распределение МНК-оценок коэффициентов в классической линейной модели

Лекция 7

СВЯЗЬ РАСПРЕДЕЛЕНИЙ ОЦЕНОК ПАРМЕТРОВ И ПРЕДСКАЗАННЫХ ЗНАЧЕНИЙ С РАСПРЕДЕЛЕНИЕМ ОШИБОК МОДЕЛИ

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

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

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

NB! РАСПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ ЗАВИСИТ ОТ СОВМЕСТНОГО РАСПРЕДЕЛЕНИЯ РЕГРЕССОРОВ И ОШИБОК

Если регрессоры не зависят от ошибок (в частности- неслучайны), а ошибки распределены нормально, то распределение оценок параметров, подогнанных значений и остатков - нормально

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Def Многомерная с.в. **Z** – **вектор,** каждая компонента которого – с.в.

Параметры распределения многомерной с.в. –это **вектор** средних, и **матрица** ковариаций

$$\boldsymbol{\mu} = \mathrm{E}\,\mathbf{Z} = \begin{pmatrix} \mathrm{E}\,Z_1 \\ \vdots \\ \mathrm{E}\,Z_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{Z}) = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1,n-1} & \sigma_{1n} \\ \sigma_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \sigma_{n-1,n} \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix} =$$

$$= \begin{bmatrix} \mathrm{Var}(Z_1) & \dots & \mathrm{cov}(Z_1, Z_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(Z_n, Z_1) & \dots & \mathrm{Var}(Z_n) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \cdots & \mathrm{corr}(Z_1, Z_n)\sigma_1\sigma_n \\ \vdots & \ddots & \vdots \\ \mathrm{corr}(Z_1, Z_n)\sigma_1\sigma_n & \dots & \sigma_n^2 \end{bmatrix}$$

Rem1 **Σ**- симметрична: **Σ´=Σ**

Rem2 **Σ**- положительна определена: для любого **Z** :**Z´ΣZ**≥0

Rem3 Существует $\Sigma^{-1}, \Sigma^{-1/2}, \Sigma^{1/2}$ (упражнение: найдите *доказательство*)

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Нормальный закон (скалярный случай):

$$\mathbf{Z} = Z, \quad \boldsymbol{\mu} = \mathrm{E}\, Z = \mu, \quad \boldsymbol{\Sigma} = \mathrm{Var}\left(\mathbf{Z}\right) = \sigma^2,$$

$$Z \sigma^2 Z = \sigma^2 Z^2 > 0$$

$$\xi \sim N\left(\mu, \sigma^2\right), \quad \mathrm{f}_{\xi}\left(x\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{\left(x-\mu\right)\left(\sigma^2\right)^{-1}\left(x-\mu\right)}{2}\right]$$

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Плотность многомерного нормального распределения:

$$\mathrm{f}_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}(\det\boldsymbol{\Sigma})^{1/2}} \cdot \exp\left[-\frac{(\mathbf{z}-\boldsymbol{\mu})'\,\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}{2}\right]$$

Если **Σ**=diag, то $\mathrm{f}_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma_i} \cdot \exp\left[-\frac{(z_i-\mu_i)^2}{2\sigma_i^2}\right]\right) = \prod_{i=1}^{n}\mathrm{f}_{Z_i}(z_i)$

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Нормальный закон (двумерный случай)

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \mathrm{E}\,\mathbf{Z} = \begin{pmatrix} \mathrm{E}\,Z_1 \\ \mathrm{E}\,Z_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{Z}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(Z_1) & \mathrm{cov}(Z_1, Z_2) \\ \mathrm{cov}(Z_2, Z_1) & \mathrm{Var}(Z_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)} \begin{bmatrix} \sigma_2^1 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}, \quad \textit{Найдите}\text{: } \boldsymbol{\Sigma}^{1/2}, \boldsymbol{\Sigma}^{-1/2}$$

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Нормальный закон (двумерный случай)

$$\boldsymbol{\xi} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma}),\quad \boldsymbol{\xi}=(\xi_1,\xi_1)\sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$$

