 28 | 120 | 1.447 | 2.079 | 28.31 | -0.317 | 0.101 | 1.133 | 11.714 | 137.224 |
| 7 | 34 | 140 | 1.531 | 2.146 | 43.52 | -9.527 | 90.77 | 28.022 | 17.714 | 313.796 |
| **Сумма** | **114** | **640** | | | | | **130.2** | **98.110** | | **837.429** |
| **Среднее** | **16.28** | **91.429** | | | | | | **14.016** | | |

**Y=-4.346+2.789*X**

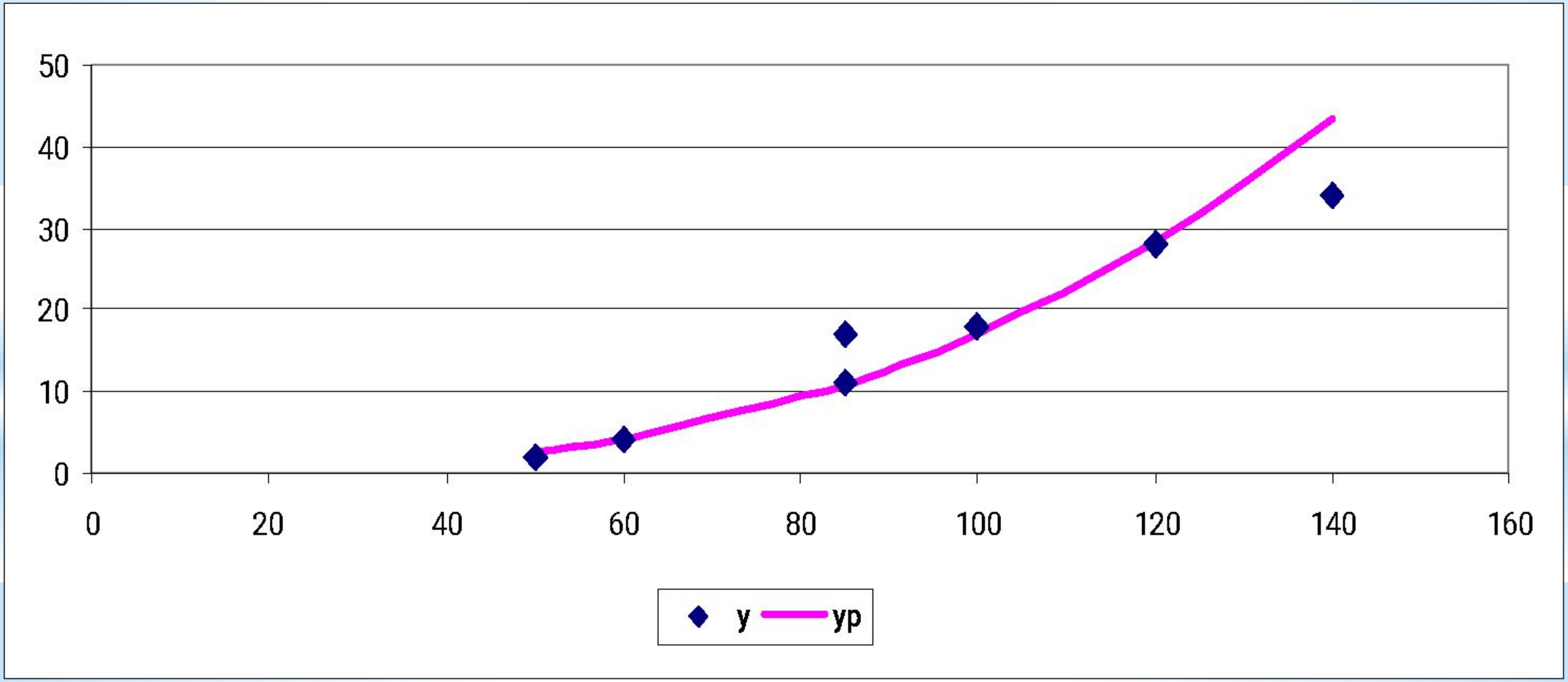
$$\lg y = -4.346 + 2.789 \cdot \lg x$$

$$y = 10^{-4.346} \cdot x^{2.789}$$

$$y = 0.000045 \cdot x^{2.789}$$

$$R^2 = 1 - \frac{130.22}{837.4} = 0.844$$

$$\bar{e}_{отн} = 14\%$$



# Показательная модель

$$y = a \cdot x^b$$

$$\lg y = \lg a + x \cdot \lg b$$

$$Y = A + B \cdot x$$

| n | y | x | lg y=Y | yp | ei | ei^2 | eiотн |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 50 | 0.301 | 3.119 | -1.119 | 1.252 | 55.954 |
| 2 | 4 | 60 | 0.602 | 4.245 | -0.245 | 0.060 | 6.125 |
| 3 | 11 | 85 | 1.041 | 9.173 | 1.827 | 3.339 | 16.611 |
| 4 | 17 | 85 | 1.230 | 9.173 | 7.827 | 61.265 | 46.042 |
| 5 | 18 | 100 | 1.255 | 14.564 | 3.436 | 11.807 | 19.089 |
| 6 | 28 | 120 | 1.447 | 26.976 | 1.024 | 1.048 | 3.657 |
| 7 | 34 | 140 | 1.531 | 49.967 | -15.967 | 254.929 | 46.960 |
| **Сумма** | **114** | **640** | | | | **333.700** | **194.439** |
| **Среднее** | **16.286** | **91.429** | | | | | **27.777** |

