метод наименьших квадратов

Лекции 5-6

Цели лекции

- Раскрыть понятие регрессии.
- Познакомиться с методом наименьших квадратов – методом построения линейного уравнения регрессии.

ЗАДАЧИ РЕГРЕССИОННОГО АНАЛИЗА

Задачи линейного регрессионного анализа состоят в том, чтобы по имеющимся статистическим данным ($\boldsymbol{x}_i$, y_i), i = 1, 2, …, n, для набора регрессоров ***X*** и зависимой переменной *Y*:

а) получить наилучшие оценки параметров модели

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \left(y_i = \beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \varepsilon_i,\ i = 1,\ldots n\right)$$

б) проверить статистические гипотезы о параметрах модели;

в) проверить, адекватность модели данным наблюдений.

ЭМПИРИЧЕСКОЕ УРАВНЕНИЕ РЕГРЕССИИ

По выборке ограниченного объема нельзя точно определить теоретические значения параметров β_k.

Можно лишь построить

эмпирическое уравнение регрессии:

$$\hat{\mathbf{y}}_i = \mathbf{x}_i \mathbf{b} \quad \left(\hat{y}_i = b_1 + x_{2i} b_2 + x_{3i} b_3, i = 1, \ldots n \right)$$

где b_k – оценки параметров β_k

эмпирические коэффициенты регрессии).

$\hat{y}_i$ – оценка условного м. о. E[*Y*/***X*** = ***x***$_i$].

ЗАДАЧА ОПРЕДЕЛЕНИЯ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Задача состоит в нахождении по выборке данных оценок b_k так, чтобы построенная линия регрессии

была ***наилучшей в определенном смысле*** среди всех других.

Решение основано на минимизации некоторого функционала:

$$g(y_i, \mathbf{x}_i, \mathbf{b}) \to \min_{\mathbf{b}}, \quad i = \overline{1, n}$$

где g – некоторая функция.

МЕТОД МИНИМИЗАЦИИ ОСТАТКОВ

Основная идея – минимизировать остатки с помощью какой-нибудь функции невязок $\mathbf{g}(\mathbf{e})$:

$$\mathbf{g}\left(y, \mathbf{x}, \mathbf{b}\right) = \mathbf{g}\left(y - \mathbf{x} \cdot \mathbf{b}\right) = \mathbf{g}\left(\mathbf{e}\right) =$$

$$= \sum_{i=1}^{n} g(y_i - \hat{y}_i) = \sum_{i=1}^{n} \mathrm{g}\left(e_i\right) \to \min$$

МЕТОД МИНИМИЗАЦИИ ОСТАТКОВ

Возможные кандидаты на роль g(e):

1) линейная функция- $g(e)=e$

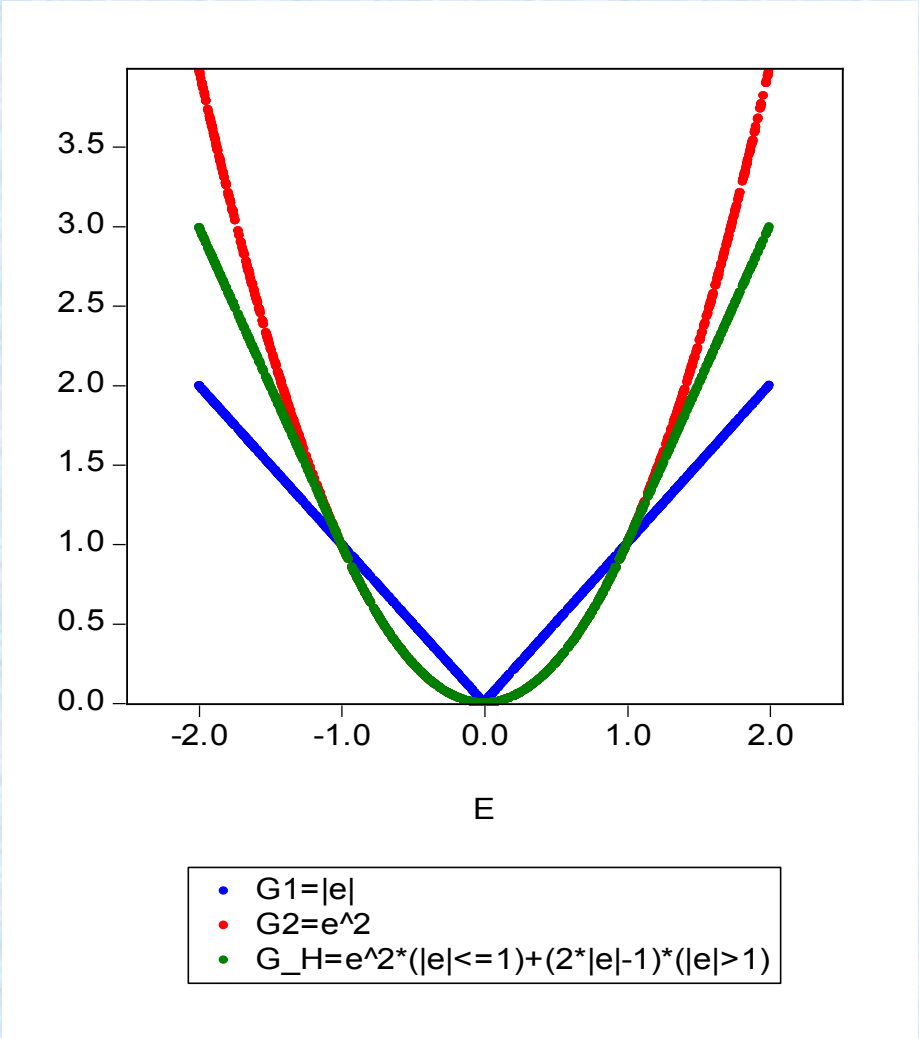
2) модуль- $g(e)=|e|$

3) квадратичная функция- $g(e)=e^2$

4) функция Хубера-

$$g(e)=\begin{cases} e^2, |e|<c \\ 2c\cdot e-c^2, e\geq c \\ -2c\cdot e-c^2, e\leq c \end{cases}$$

МЕТОД МИНИМИЗАЦИИ ОСТАТКОВ



МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Наиболее распространена ***методом наименьших квадратов*** (***МНК***), использующий в качестве функции невязок – квадратичную функцию отклонений:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 \rightarrow \min$$

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 \to \min$$

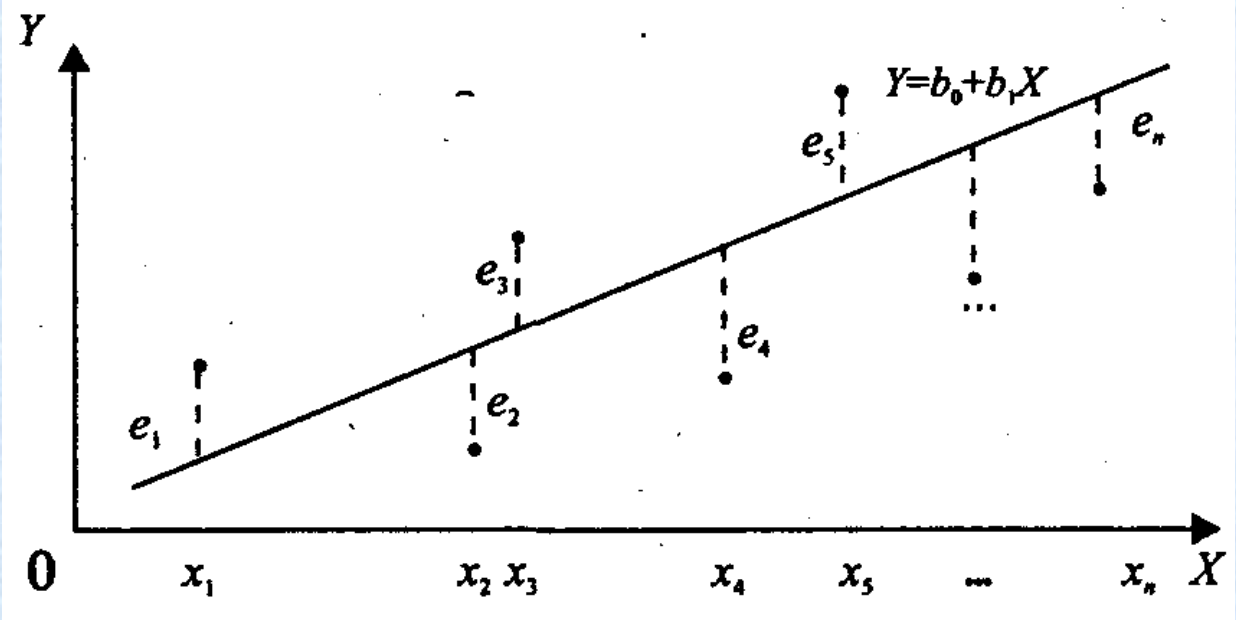
Основные особенности МНК:

1) Он наиболее простой с вычислительной точки зрения.

2) Оценки коэффициентов регрессии по МНК при определенных предпосылках обладают рядом оптимальных свойств.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Пусть по выборке данных (x_i, y_i), i = 1, 2, …, n,требуется определить оценки b_1 и b_2 эмпирического уравнения регрессии:



МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

В этом случае минимизируется функция:

$$Q(b_1, b_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2.$$

Т.к. функция $Q(b_0,b_1)$ непрерывна, выпукла и ограничена снизу, то она имеет минимум.

Необходимым условием минимума $Q(b_1,b_2)$ является равенство нулю ее частных производных по неизвестным параметрам b_1 и b_2.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Приравняем нулю частные производные и затем разделим на *n* оба уравнения:

$$\begin{cases} \dfrac{\partial Q}{\partial b_1} = -2\sum (y_i - b_1 - b_2 x_i) = 0 \\ \dfrac{\partial Q}{\partial b_2} = -2\sum (y_i - b_1 - b_2 x_i) x_i = 0 \end{cases} \Rightarrow \begin{cases} n b_1 + b_2 \sum x_i = \sum y_i \\ b_1 \sum x_i + b_2 \sum x_i^2 = \sum x_i y_i \end{cases}$$

$$\Rightarrow \begin{cases} b_1 + b_2 \bar{x} = \bar{y} \\ b_1 \bar{x} + b_2 \overline{x^2} = \overline{xy} \end{cases}$$

ОЦЕНКИ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ (МНК-оценки, OLS-estimation)