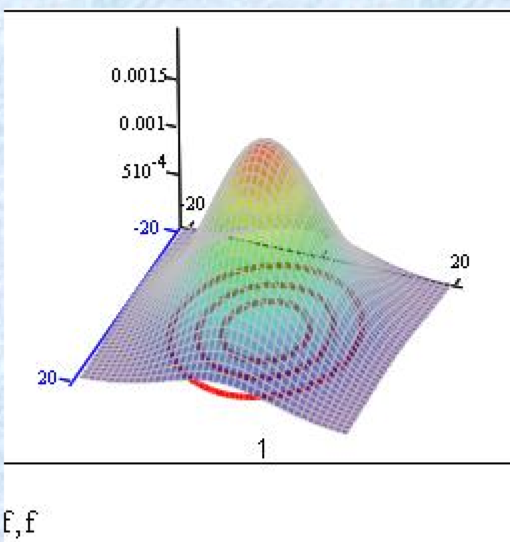
$$\mathrm{f}_{\xi}(z_1,z_2)=\frac{1}{(2\pi)^{2/2}\sqrt{(1-\rho^2)\sigma_1^2\sigma_2^2}}\cdot\exp\left[-\frac{\begin{pmatrix}z_1-\mu_1\\ z_2-\mu_2\end{pmatrix}'\begin{bmatrix}\sigma_2^1 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}\begin{pmatrix}z_1-\mu_1\\ z_2-\mu_2\end{pmatrix}}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right]$$

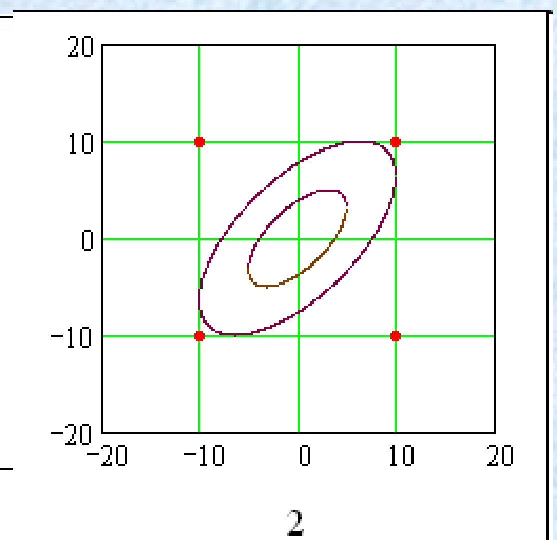
$$\mathrm{f}_{\xi}(z_1,z_2)=\frac{1}{(2\pi)^{2/2}\sqrt{(1-\rho^2)\sigma_1^2\sigma_2^2}}\cdot\exp\left[-\frac{\left(\frac{z_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{z_1-\mu_1}{\sigma_1}\right)\left(\frac{z_2-\mu_2}{\sigma_2}\right)+\left(\frac{z_2-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]$$

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Нормальный закон (двумерный случай)

$$\mathrm{f}_{\xi}(z_1, z_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_1^2\sigma_2^2}} \cdot \exp\left[-\frac{\begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}' \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}}{2}\right]$$





МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Линейное преобразование многомерного нормального вектора:

$$\mathbf{Z} \sim \mathrm{N}\left(\boldsymbol{\mu}, \Sigma\right)$$

$$\mathbf{X} = \boldsymbol{\Lambda} + \mathbf{B}\mathbf{Z} \sim \mathrm{N}\left(\boldsymbol{\Lambda} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}'\right)$$ -

нормальное распределение устойчиво по аффиным преобразованиям

Упражнение:

Докажите эту формулу

Аналогии:

Одномерный случай:

$$Z \sim \mathrm{N}\left(\mu, \sigma^2\right)$$

$$X = \alpha + \beta Z \sim \mathrm{N}\left(\alpha + \beta\mu, \beta\sigma^2\beta\right)$$

СВЯЗЬ РАСПРЕДЕЛЕНИЙ ОЦЕНОК ПАРМЕТРОВ И ПРЕДСКАЗАННЫХ ЗНАЧЕНИЙ С РАСПРЕДЕЛЕНИЕМ ОШИБОК МОДЕЛИ

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

$$\widehat{\mathbf{Y}} = \mathbf{PY}$$

$$\mathbf{e} = \mathbf{MY}$$

NB! РАСПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ ЗАВИСИТ ОТ СОВМЕСТНОГО РАСПРЕДЕЛЕНИЯ РЕГРЕССОРОВ И ОШИБОК

Если регрессоры не зависят от ошибок (в частности-неслучайны), а ошибки распределены нормально, то распределение оценок параметров, подогнанных значений и остатков - нормально