**Y=-0.161+0.0133*x**

$$\lg y = -0.161 + 0.0133 \cdot x$$

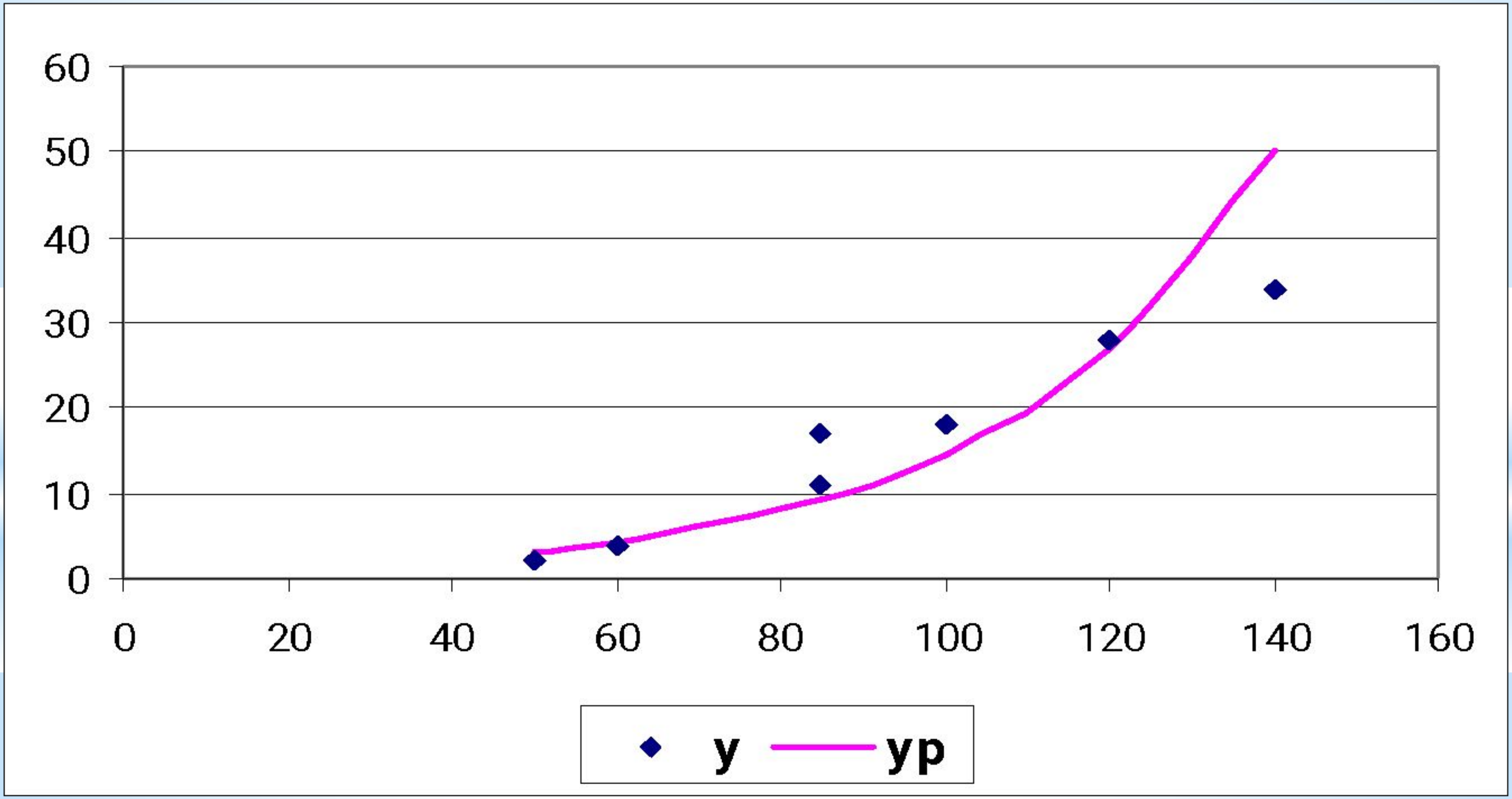
$$y = 10^{-0.161+0.0133x}$$

$$y = 10^{-0.161} \cdot (10^{0.0133})^{x}$$

$$y = 0.668 \cdot 1.0313^{x}$$

$$R^2 = 1 - \frac{333.7}{837.4} = 0.776$$

$$\bar{e}_{отн} \approx 28\%$$



# Гиперболическая модель

$$y = a + \frac{b}{x} \qquad X = \frac{1}{x} \qquad y = a + b \cdot X$$

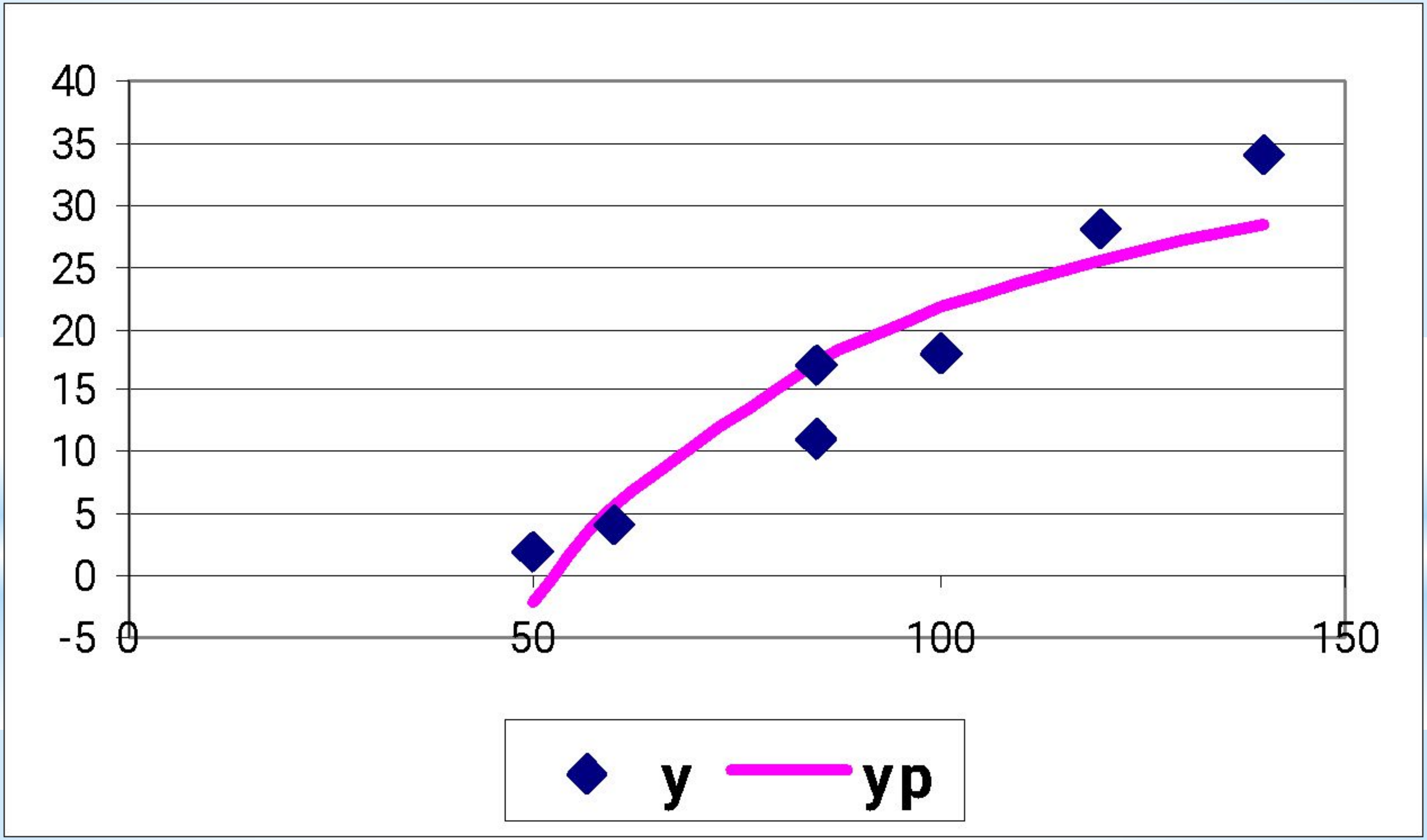
| n | y | x | 1/x=X | yp | ei | ei^2 | eiотн |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 50 | 0.02 | -2.298 | 4.2983 | 18.475 | 214.9137 |
| 2 | 4 | 60 | 0.0167 | 5.6834 | -1.683 | 2.8338 | 42.08507 |
| 3 | 11 | 85 | 0.0118 | 17.421 | -6.421 | 41.231 | 58.37421 |
| 4 | 17 | 85 | 0.0118 | 17.421 | -0.421 | 0.1774 | 2.477429 |
| 5 | 18 | 100 | 0.01 | 21.647 | -3.647 | 13.299 | 20.25976 |
| 6 | 28 | 120 | 0.0083 | 25.638 | 2.3624 | 5.581 | 8.437161 |
| 7 | 34 | 140 | 0.0071 | 28.488 | 5.5118 | 30.38 | 16.21119 |
| **Сумма** | **114** | **640** | | | | **111.98** | **362.7585** |
| **Среднее** | **16.286** | **91.429** | | | | | **51.82265** |