Решив последнюю систему уравнений, получим:

$$b_2 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{C\tilde{o}v(x,y)}{Var(x)} = r_{xy}\sqrt{\frac{Var(y)}{Var(x)}}$$

$$b_1 = \bar{y} - b_2 \bar{x}$$

ОЦЕНКИ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ (МНК-оценки, OLS-estimation)

$$b_2 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$$

$$b_2 = \frac{\frac{1}{n}\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Вид формулы в отклонениях:

$$x_i = X_i - \bar{X},\ \bar{x} = 0; \quad y_i = Y_i - \bar{Y},\ \bar{y} = 0.$$

$$b_2 = \frac{\sum x_i y_i}{\sum x_i^2}; \quad b_1 = \bar{y} - \bar{x} b_2 = 0$$

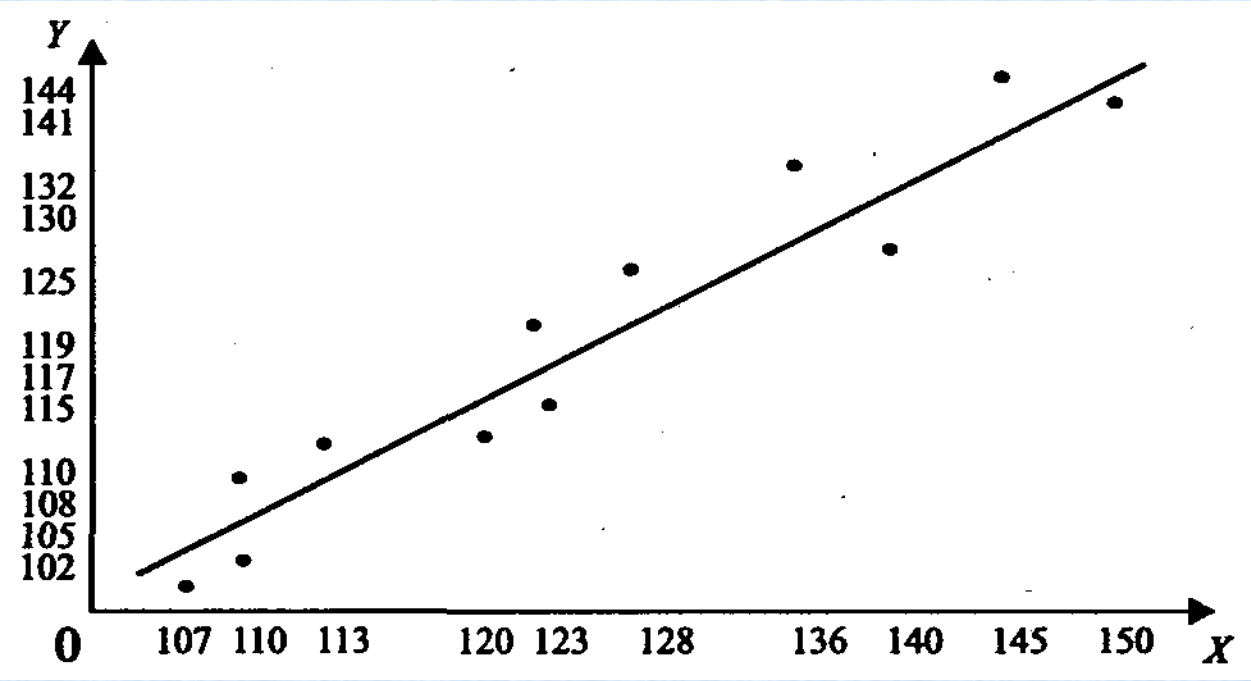
Пример построения уравнения регрессии

При анализе зависимости объема потребления *Y* (у.е.) домохозяйства от располагаемого дохода *X* (у.е.) отобрана выборка объема *n* = 12 (помесячно в течение года), результаты которой приведены в таблице:

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Пример построения уравнения регрессии

Для определения вида зависимости построим корреляционное поле:



Пример построения уравнения регрессии

По расположению точек на корреляционном поле делаем предположение о линейной зависимости:

$$\hat{Y} = b_1 + b_2 X.$$

Согласно МНК, имеем:

$$b_2 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$b_1 = \bar{y} - b_1 \bar{x}$$

Пример
Таблица расчетов по МНК

№№	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\hat{y}_i$	e_i	e_i^2
1	107	102	11449	10914	10404	103,58	-1,583	2,507
2	109	105	11881	11445	11025	105,46	-0,455	0,207
3	110	108	12100	11880	11664	106,39	1,609	2,587
4	113	110	12769	12430	12100	109,20	0,800	0,641
5	120	115	14400	13800	13225	115,75	-0,752	0,566
6	122	117	14884	14274	13689	117,62	-0,624	0,390
7	123	119	15129	14637	14161	118,56	0,440	0,193
8	128	125	16384	16000	15625	123,24	1,759	3,094
9	136	132	18496	17952	17424	130,73	1,270	1,614
10	140	130	19600	18200	16900	134,47	-4,474	20,015
11	145	141	21025	20445	19881	139,15	1,846	3,407
12	150	144	22500	21600	20736	143,83	0,165	0,027
Сумма	1503	1448	190617	183577	176834	-	$1{,}4 \cdot 10^{-14}$	35,249
Среднее	125,25	120,67	15884,75	15298,1	14736,2	-	-	-

Пример построения уравнения регрессии

По расположению точек на корреляционном поле делаем предположение о линейной зависимости: $\hat{Y} = b_1 + b_2 X.$

Согласно МНК, имеем:

$$b_2 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{15298,08 - 125,25 \cdot 120,67}{15884,75 - (125,25)^2} = \frac{184,583}{197,188} = 0,9361$$

$$b_1 = \overline{y} - b_2 \overline{x} = 120,67 - 0,9361 \cdot 125,25 = 3,423$$

Пример построения уравнения регрессии

Т.о., уравнение парной линейной регрессии имеет вид:

$$\hat{Y} = 3{,}423 + 0{,}9361X$$

Изобразим данную прямую регрессии на корреляционном поле.

По этому уравнению рассчитаем $\hat{y}_i$, а также $e_i = y_i - \hat{y}_i$.

Для анализа степени линейной зависимости вычислим:

$$r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\sqrt{\overline{y^2} - \overline{y}^2}} = \frac{184{,}1625}{14{,}04 \cdot 13{,}23} = 0{,}9914$$

Отсюда можно сделать вывод о сильной прямой линейной зависимости между переменными.

ОБЩЕЕ РЕШЕНИЕ ДЛЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

$$\text{MSPE (BLP)}: \quad \min_{\boldsymbol{\beta}} \mathrm{E}\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)^2$$

FOC (УСЛОВИЯ ПЕРВОГО ПОРЯДКА): $-2\,\mathrm{E}\,\mathbf{X}'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)=0$

SOC (УСЛОВИЯ ВТОРОГО ПОРЯДКА):

$\left(\mathbf{X}'\mathbf{X}\right)'=\mathbf{X}'\mathbf{X}$ симметрична $\Rightarrow$ пол. определена

РЕШЕНИЕ:

$$\mathrm{E}\left(\mathbf{X}'\mathbf{Y}\right)-\mathrm{E}\left(\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right)=0 \quad \Leftrightarrow \quad \boldsymbol{\beta}=\mathrm{E}^{-1}\left(\mathbf{X}'\mathbf{X}\right)\cdot\mathrm{E}\left(\mathbf{X}'\mathbf{Y}\right)$$

Выборочные оценки:

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

РЕШЕНИЕ СУЩЕСТВУЕТ ТОЛЬКО ПРИ НЕОСОБЕННОЙ МАТРИЦЕ- РЕГРЕССОРЫ ДОЛЖНЫ БЫТЬ НЕЗАВИСИМЫ

МАТРИЧНЫЙ ВИД ОЦЕНКИ КОЭФФИЦИЕНТОВ

FOC (УСЛОВИЯ ПЕРВОГО ПОРЯДКА): $-2\,\mathrm{E}\,\mathbf{X}'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)=0$

$$\mathbf{e}=\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}$$

МНК эквивалентен ортогональности матрицы **X** и вектора **e**:

$$\mathbf{X}^T\mathbf{e}=\mathbf{0} \qquad \mathbf{X}^T(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{0} \Rightarrow \hat{\boldsymbol{\beta}}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

МАТРИЧНЫЙ ВИД ОЦЕНКИ КОЭФФИЦИЕНТОВ

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}; \quad \mathbf{X} = \left[\mathbf{X}_1 \ldots \mathbf{X}_k\right] = \begin{bmatrix} X_{11} \cdots X_{1k} \\ \vdots \quad \ddots \quad \vdots \\ X_{N1} \cdots X_{Nk} \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \Rightarrow \mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{u} \Rightarrow$$

$$\Rightarrow \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}$$

$$\boldsymbol{\beta} \approx \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \hat{\boldsymbol{\beta}}$$

ПАРНАЯ РЕГРЕССИЯ

$$\mathbf{X} = \left[\mathbf{X}_1 \ldots \mathbf{X}_k\right] = \left[\mathbf{I}; \mathbf{X}_2\right] = \begin{bmatrix} 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{N2} \end{bmatrix}$$

$$\mathbf{X}' = \begin{bmatrix} \mathbf{I}' \\ \mathbf{X}_2' \end{bmatrix} = \begin{bmatrix} 1 & \ldots & 1 \\ X_{12} & \ldots & X_{N2} \end{bmatrix}$$

ПАРНАЯ РЕГРЕССИЯ

$$\mathbf{X} = \left[\mathbf{X}_1 \ldots \mathbf{X}_k\right] = \left[\mathbf{I}; \mathbf{X}_2\right] = \begin{bmatrix} 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{N2} \end{bmatrix}$$

$$\mathbf{X}' = \begin{bmatrix} \mathbf{I} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} 1 & \ldots & 1 \\ X_{12} & \ldots & X_{N2} \end{bmatrix}; \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \ldots & \sum X_{2i} \\ \sum X_{2i} & & \sum X_{2i}^2 \end{bmatrix}; \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum Y_i \\ \sum X_{2i} Y_i \end{bmatrix}$$

ПАРНАЯ РЕГРЕССИЯ

$$\mathbf{X}' = \begin{bmatrix} \mathbf{I} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ X_{12} & \dots & X_{N2} \end{bmatrix};$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum Y_i \\ \sum X_{2i} Y_i \end{bmatrix}; \quad \mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \sum X_{2i} \\ \sum X_{2i} & \sum X_{2i}^2 \end{bmatrix}$$

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \frac{1}{N\sum X_{2i}^2 - \left(\sum X_{2i}\right)^2} \begin{bmatrix} \sum X_{2i}^2 & -\sum X_{2i} \\ -\sum X_{2i} & N \end{bmatrix}$$

ПАРНАЯ РЕГРЕССИЯ

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}=\frac{1}{N\sum X_{2i}^{2}-\left(\sum X_{2i}\right)^{2}}\begin{bmatrix}\sum X_{2i}^{2} & -\sum X_{2i}\\ -\sum X_{2i} & N\end{bmatrix}\begin{bmatrix}\sum Y_{i}\\ \sum X_{2i}Y_{i}\end{bmatrix}=$$

ПАРНАЯ РЕГРЕССИЯ

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{N\sum X_{2i}^2 - \left(\sum X_{2i}\right)^2}\begin{bmatrix}\sum X_{2i}^2 & -\sum X_{2i} \\ -\sum X_{2i} & N\end{bmatrix}\begin{bmatrix}\sum Y_i \\ \sum X_{2i}Y_i\end{bmatrix} =$$

$$= \frac{1}{N^2\,\mathrm{Var}\left(\mathbf{X}_2\right)}\begin{bmatrix}\sum X_{2i}^2\sum Y_i - \sum X_{2i}\sum X_{2i}Y_i \\ -\sum X_{2i}\sum Y_i + N\sum X_{2i}Y_i\end{bmatrix}$$

$$\Rightarrow \hat{\beta}_2 = \frac{-\sum X_{2i}\sum Y_i + N\sum X_{2i}Y_i}{N^2\,\mathrm{Var}\left(\mathbf{X}_2\right)} = \frac{\mathrm{cov}\left(\mathbf{X}_2,\mathbf{Y}\right)}{\mathrm{Var}\left(\mathbf{X}_2\right)}$$

ПАРНАЯ РЕГРЕССИЯ

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}=\frac{1}{N^2\,\mathrm{Var}\left(\mathbf{X}_2\right)}\begin{bmatrix}\sum X_{2i}^2\sum Y_i-\sum X_{2i}\sum X_{1i}Y_i\\ -\sum X_{2i}\sum Y_i+N\sum X_{2i}Y_i\end{bmatrix}$$

$$\Rightarrow\hat{\beta}_2=\frac{-\sum X_{2i}\sum Y_i+N\sum X_{2i}Y_i}{N^2\,\mathrm{Var}\left(\mathbf{X}_2\right)}=\frac{\mathrm{cov}\left(\mathbf{X}_2,\mathbf{Y}\right)}{\mathrm{Var}\left(\mathbf{X}_2\right)}$$

Упр.1 проверить: $$\hat{\beta}_1=\frac{\sum X_{2i}^2\sum Y_i-\sum X_{2i}\sum X_{2i}Y_i}{N^2\,\mathrm{Var}\left(\mathbf{X}_2\right)}=\bar{\mathbf{Y}}-\bar{\mathbf{X}}_2\hat{\beta}_2$$

Упр.2 Выведите формулы для коэффициентов, если:

$$\mathbf{Y}=\beta_1\mathbf{X}_1+\beta_2\mathbf{X}_2+u$$

$$x_1=\begin{cases}1, & \text{для первых N/2 наблюдений}\\ 0, & \text{для остальных}\end{cases}$$

$$x_2=\begin{cases}0, & \text{для первых N/2 наблюдений}\\ 1, & \text{для остальных}\end{cases}$$

Выводы

1. Оценки МНК являются функциями от выборки, что позволяет их легко рассчитать.