РАСПРЕДЕЛЕНИЕ ОЦЕНОК КОЭФФИЦИЕНТОВ

$$\varepsilon \in \mathrm{N}\left(\mathbf{0};\sigma^2\mathbf{I}\right), \quad \boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \sigma^2\mathbf{I},$$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon, \quad \boldsymbol{\Lambda} = \boldsymbol{\beta}, \quad \mathbf{B} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

РАСПРЕДЕЛЕНИЕ ОЦЕНОК КОЭФФИЦИЕНТОВ

$$\varepsilon \in \mathrm{N}\left(\mathbf{0};\sigma^2\mathbf{I}\right),\quad \boldsymbol{\mu}=\mathbf{0},\quad \boldsymbol{\Sigma}=\sigma^2\mathbf{I},$$

$$\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon,\quad \boldsymbol{\Lambda}=\boldsymbol{\beta},\quad \mathbf{B}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

$$\hat{\boldsymbol{\beta}}\in \mathrm{N}\left(\boldsymbol{\beta}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\cdot\mathbf{0};\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\sigma^2\mathbf{I}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)'\right)$$

$$=\mathrm{N}\left(\boldsymbol{\beta};\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

$$\mathrm{E}\left(\hat{\boldsymbol{\beta}}\right)=\boldsymbol{\beta}$$

$$\mathrm{COV}\left(\hat{\boldsymbol{\beta}}\right)=\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}=\frac{\sigma^2}{N}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}$$

РАСПРЕДЕЛЕНИЕ ПРЕДСКАЗАННЫХ ЗНАЧЕНИЙ

$$\varepsilon \in \mathrm{N}\left(\mathbf{0};\sigma^2\mathbf{I}\right), \quad \boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \sigma^2\mathbf{I},$$

$$\widehat{\mathbf{Y}} = \mathbf{PY} = \mathbf{PX\beta} + \mathbf{P\varepsilon} = \mathbf{X\beta} + \mathbf{P\varepsilon}, \quad \boldsymbol{\Lambda} = \mathbf{X\beta}, \quad \mathbf{B} = \mathbf{P}$$

$$\widehat{\mathbf{Y}} \in \mathrm{N}\left(\mathbf{X\beta} + \mathbf{P}\cdot\mathbf{0};\mathbf{P}\cdot\sigma^2\mathbf{I}\cdot\mathbf{P}'\right) = \mathrm{N}\left(\mathbf{X\beta} + \mathbf{P}\cdot\mathbf{0};\sigma^2\mathbf{P}\right)$$

$$= \mathrm{N}\left(\mathbf{X\beta};\sigma^2\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)$$

$$\mathrm{E}\left(\widehat{\mathbf{Y}}\right) = \mathbf{X\beta}$$

$$\mathrm{COV}\left(\widehat{\mathbf{Y}}\right) = \sigma^2\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \frac{\sigma^2}{N}\mathbf{X}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\mathbf{X}'$$

РАСПРЕДЕЛЕНИЕ ОСТАТКОВ

$$\varepsilon \in \mathrm{N}\left(\mathbf{0};\sigma^2\mathbf{I}\right),\quad \boldsymbol{\mu} = \mathbf{0},\quad \boldsymbol{\Sigma} = \sigma^2\mathbf{I},$$

$$\mathbf{e} = \mathbf{MY} = \mathbf{MX\beta} + \mathbf{M\varepsilon} = \mathbf{M\varepsilon},\quad \boldsymbol{\Lambda} = \mathbf{0},\quad \mathbf{B} = \mathbf{M}$$