$$y = 45{,}59 - 2395 \cdot X$$

$$R^2 = 1 - \frac{111{,}98}{837{,}37} = 0.866$$

$$y = 45{,}59 - \frac{2395}{x}$$

$$\bar{e}_{отн} \approx 51{,}8\%$$

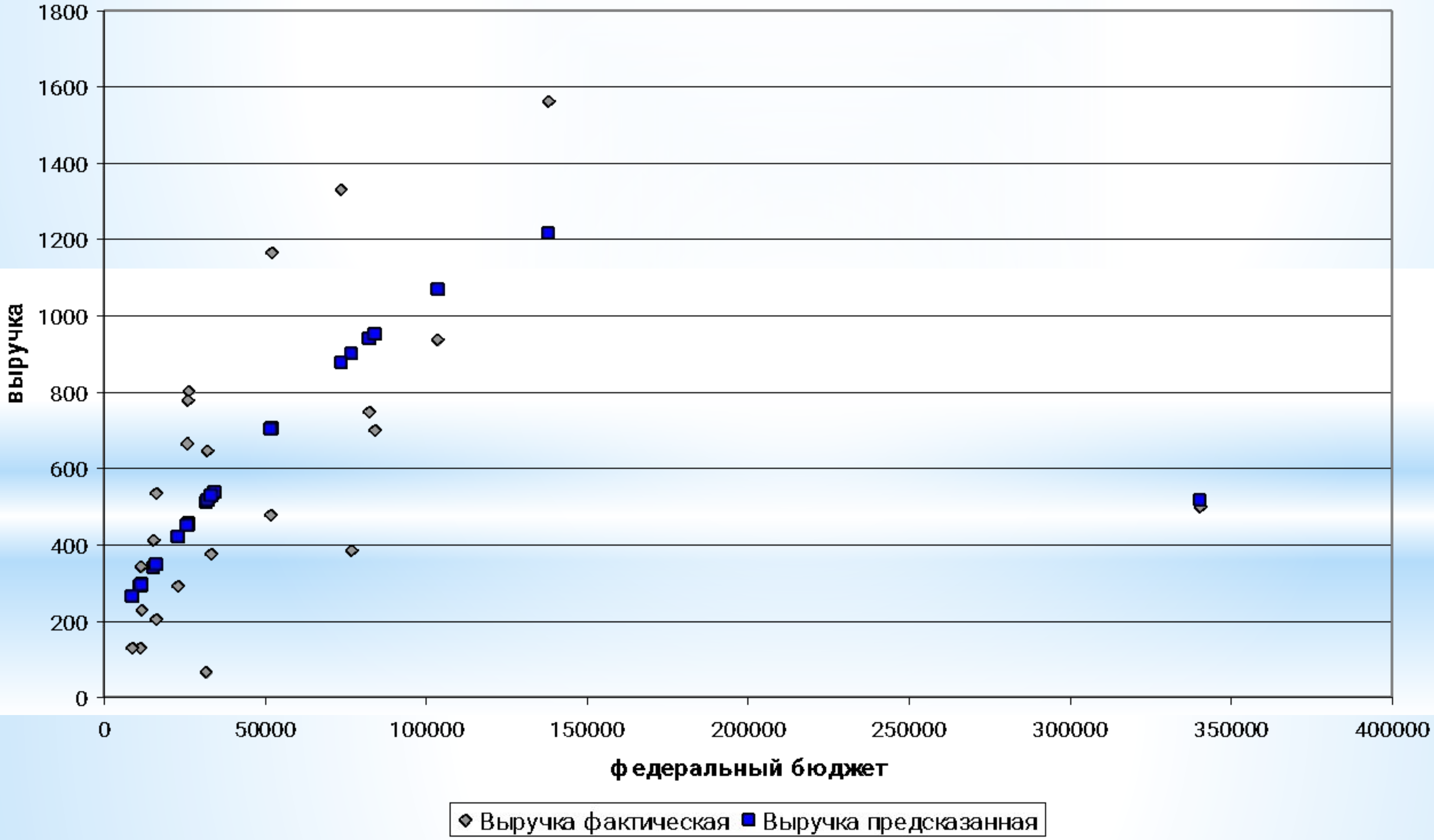


# Сравнение моделей

| Модель | $R^2$ | $\bar{e}_{отн.}$ |
|---|---|---|
| **Степенная** | **0.844** | **14.02** |
| Показательная | 0.776 | 27.78 |
| Гиперболическая | 0.866 | 51.8 |