2. Оценки МНК являются точечными оценками теоретических коэффициентов регрессии.

3. Эмпирическая прямая регрессии обязательно проходит через точку $(\overline{x}, \overline{y}).$

Выводы

4. Эмпирическое уравнение регрессии построено так, что $\sum e_i = 0, \quad \bar{e} = 0.$

5. Случайные отклонения e_i не коррелированы с наблюдаемыми значениями y_i зависимой переменной Y.

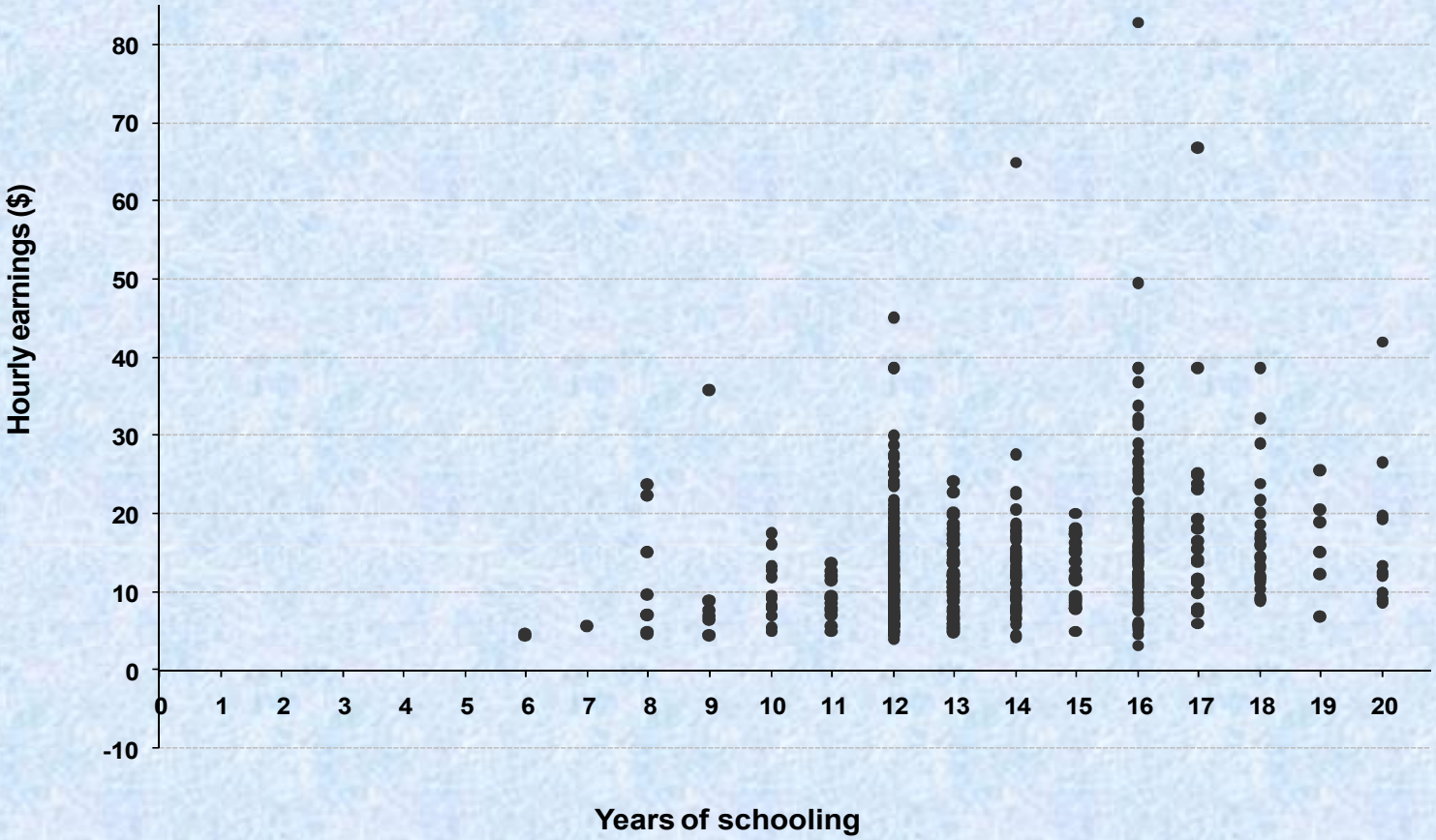
6. Случайные отклонения e_i не коррелированы с наблюдаемыми значениями x_i независимой переменной X.

Другие методы определения коэффициентов регрессии

Другие методы определения коэффициентов регрессии:

- метод моментов (ММ)
- метод максимального правдоподобия (ММП).

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



Данные 1994 г. о заработной плате и уровне образования по 570 респондентам National Longitudinal Survey of Youth.

12 лет – средняя школа

13-16 лет – колледж (бакалавриат)

17-18 лет – университет (магистратура)

19-20 лет - PhD

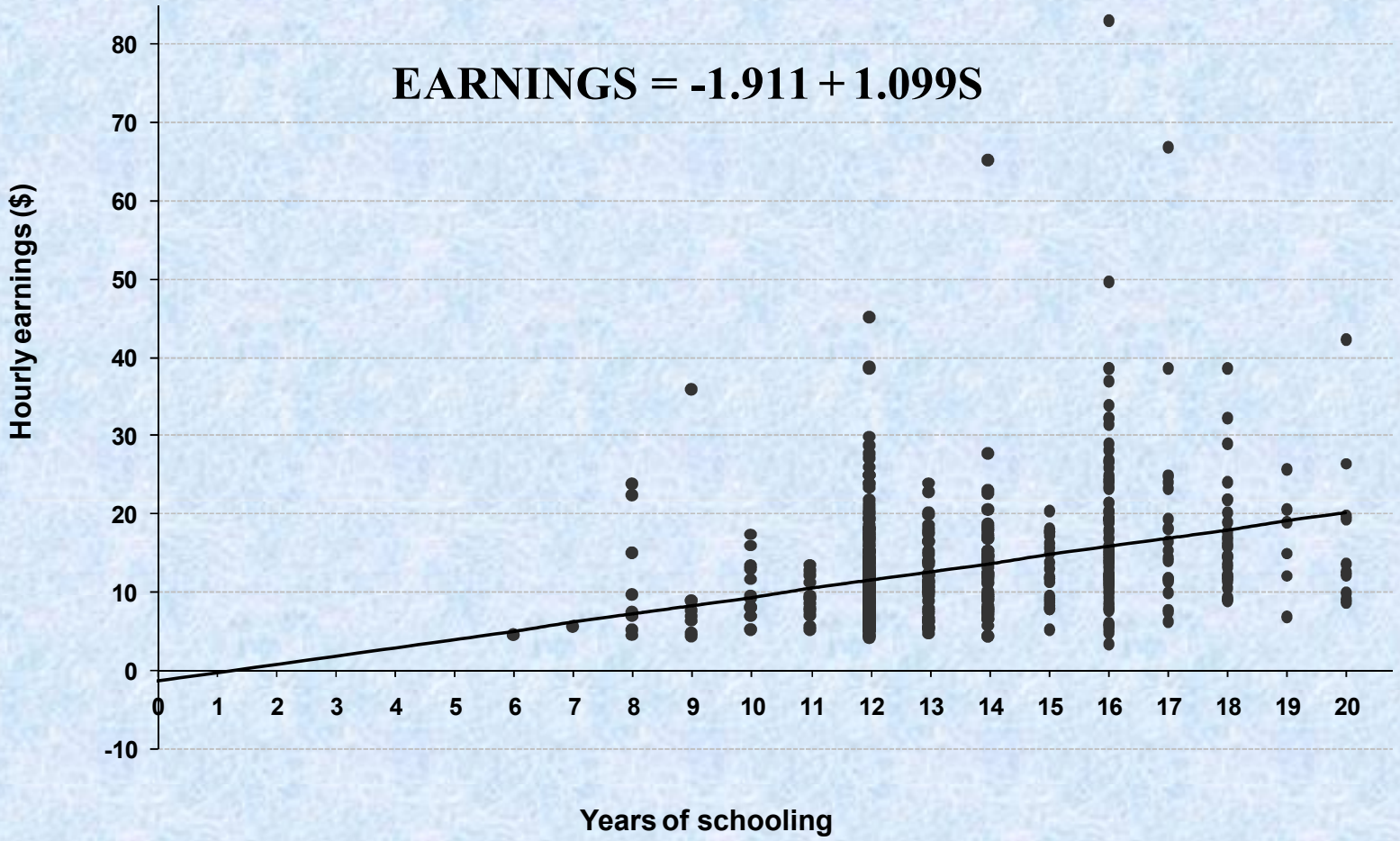
ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

Dependent Variable: EARNINGS
Method: Least Squares
Date: 09/20/08 Time: 21:59
Sample: 1 570
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
S	1.098764	0.132487	8.293371	0.0000
C	-1.910908	1.820813	-1.049481	0.2944

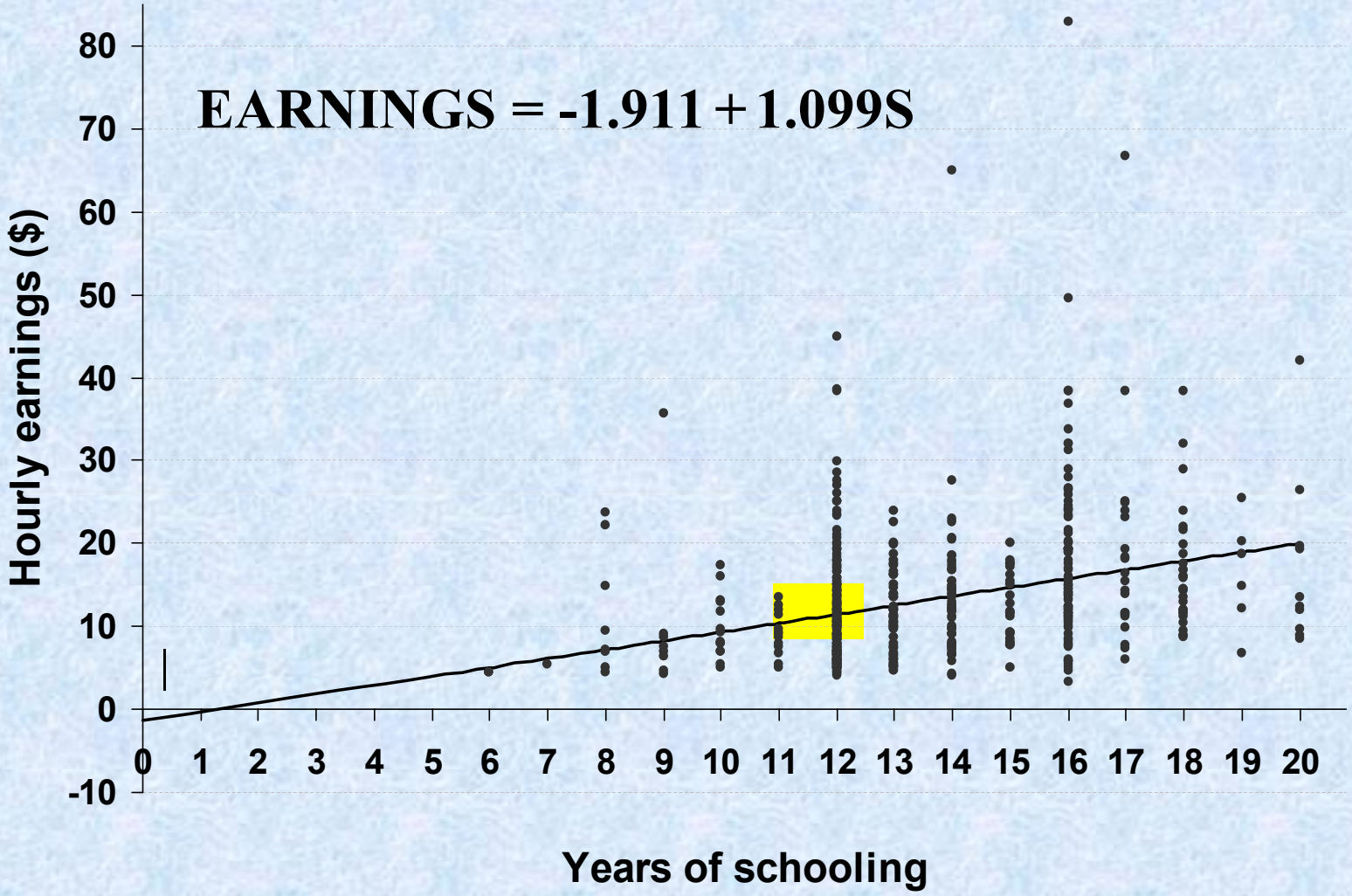
R-squared	0.103086	Mean dependent var	13.11782
Adjusted R-squared	0.101506	S.D. dependent var	8.214719
S.E. of regression	7.786642	Akaike info criterion	6.946199
Sum squared resid	34438.86	Schwarz criterion	6.961447
Log likelihood	-1977.667	F-statistic	65.28223
Durbin-Watson stat	1.933596	Prob(F-statistic)	0.000000