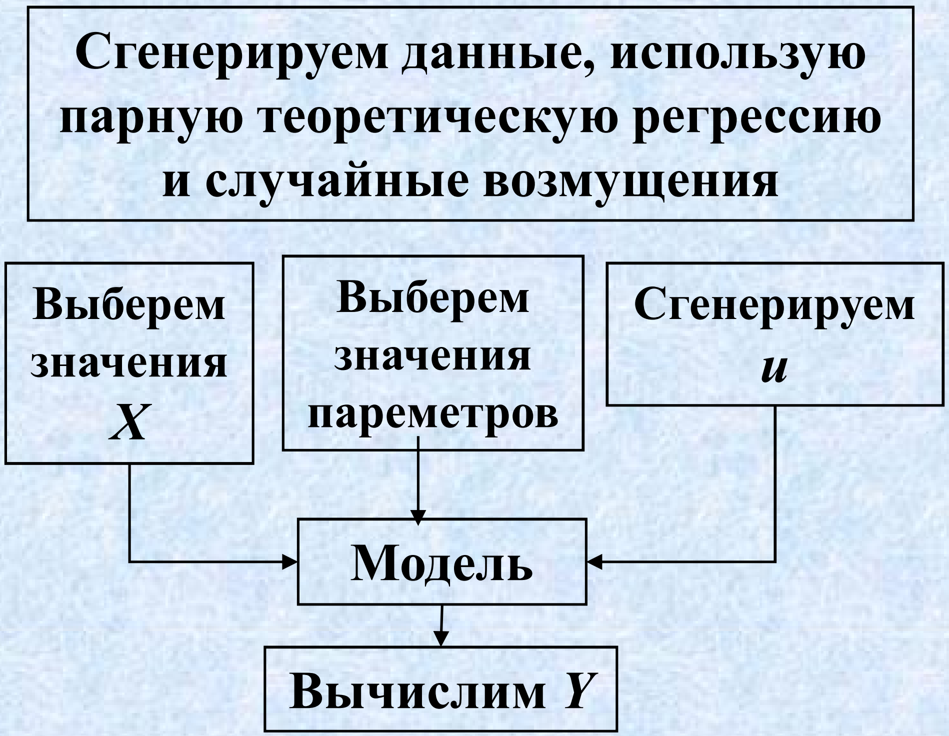
$$\mathbf{e} \in \mathrm{N}\left(\mathbf{0} + \mathbf{M}\cdot\mathbf{0};\mathbf{M}\cdot\sigma^2\mathbf{I}\cdot\mathbf{M}'\right)$$

$$= \mathrm{N}\left(\mathbf{0};\sigma^2\mathbf{M}\right)$$

$$\mathrm{E}\left(\mathbf{e}\right) = \mathbf{0}$$

$$\mathrm{COV}\left(\mathbf{e}\right) = \sigma^2\mathbf{M} = \sigma^2\left(\mathbf{I} - \frac{1}{N}\mathbf{X}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\mathbf{X}'\right)$$