# 4. Причины ложных результатов регрессионного анализа

1. ***<u>Грубое искажение вида модели или оценок ее параметров</u>***

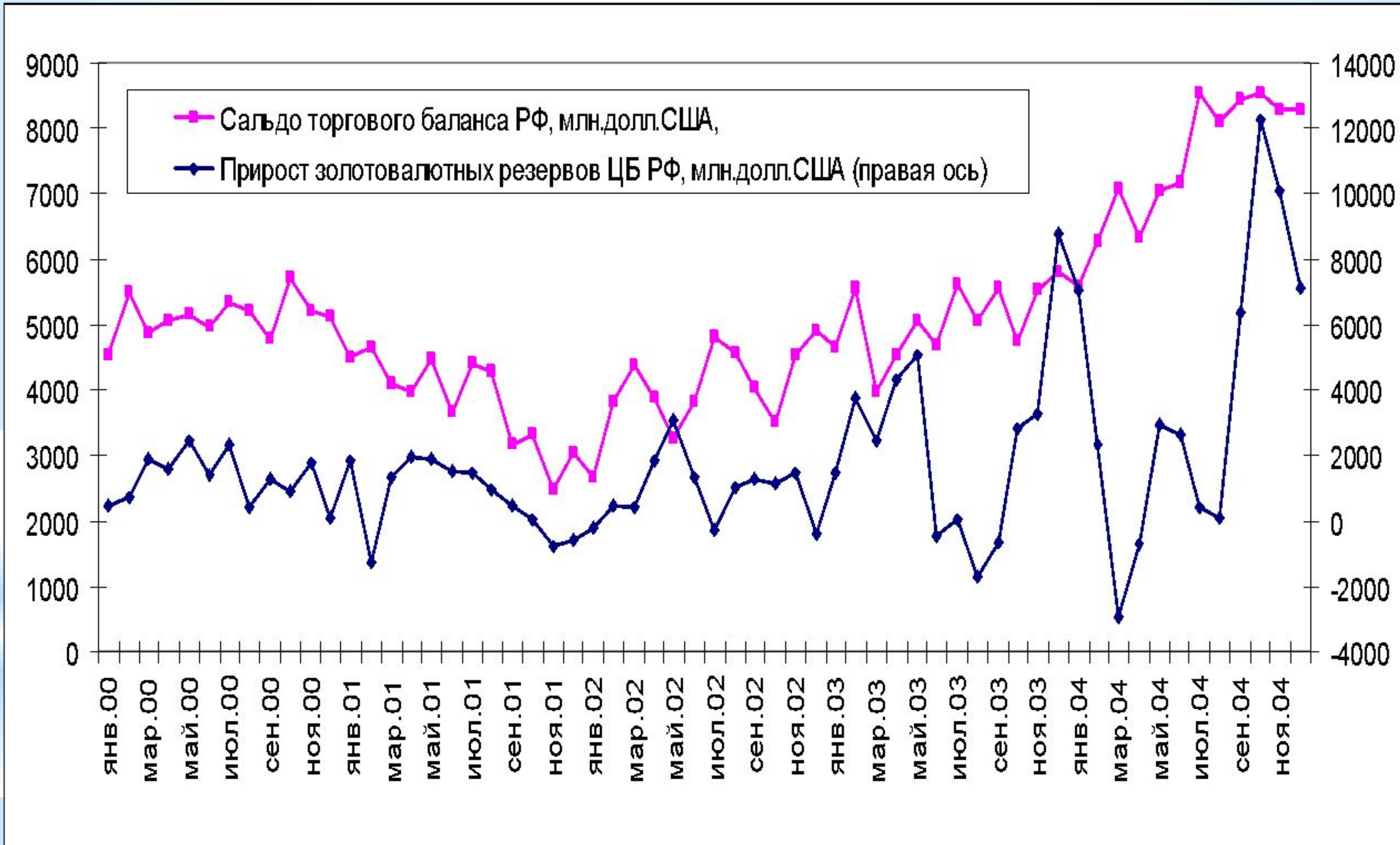


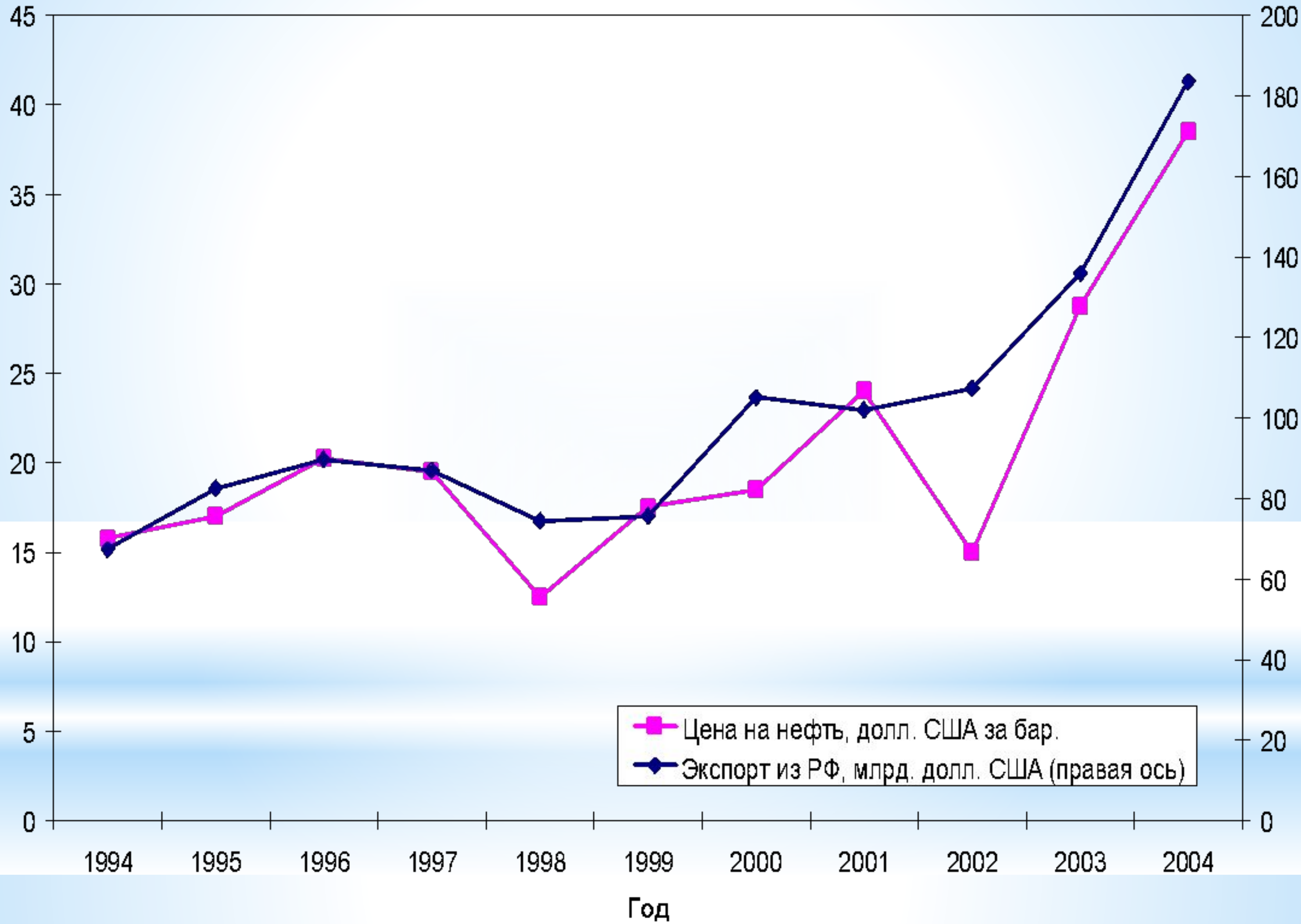
2. ***Несоответствие результатов корреляционного и регрессионного анализа***

3. ***Отрицательные расчетные значения эндогенной переменной***

4. ***Интервал прогноза с отрицательной нижней границей или слишком широкий***

# 5. Эффект ложной регрессии





45
40
35
30
25
20
15
10
5
0
200
180
160
140
120
100
80
60
40
20
0
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
Год
Цена на нефть, долл. США за бар.
Экспорт из РФ, млрд. долл. США (правая ось)

СПАСИБО

ЗА ВНИМАНИЕ!