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



Интерпретация коэффициентов зависит от единиц измерения!!!

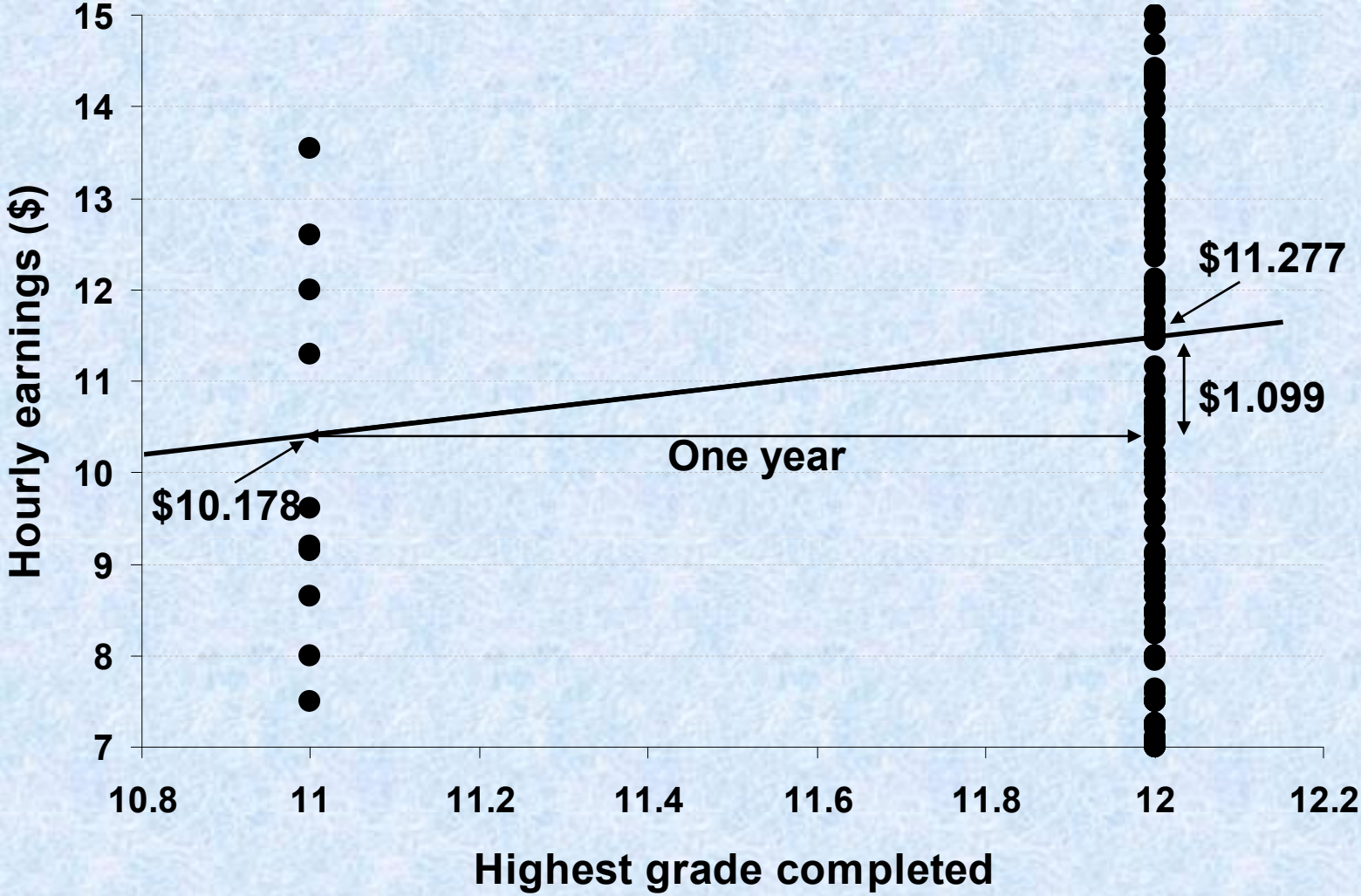
ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



S – измеряется в годах,

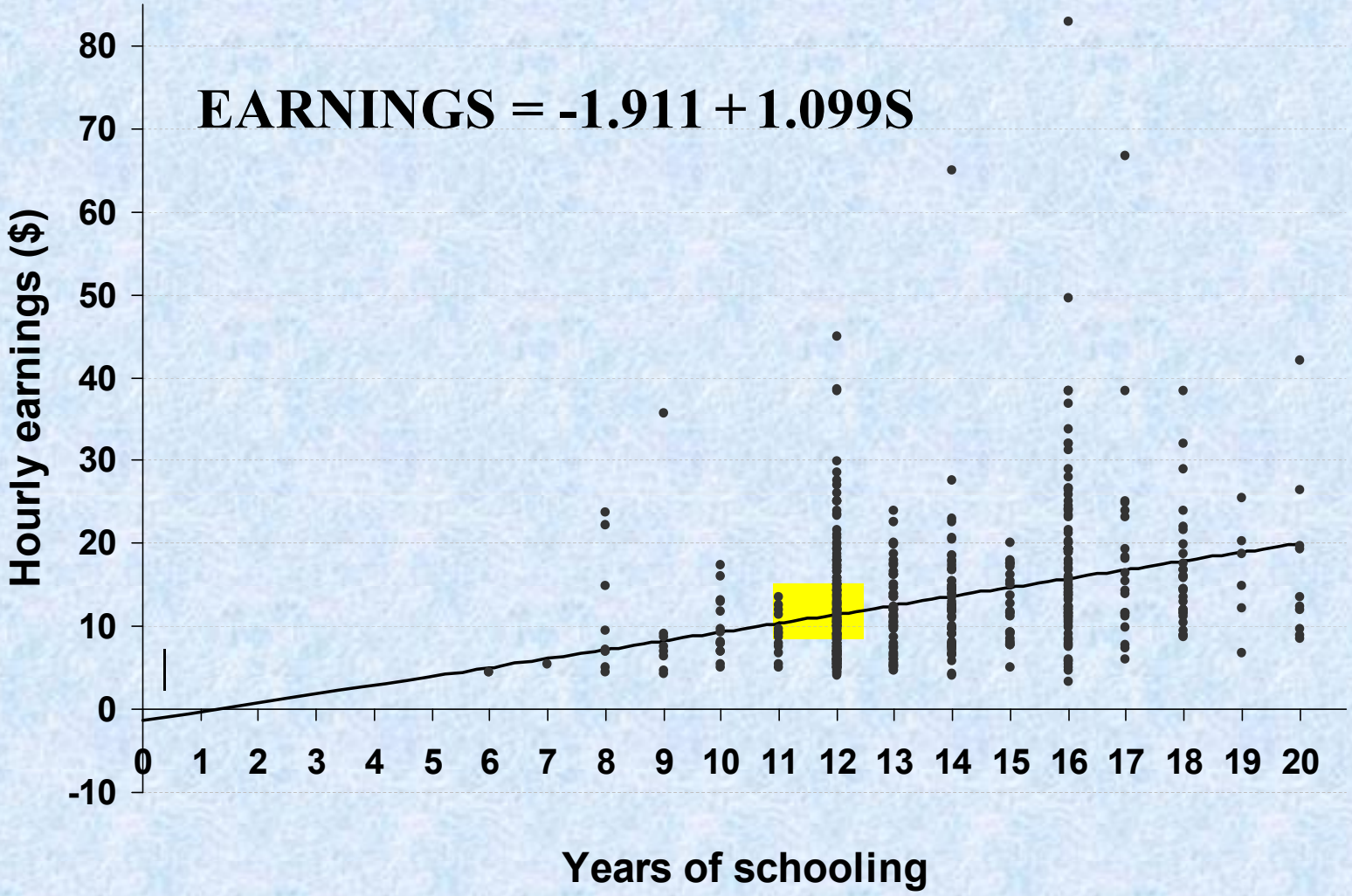
Earnings - в $/час

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



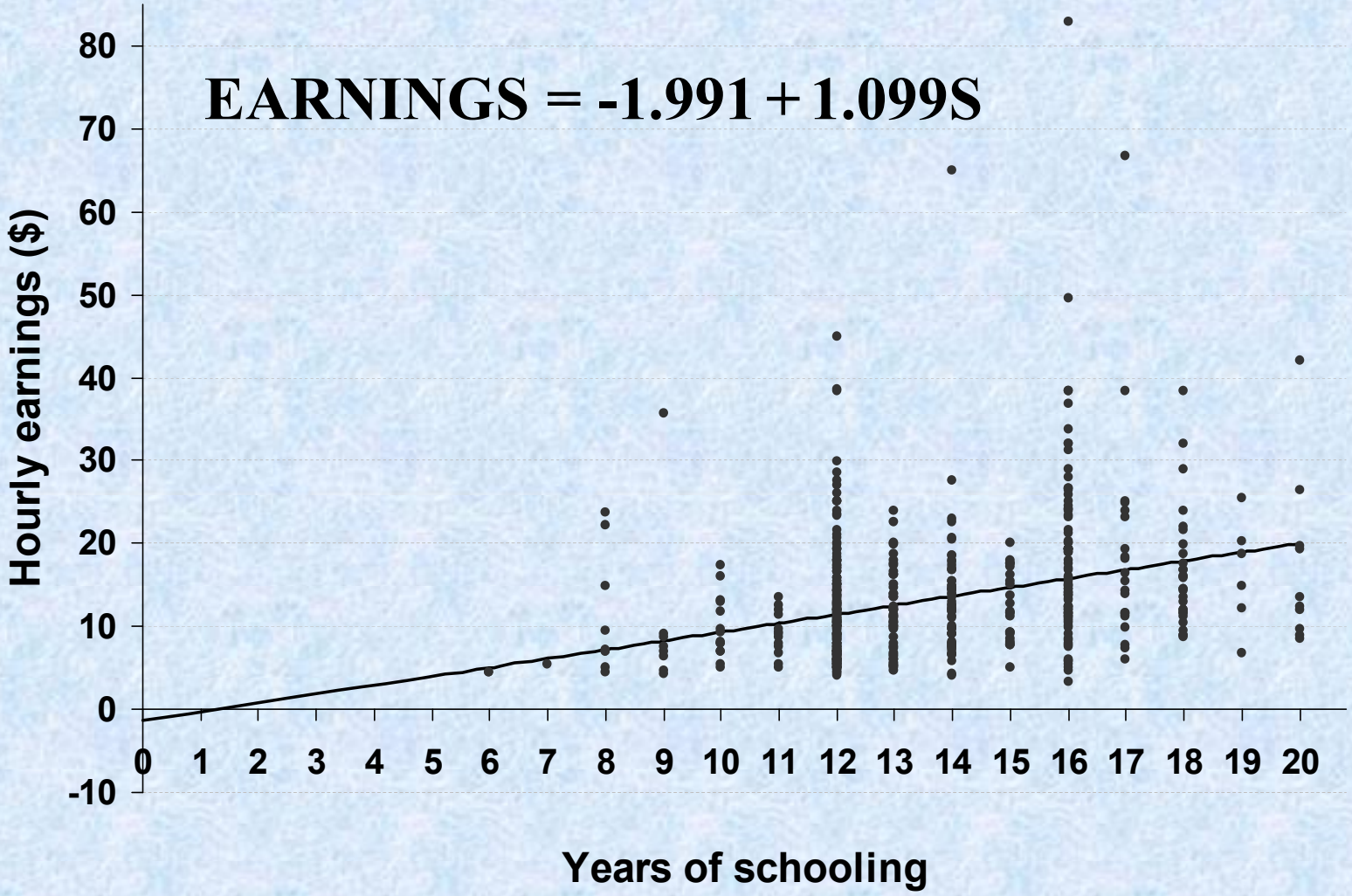
Увеличении уровня образования с 11 лет до 12 (окончание средней школы) приведет в среднем к увеличению почасовой заработной платы на $1.099, с $10.413 до $11.486