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



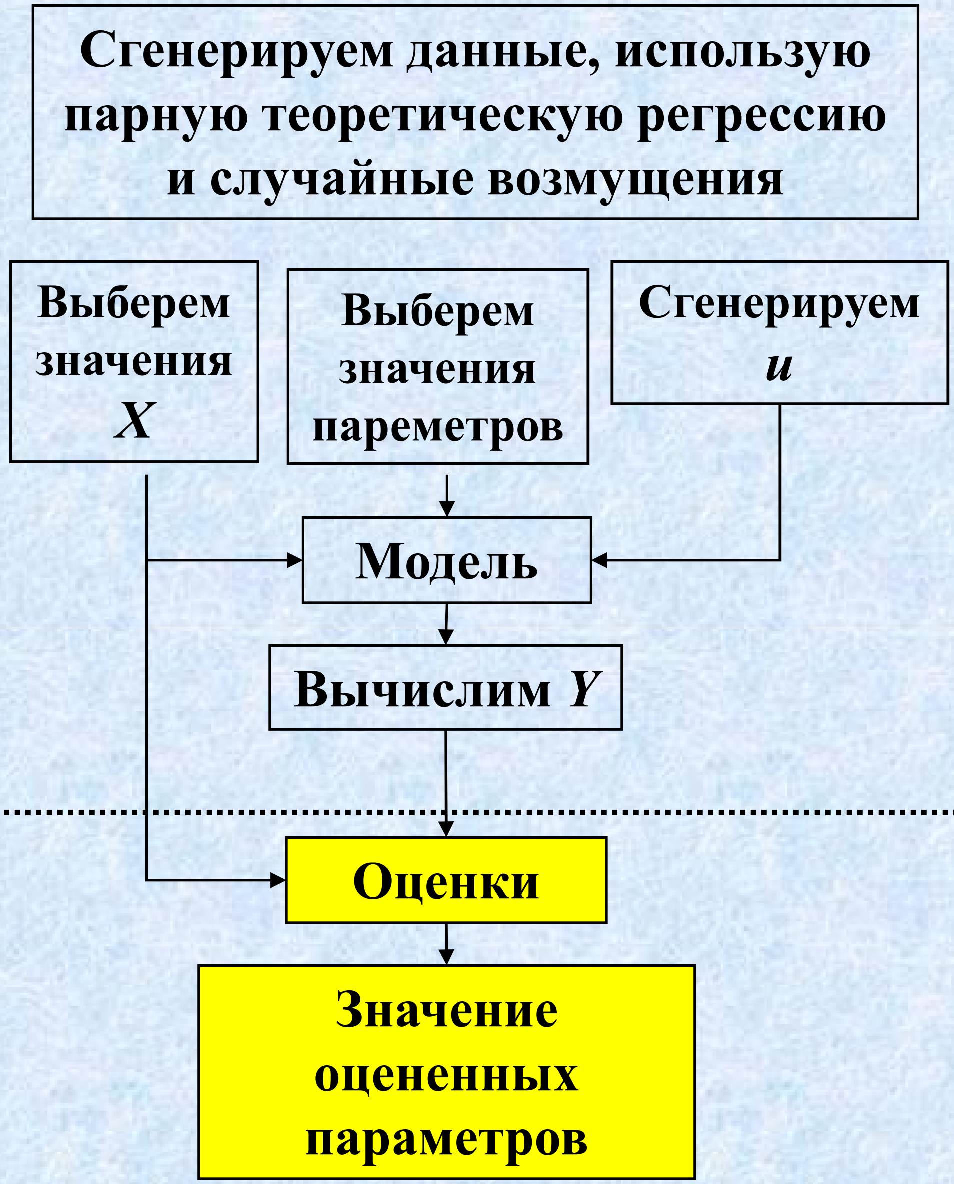
Применим метод Монте-Карло к исследованию распределения оценок коэффициентов парной регрессии.

Выберем функциональную зависимость между *Y* и *X* (выберем коэффициенты теоретической регрессии).

Случайные ошибки *u* будем генерировать из нормальной совокупности.