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



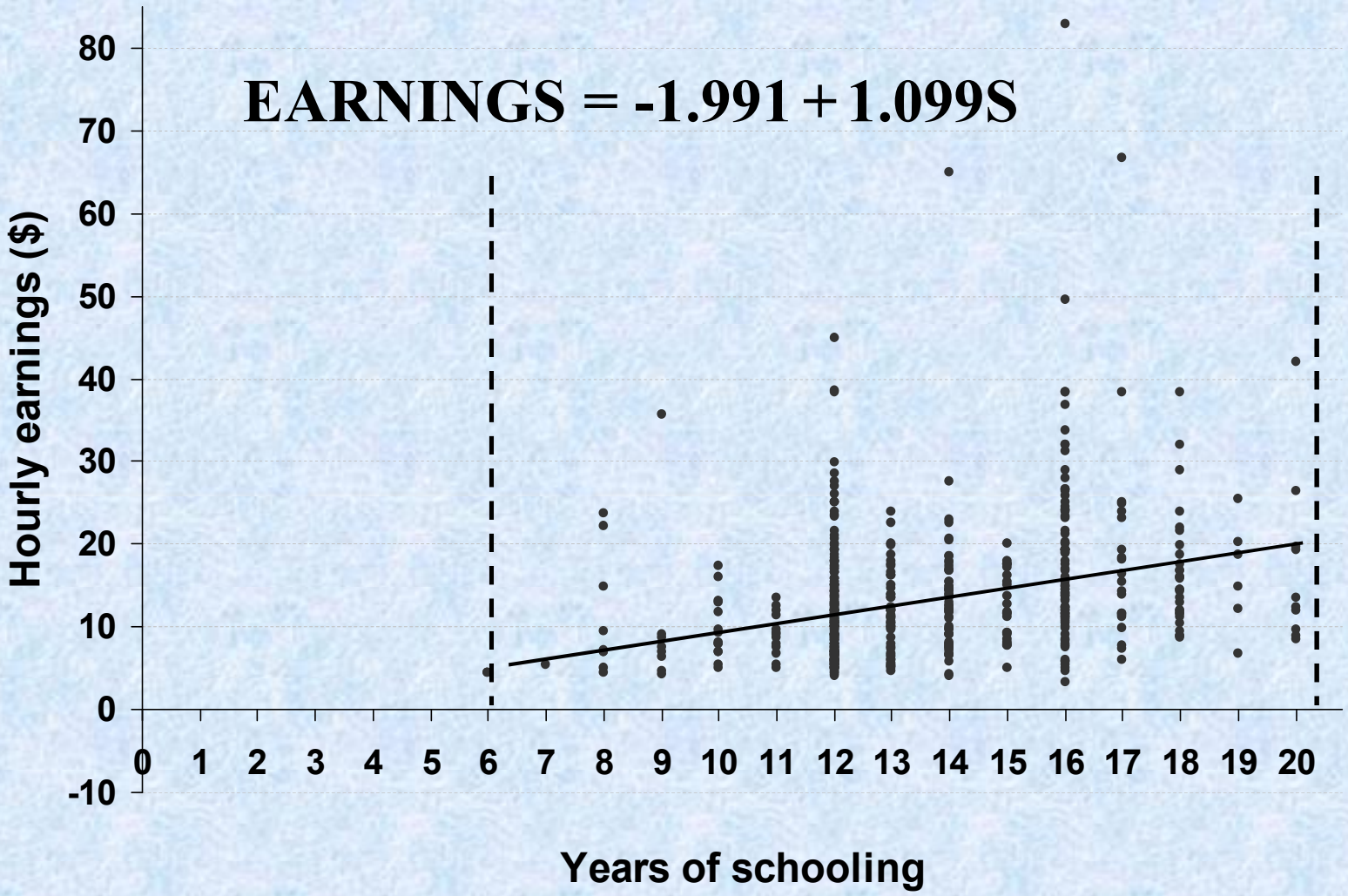
Значение коэффициента наклона правдоподобно для среднего уровня, но неправдоподобно для малого и большого числа лет обучения

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



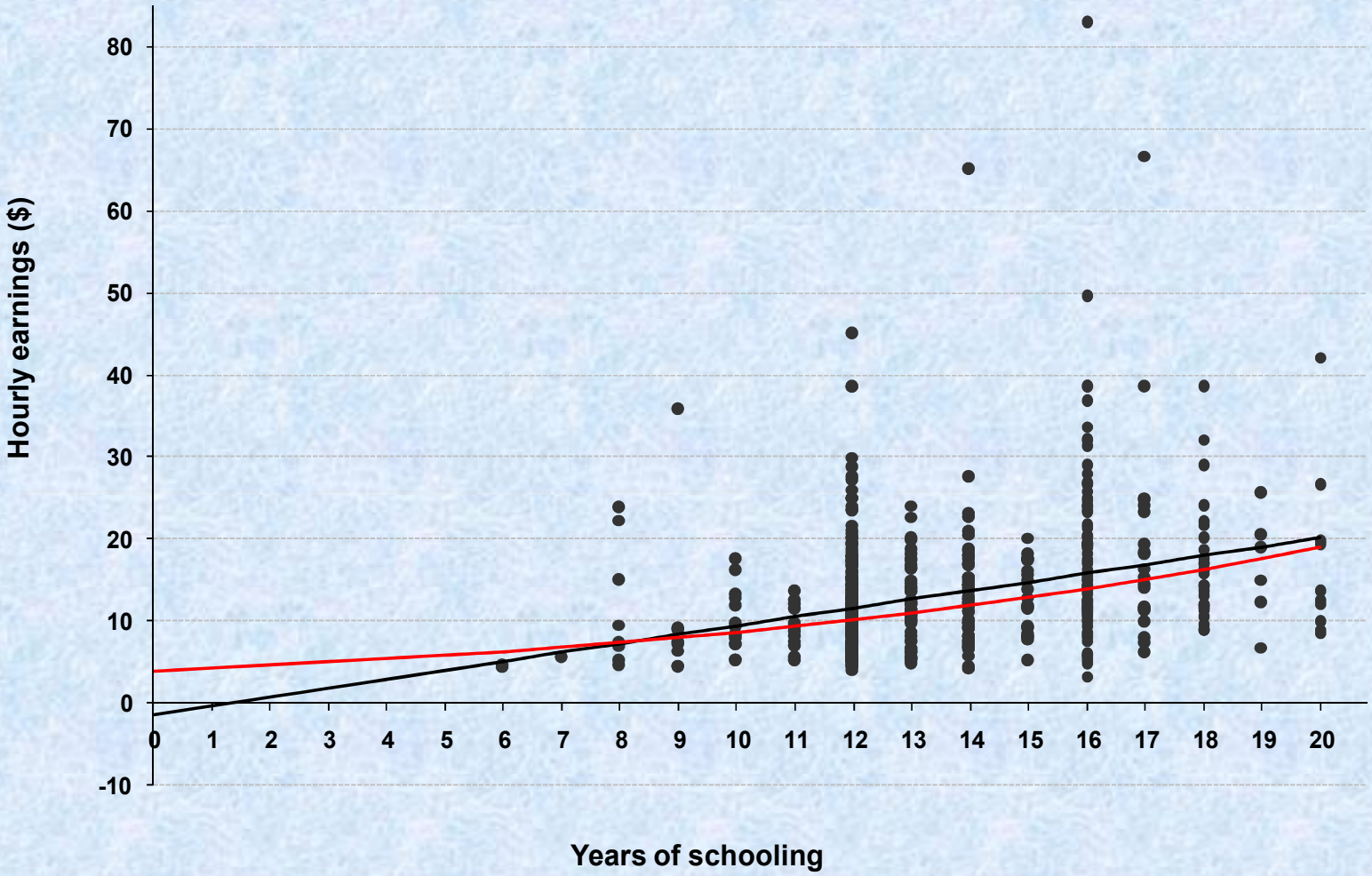
Должен ли индивид платить за право работы $1.99 в час, если он не имеет образования?.

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



Экстраполировать результаты эконометрического анализа далеко за пределы рабочей выборки нельзя!!! .

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ



Скорее всего зависимость почасовой заработной платы от количества лет обучения описывается нелинейным законом

МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Множественная регрессия имеет вид:

$$\mathrm{E}[Y / x_1, x_2, \ldots, x_m] = f(x_1, x_2, \ldots, x_m)$$

Уравнение множественной регрессии:

$$Y = f(\beta, \mathbf{X}) + \varepsilon$$

где $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ – вектор объясняющих переменных,

β – вектор параметров (подлежащих определению),

ε – вектор случайных ошибок (отклонений),

Y – зависимая переменная.

ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Теоретическое уравнение линейной множественной регрессии:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon$$

или для индивидуальных наблюдений:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

$i = 1, 2, \ldots, n$, $n \geq k$, $m = n–k$ – число степеней свободы

Для обеспечения статистической надежности должно выполняться условие: $n > 3k$

АНАЛИЗ ПРЕДЕЛЬНОГО ВКЛАДА ФАКТОРОВ

Множественная регрессия позволяет разложить суммарное влияние факторов на составные части, точнее выявив предельный вклад каждого фактора

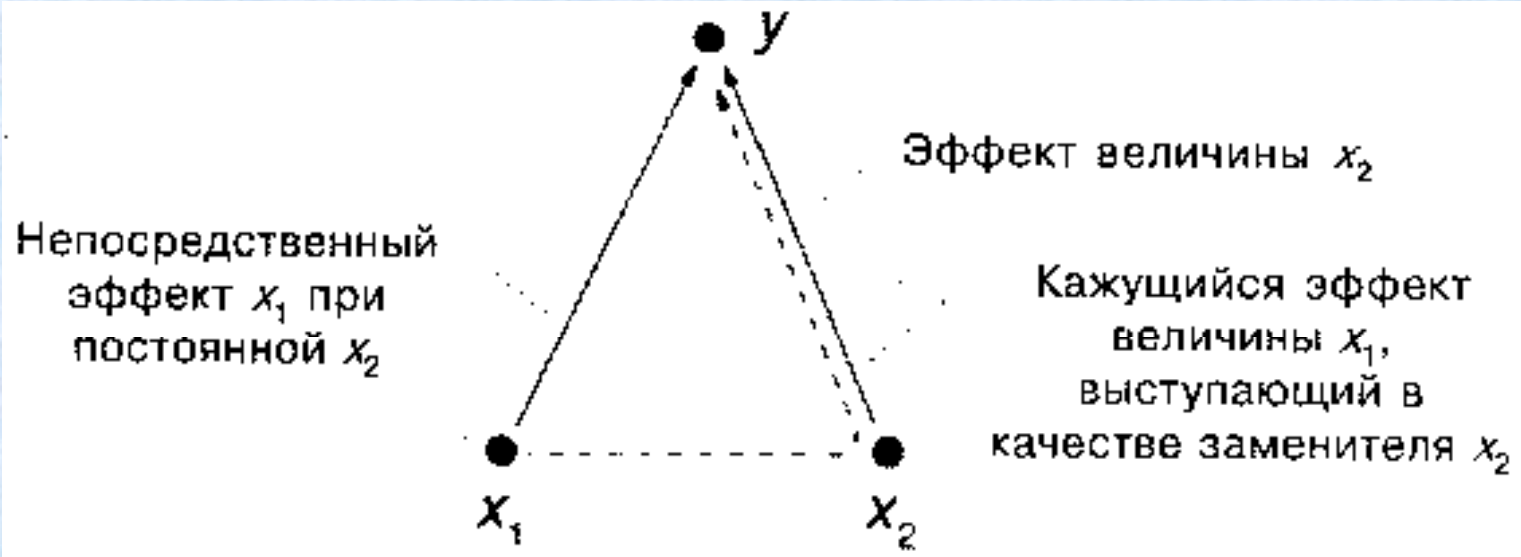
ИНТЕРПРЕТАЦИЯ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Интерпретация: коэффициент регрессии при переменной X_1 выражает предельный прирост зависимой переменной при изменении переменной X_1 , при условии постоянства других переменных:

$$\beta_2 = \frac{dY}{dX_2} \approx \frac{\Delta Y}{\Delta X_2}, \quad X_3 = const$$

ОСОБЕННОСТИ ПРОЯВЛЕНИЯ СВЯЗЕЙ В МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

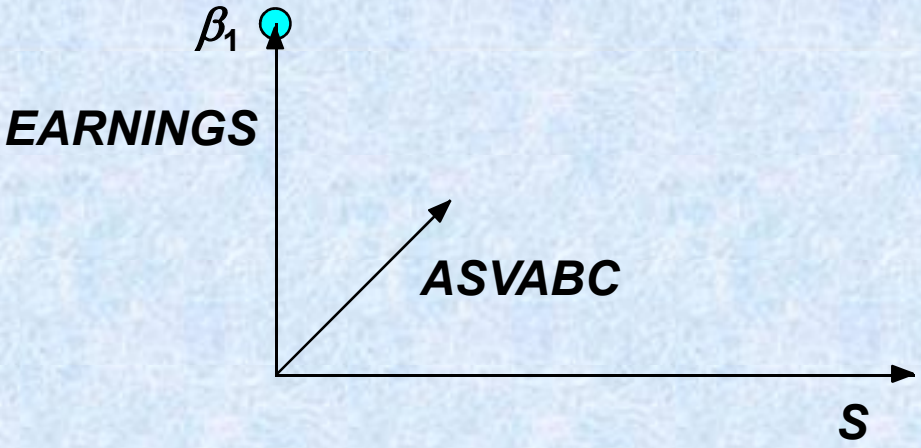


Из-за наличия вторичных связей качество оценок страдает - оценки оказываются менее эффективными.

В случае исключения значимой переменной X2 часть изменений Y за счет X2 будет приписана X1 , если переменная X1 может замещать X2. В результате оценка значения $\beta 1$ будет смещена.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 ASVABC + u$$

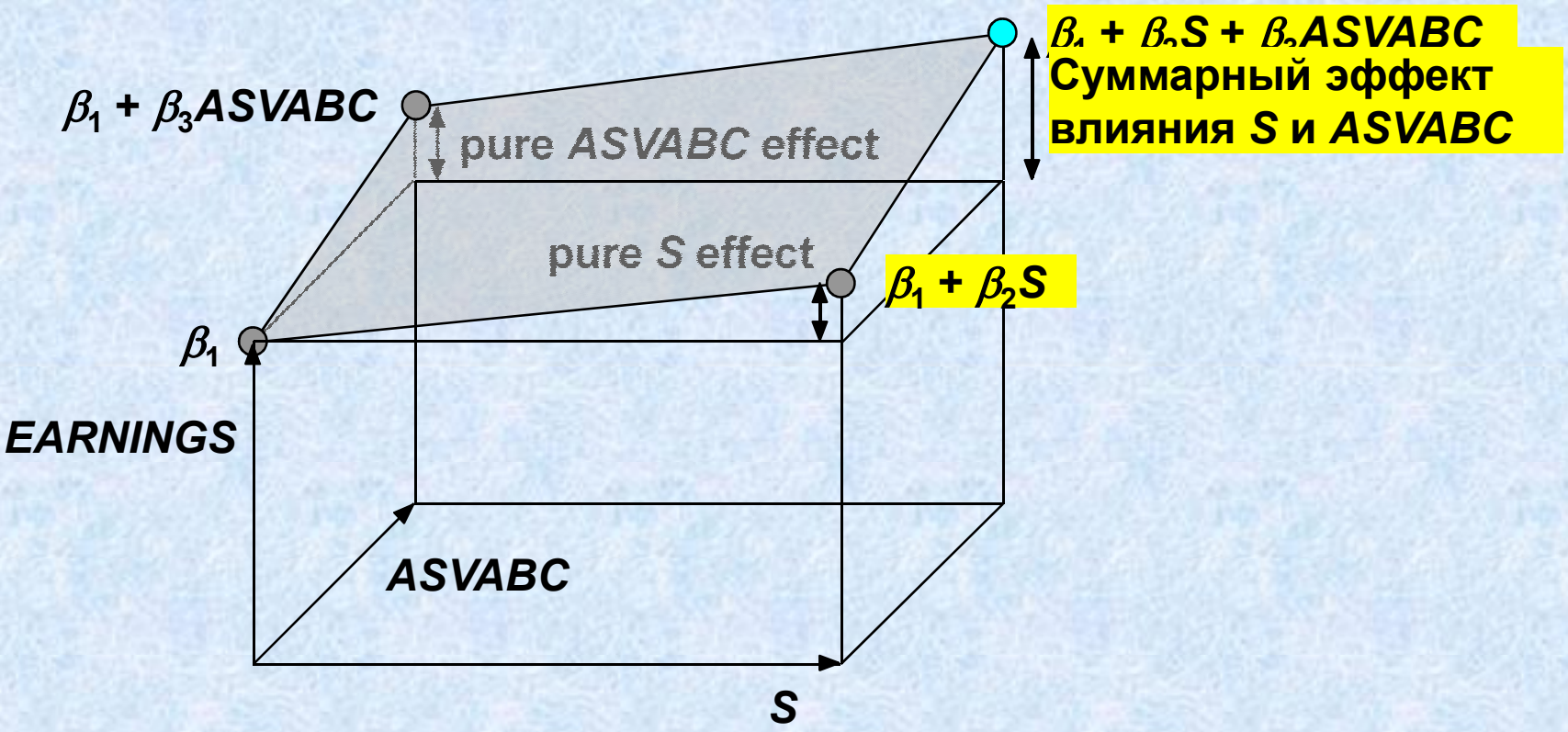


Геометрическая интерпретация разложения суммарного влияния на почасовую ставку заработной платы количества лет обучения и результатов теста на способности.

Константа β_1 соответствует ставке заработной платы тех респондентов, кто никогда не учился, и показал нулевые результаты по тесту

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

$EARNINGS = \beta_1 + \beta_2 S + \beta_3 ASVABC + u$

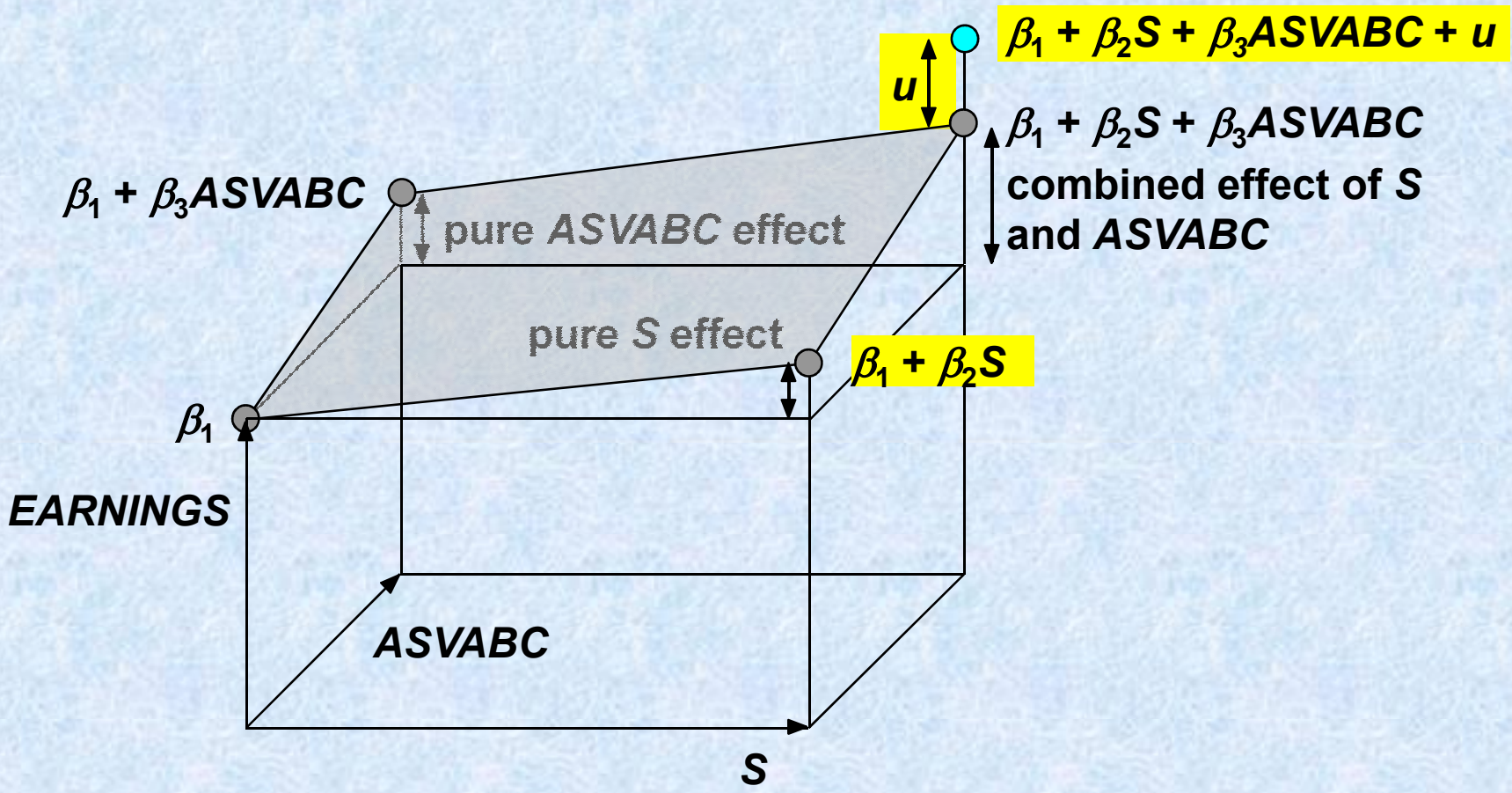


При различных сочетаниях величин факторов *S* и *ASVABC* будет определенный прирост почасовой заработной платы *EARNINGS* по сравнению со стартовой (для неспособных и не образованных) в соответствии с линейной связью: $EARNINGS = \beta_1 + \beta_2 S + \beta_3 ASVABC$.