***«Наблюдаемые»* значения *Y* будем получать как сумму функциональной зависимости и случайной ошибки.**

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

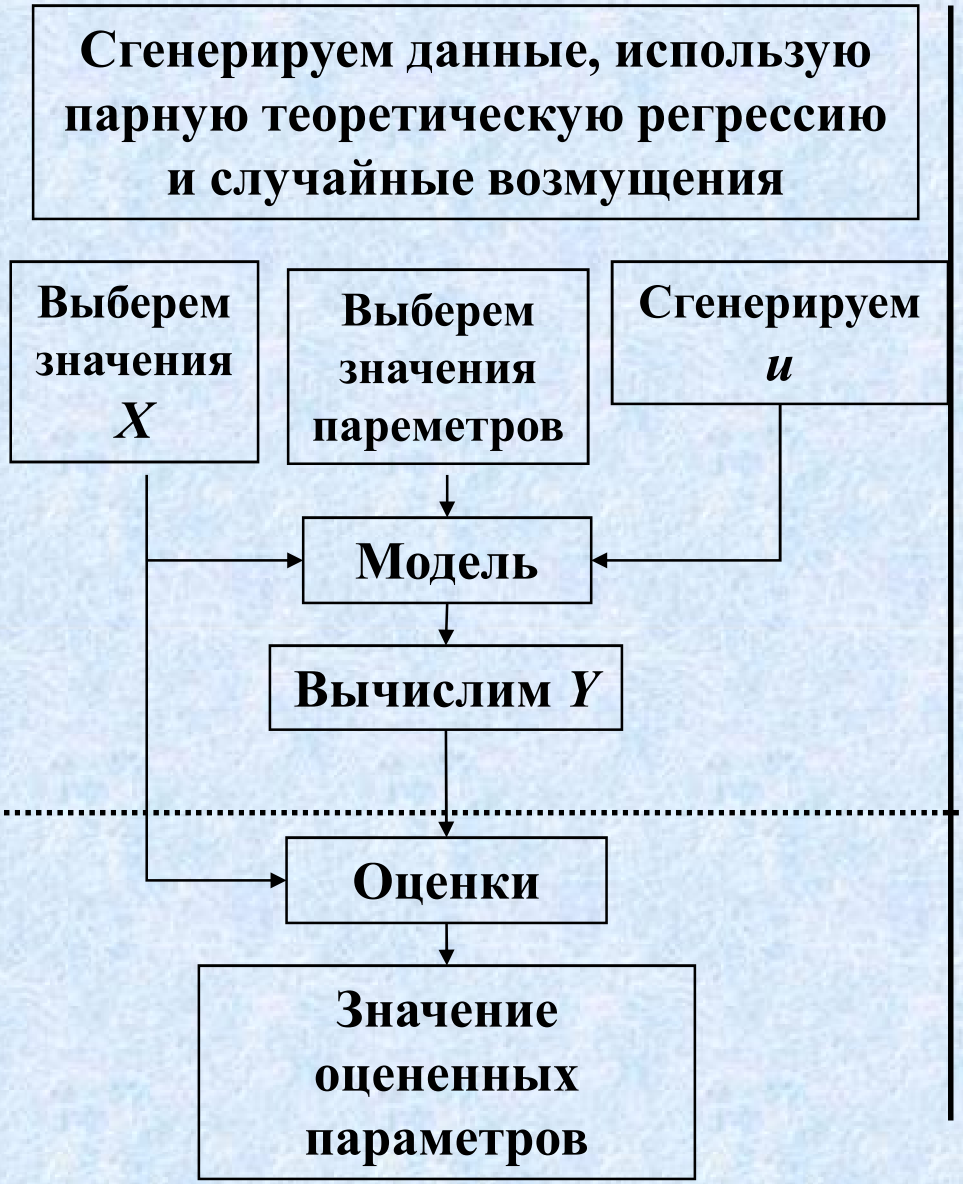


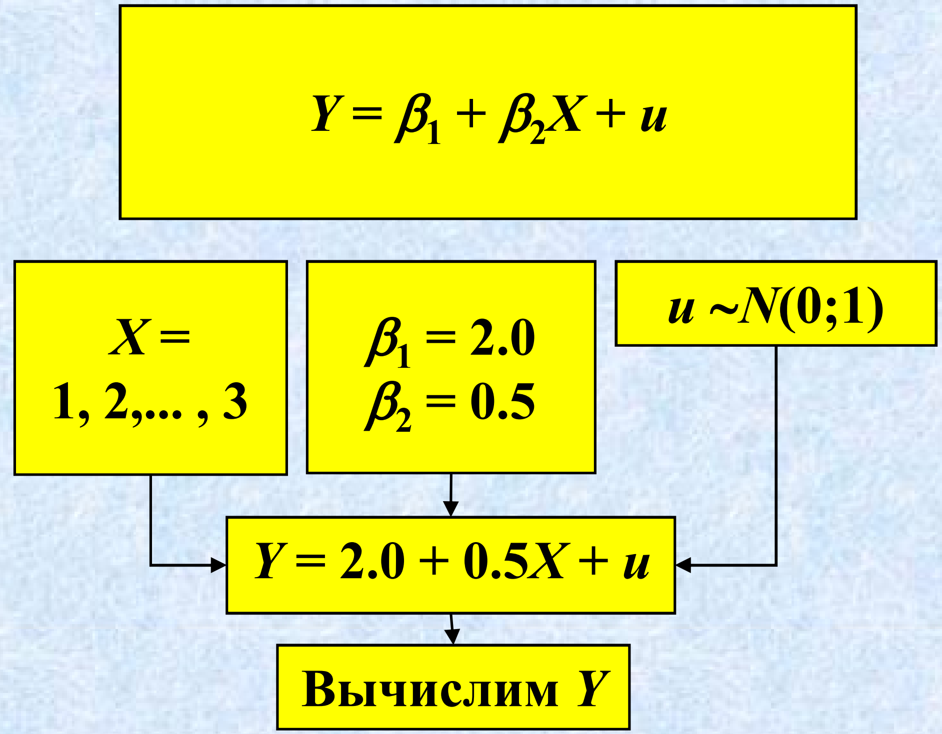
Оценим коэффициенты регрессии только по «*наблюдаемым*» значениям *Y* и значениям *X*.

Повторяя процесс для тех же самых *X* , но нового рядя случайных ошибок, будем получать новые значения оценок коэффициентов теоретической регрессии.