Пока что мы считаем факторы некоррелированными (отсутствие вторичных связей)

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

$EARNINGS = \beta_1 + \beta_2 S + \beta_3 ASVABC + u$



Стохастическое слагаемое *u*, вызывает статистический разброс значений «наблюдаемой» заработной паты при одних и тех же параметрах.

Значение *u*, как всегда, ненаблюдаемо

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Dependent Variable: EARNINGS
Method: Least Squares
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
S	0.739037	0.160622	4.601103	0.0000
ASVABC	0.154534	0.042949	3.598121	0.0003
C	-4.624749	2.013200	-2.297213	0.0220

R-squared	0.123597	Mean dependent var	13.11782
Adjusted R-squared	0.120505	S.D. dependent var	8.214719
S.E. of regression	7.703877	Akaike info criterion	6.926574
Sum squared resid	33651.29	Schwarz criterion	6.949445
Log likelihood	-1971.073	F-statistic	39.98123
Durbin-Watson stat	1.962011	Prob(F-statistic)	0.000000

EARNINGS = 0.739*S + 0.155*ASVABC - 4.625

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Variable	Coefficient	Std. Error	t-Statistic	Prob.
S	0.739037	0.160622	4.601103	0.0000
ASVABC	0.154534	0.042949	3.598121	0.0003
C	-4.624749	2.013200	-2.297213	0.0220

S	1.073055	0.132450	8.101575	0.0000
C	-1.391004	1.820305	-0.764160	0.4451

	S	ASVABC
S	1.000000	0.577950
ASVABC	0.577950	1.000000

Завышенное влияние S из-за положительной корреляции с ASVABC, которая также влияет положительно

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Dependent Variable: EARNINGS
Method: Least Squares
Date: 09/21/08 Time: 13:37
Sample: 1 570
Included observations: 570

Остатки от этой регрессии:
EARN

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ASVABC	0.268743	0.035666	7.534995	0.0000
C	-0.359883	1.818571	-0.197893	0.8432

Dependent Variable: S
Method: Least Squares
Date: 09/21/08 Time: 13:43
Sample: 1 570
Included observations: 570

Остатки от этой регрессии:
ES

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ASVABC	0.154538	0.009156	16.87857	0.0000
C	5.770845	0.466847	12.36131	0.0000

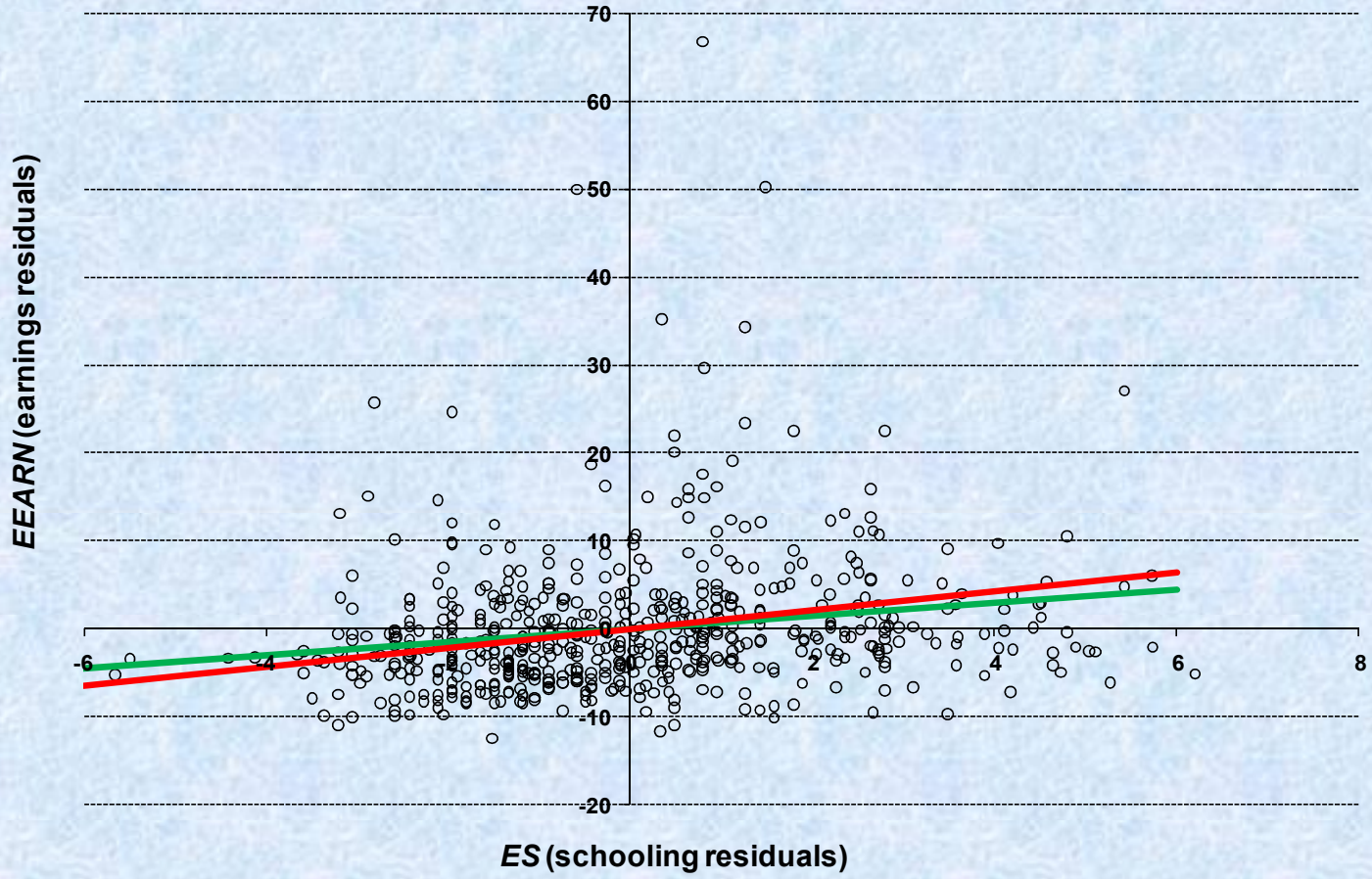
МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

EEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ES	.7390366	.1604802	4.605	0.000	.4238296	1.054244
_cons	-5.99e-09	.3223957	0.000	1.000	-.6332333	.6332333

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
S	.7390366	.1606216	4.601	0.000	.4235506	1.054523
ASVAB	.1545341	.0429486	3.598	0.000	.0701764	.2388918
_cons	-4.624749	2.0132	-2.297	0.022	-8.578989	-.6705095

Значения показателей в трехфакторной модели и в парной регрессии после элиминирования третьего фактора- равны

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ



красная линия тренда – трехфакторная регрессия
зеленая линия тренда – парная регрессия

Оценки параметров линейной множественной регрессии

Эмпирическое уравнение регрессии:

$$\hat{Y} = b_1 + b_2 X_2 + \ldots + b_k X_k$$

$$\hat{y}_i = b_1 + b_2 x_{i2} + \ldots + b_k x_{ik}$$

Самый распространенный метод оценки параметров – МНК

$$b_j,\ j = \overline{1,k}:\ \sum_{i=1}^{n} \underbrace{\left(y_i - (b_1 + \sum_{j=2}^{k} b_j x_{ij}) \right)}_{\hat{y}_i}^2 = \sum_{i=1}^{n} e_i^2 \to \min$$

Множественная регрессия
Метод МНК

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i}$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i}$$

$$RSS = \sum e_i^2 = \sum (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i})^2 \to \min$$

Значения коэффициентов в уравнении подогнанных значений определяются исходя из того же принципа минимизации суммы квадратов невязок между наблюдаемым значением и расчетным.

Множественная регрессия
Метод МНК

$$RSS = \sum e_i^2 = \sum (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i})^2$$

$$= \sum (Y_i^2 + b_1^2 + b_2^2 X_{2i}^2 + b_3^2 X_{3i}^2 - 2b_1 Y_i - 2b_2 X_{2i} Y_i$$

$$- 2b_3 X_{3i} Y_i + 2b_1 b_2 X_{2i} + 2b_1 b_3 X_{3i} + 2b_2 b_3 X_{2i} X_{3i})$$

$$= \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_{2i}^2 + b_3^2 \sum X_{3i}^2 - 2b_1 \sum Y_i$$

$$- 2b_2 \sum X_{2i} Y_i - 2b_3 \sum X_{3i} Y_i + 2b_1 b_2 \sum X_{2i}$$

$$+ 2b_1 b_3 \sum X_{3i} + 2b_2 b_3 \sum X_{2i} X_{3i}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \qquad \frac{\partial RSS}{\partial b_2} = 0 \qquad \frac{\partial RSS}{\partial b_3} = 0$$

Для нахождения кандидатов на роль оцененных коэффициентов используем условия первого порядка

Множественная регрессия
Метод МНК

$$b_1 = \overline{Y} - b_2\overline{X}_2 - b_3\overline{X}_3$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2}$$

$$b_3 = \frac{\mathrm{Cov}(X_3, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2}$$

На слайде представлены решения для трехфакторной модели. Обратите внимание на то, что принцип вычисления константы остался тем же, что и в парной модели (и с любым количеством факторов)

Не напоминают ли вам что-либо приводимые выражения?

Упражнение1: доказать формулы

Множественная регрессия
Метод МНК

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - \left[\mathrm{Cov}(X_2, X_3)\right]^2}$$

$$b_2 = \frac{\dfrac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3)} - \dfrac{\mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3)}}{1 - \dfrac{\left[\mathrm{Cov}(X_2, X_3)\right]^2}{\mathrm{Var}(X_2)\mathrm{Var}(X_3)}}$$

Если между регрессорами нет связи (коэффициент корреляции равен нулю), то коэффициент в множественной регрессии совпадает с коэффициентом в парной регрессии

Множественная регрессия
Метод МНК

$$\sqrt{\frac{\mathrm{Var}(X_2)}{\mathrm{Var}(Y)}} \cdot b_2 = \frac{\dfrac{\mathrm{Cov}(X_2, Y)}{\sqrt{\mathrm{Var}(X_2)\mathrm{Var}(Y)}} - \dfrac{\mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\sqrt{\mathrm{Var}(X_3)\mathrm{Var}(Y)}\sqrt{\mathrm{Var}(X_2)\mathrm{Var}(X_3)}}}{1 - \left[\mathrm{Corr}(X_2, X_3)\right]^2}$$

$$\sqrt{\frac{\mathrm{Var}(X_2)}{\mathrm{Var}(Y)}} \cdot b_2 = \frac{r_{YX_2} - r_{YX_3} r_{YX_3}}{1 - r^2_{X_2X_3}}$$

$$b_2 = r_{YX_2 \cdot X_3} \sqrt{\frac{\mathrm{Var}(Y)}{\mathrm{Var}(X_2)}}$$

Поскольку в множественной регрессии коэффициенты отражают связь каждого регрессора с зависимой переменной, то они пропорциональны частным коэффициентам корреляции

ЛИНЕЙНОСТЬ МОДЕЛИ РЕГРЕССИИ

- Линейность по регрессорам:

$$Y = \mathbf{X}\beta + u; \quad \frac{dY}{dX_i} = \beta_i$$

КОЭФФИЦИЕНТЫ ОТРАЖАЮТ ПРЕДЕЛЬНЫЕ ЭФФЕКТЫ!!!

REM 1: при нелинейном предикторе смысл коэффициентов иной!!!

REM 2: иногда возможна линеаризация модели:

$$Y = X_1\beta_1 + \ln(X_2)\beta_2 + u$$
$$Z_1 = X_1; \quad Z_2 = \ln(X_2)$$
$$Y = Z_1\beta_1 + Z_2\beta_2 + u$$

ЛИНЕЙНОСТЬ МОДЕЛИ РЕГРЕССИИ

- Линейность по параметрам:

$$Y = \mathbf{X}\beta + u; \quad \mathrm{E}\,u = \beta_0; \quad \mathrm{D}\,u = \sigma^2$$

$$\tilde{u} = \frac{u - \beta_0}{\sigma}; \quad \mathrm{E}\,\tilde{u} = 0; \quad \mathrm{D}\,\tilde{u} = 1$$

$$Y = \beta_0 + \mathbf{X}\beta + \sigma\tilde{u}$$

Модель линейна по коэффициентам и по дисперсии ошибок

REM 1:

Пример линеаризуемой модели:

$$Y = AK^{\beta_1}L^{\beta_2}u;$$

$$\tilde{Y} = \ln(Y); \quad \tilde{K} = \ln(K); \quad \tilde{L} = \ln(L);$$

$$\tilde{u} = \ln(u); \quad \tilde{\tilde{u}} = \frac{\tilde{u} - \mathrm{E}\,\tilde{u}}{\sqrt{\mathrm{D}\,\tilde{u}}};$$

$$\beta_0 = \ln A + \mathrm{E}\,\tilde{u}; \quad \sigma = \sqrt{\mathrm{D}\,\tilde{u}}$$

$$\tilde{Y} = \beta_0 + \tilde{K}\beta_1 + \tilde{L}\beta_2 + \sigma\tilde{\tilde{u}}$$

REM 2:

Пример нелинеаризуемой модели:

$$Y = K^{\beta_1}L^{\beta_2} + u$$

ЛИНЕЙНОСТЬ МОДЕЛИ РЕГРЕССИИ

- Линейность коэффициентов и предикторов (линейные формы от объясняемой переменной):

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y; \quad \hat{Y} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y$$

$$e = \mathbf{Y} - \hat{\mathbf{Y}} = \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)Y$$

Для парной регрессии

(все функции- линейные преобразования Y):

$$\hat{\beta} = \frac{\operatorname{cov}\left(X,Y\right)}{\operatorname{Var}\left(\mathrm{X}\right)}; \quad \hat{\alpha} = \bar{Y} - \bar{X}\frac{\operatorname{cov}\left(X,Y\right)}{\operatorname{Var}\left(\mathrm{X}\right)}$$

$$\hat{Y} = \bar{Y} - \left(\bar{X} - X\right)\frac{\operatorname{cov}\left(X,Y\right)}{\operatorname{Var}\left(\mathrm{X}\right)}$$

ОБЩЕЕ РЕШЕНИЕ ДЛЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

$$\text{MSPE (BLP)}: \quad \min_{\boldsymbol{\beta}} \mathrm{E}\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)^2$$

FOC (УСЛОВИЯ ПЕРВОГО ПОРЯДКА): $-2\,\mathrm{E}\,\mathbf{X}'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)=0$

SOC (УСЛОВИЯ ВТОРОГО ПОРЯДКА):

$\left(\mathbf{X}'\mathbf{X}\right)'=\mathbf{X}'\mathbf{X}$ симметрична $\Rightarrow$ пол. определена

РЕШЕНИЕ:

$$\mathrm{E}\left(\mathbf{X}'\mathbf{Y}\right)-\mathrm{E}\left(\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right)=0 \Leftrightarrow \mathbf{X}'\mathbf{Y}=\mathbf{X}'\mathbf{X}\cdot\boldsymbol{\beta}$$

Выборочные оценки:

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

РЕШЕНИЕ СУЩЕСТВУЕТ ТОЛЬКО ПРИ НЕОСОБЕННОЙ МАТРИЦЕ- РЕГРЕССОРЫ ДОЛЖНЫ БЫТЬ НЕЗАВИСИМЫ

Оценка параметров классической регрессионной модели МНК

Матричная форма СЛАУ:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_k \end{pmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ . & . & . & . \\ 1 & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

$$\mathbf{E} = \left(e_1\ e_2 \dots\ e_n\right)^T$$

Оценка параметров классической регрессионной модели МНК

$$\mathbf{X}^T\mathbf{X}\mathbf{B} = \mathbf{X}^T\mathbf{Y} \Rightarrow \mathbf{B} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum x_{i1} & \cdots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \cdots & \sum x_{i1}x_{ik} \\ . & . & . & . \\ \sum x_{ik} & \sum x_{i1}x_{ik} & \cdots & \sum x_{ik}^2 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \cdots \\ \sum y_i x_{im} \end{bmatrix}$$

ГЕОМЕТРИЧЕСКИЙ СМЫСЛ МЕТОДА OLS

- Вектор **Y** раскладывается на составляющие из непересекающихся подпространств- пространства регрессоров и ортогонального к нему (**остатки и регрессоры некоррелированны**):

$$Y = \mathbf{X}\boldsymbol{\beta} + u = \hat{Y} + e = \mathbf{X}\hat{\boldsymbol{\beta}} + e$$

$$\mathbf{FOC:}\quad \mathrm{E}\,\mathbf{X}'\left(Y - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = 0 \Leftrightarrow \mathrm{E}\,\mathbf{X}'e = \mathbf{0} \Leftrightarrow$$

$$\Leftrightarrow \mathbf{X}'\mathbf{e} = \sum_t \mathbf{X}'_{\cdot t}\left(Y - \mathbf{X}\hat{\boldsymbol{\beta}}\right)_t = \mathbf{X}'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) =$$

$$= \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{0} \Leftrightarrow$$

$$\Leftrightarrow \mathbf{X}'_i\mathbf{e} = 0_i \Leftrightarrow \sum_t X_{it}e_t = 0$$

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_N \end{bmatrix} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

ПРОЕКТОРЫ

Проектор **P –** проектор на пространство регрессоров:

$$\hat{Y} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y; \quad \mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'; \quad \hat{Y} = \mathbf{P}Y$$

Проектор **M** – проектор на пространство ортогональное регрессорам:

$$\mathbf{M} = \mathbf{I} - \mathbf{P}; \quad e = \mathbf{M}Y$$

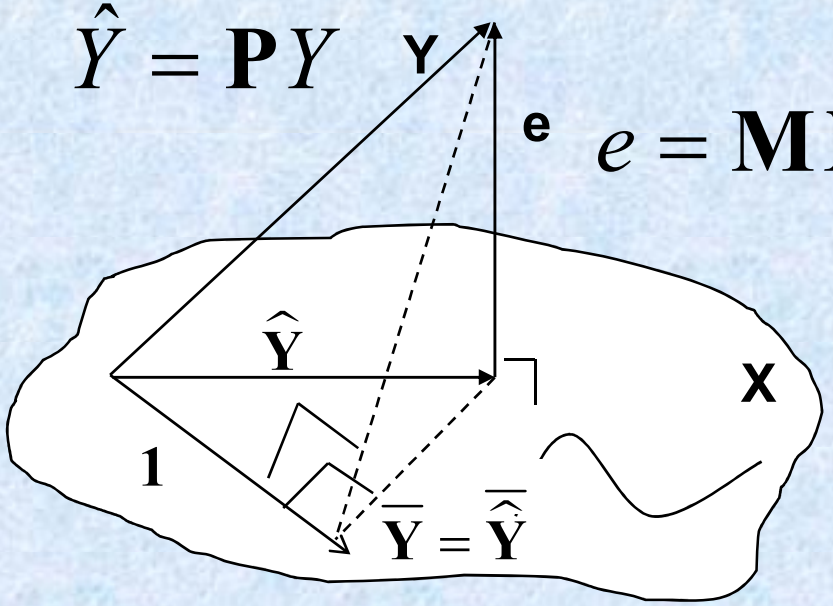
ПРИМЕР

$$\mathbf{X} = \mathbf{1}_{n\text{x}1}, \quad \mathbf{P} = \mathbf{1}\left(\mathbf{1}'\cdot\mathbf{1}\right)^{-1}\mathbf{1}' = \frac{1}{n}\cdot\mathbf{1}\cdot\mathbf{1}' = \frac{1}{n}\mathbf{1}_{n\text{x}n}; \quad \hat{\mathbf{Y}} = \mathbf{P}\mathbf{Y} = \overline{Y}\cdot\mathbf{1}$$

ГЕОМЕТРИЧЕСКИЙ СМЫСЛ МЕТОДА OLS

Если среди регрессоров есть константа, то
А) остатки в среднем равны нулю,

Б) среднее зависимой переменной и ее предсказанного значения- равны:



$$0 = \mathbf{1}'\mathbf{e} = [1 \ldots 1] \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} = \sum_i e_i$$

$$\sum_i Y_i = \mathbf{1}'\mathbf{Y} = \mathbf{1}'\left(\hat{\mathbf{Y}} + \mathbf{e}\right) = \sum_i \hat{Y}_i + \sum_i e_i = \sum_i \hat{Y}_i$$

Минимизация суммы квадратов остатков- поиск вектора наименьшей длины- это нормаль к пространству регрессоров

ПРОЕКТОРЫ

Проектор **P –** проектор на пространство регрессоров:

$$\hat{Y} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'Y; \quad \mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'; \quad \hat{Y} = \mathbf{P}Y$$

Проектор **M** – проектор на пространство ортогональное регрессорам:

$$\mathbf{M} = \mathbf{I} - \mathbf{P}; \quad e = \mathbf{M}Y$$

ПРИМЕР

$$\mathbf{X} = \mathbf{1}_{n\text{x}1}, \quad \mathbf{P} = \mathbf{1}\left(\mathbf{1}'\cdot\mathbf{1}\right)^{-1}\mathbf{1}' = \frac{1}{n}\cdot\mathbf{1}\cdot\mathbf{1}' = \frac{1}{n}\mathbf{1}_{n\text{x}n}; \quad \hat{\mathbf{Y}} = \mathbf{P}\mathbf{Y} = \overline{Y}\cdot\mathbf{1}$$

ПРОЕКТОРЫ

СВОЙСТВА ПРОЕКТОРОВ:

1. $\mathbf{M} + \mathbf{P} = \mathbf{I}$ полнота

2. $\mathbf{P}^2 = \mathbf{P}; \quad \mathbf{M}^2 = \mathbf{M}$ иденпотентность

3. $\mathbf{P}' = \mathbf{P}; \quad \mathbf{M}' = \mathbf{M}$ ортонормированность

4. $\mathbf{PM} = \mathbf{MP} = \mathbf{0}$ ортогональность

5. $\mathbf{PX} = \mathbf{X}; \quad \mathbf{P}\hat{Y} = \hat{Y}; \quad \mathbf{MX} = \mathbf{0}; \quad \mathbf{M}\hat{Y} = \mathbf{0}$

6. $\mathbf{M}u = e; \quad \mathbf{M}e = e; \quad \mathbf{P}e = \mathbf{0}$

ПРОЕКТОРЫ

СВОЙСТВА ПРОЕКТОРОВ:

Следствие :

a) $\mathbf{P}_{\mathbf{1}_{n\times 1}}\,\mathbf{e}=\mathbf{0}$

b) $\mathbf{X}'\mathbf{e}=\mathbf{X}'\mathbf{M}\mathbf{Y}=\mathbf{X}'\mathbf{M}'\mathbf{Y}=\left(\mathbf{M}\mathbf{X}\right)'\,\mathbf{Y}=\mathbf{0}$

c) $\mathbf{P}_{\mathbf{1}_{n\times 1}}\,Y=\hat{Y}=\overline{Y}\cdot\mathbf{1}_{n\times 1}, \quad \mathbf{P}_{\mathbf{1}_{n\times 1}}\,\hat{Y}=\overline{\hat{Y}}\cdot\mathbf{1}_{n\times 1}=\hat{Y}=\overline{Y}\cdot\mathbf{1}_{n\times 1}$

Если среди регрессоров есть константа, то

А) остатки и регрессоры- ортогональны

Б) остатки в среднем равны нулю

В) среднее зависимой переменной и ее предсказанного значения- равны

Конец лекции