Проверим статистически нормальность, несмещенность и состоятельность оценок.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ





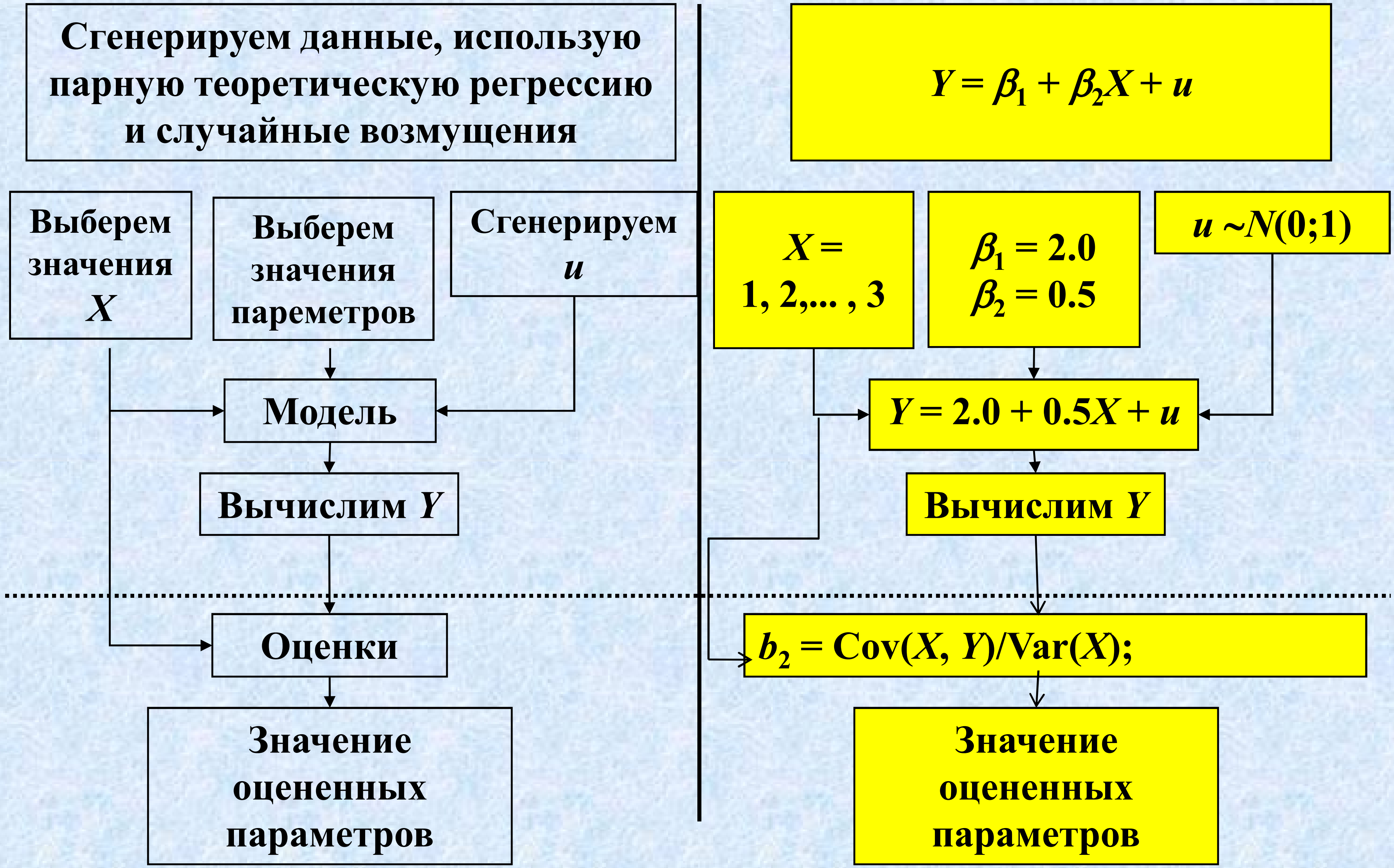
В нашем эксперименте Монте-Карло совокупность состоит из 20 наблюдений.

X **принимает значения 1, 2, ..., 20.**

$\beta_1 = 2.0$; $\beta_2 = 0.5$.

Случайные ошибки выбираются из стандартного нормального распределения

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

$$Y = 2.0 + 0.5X + u$$

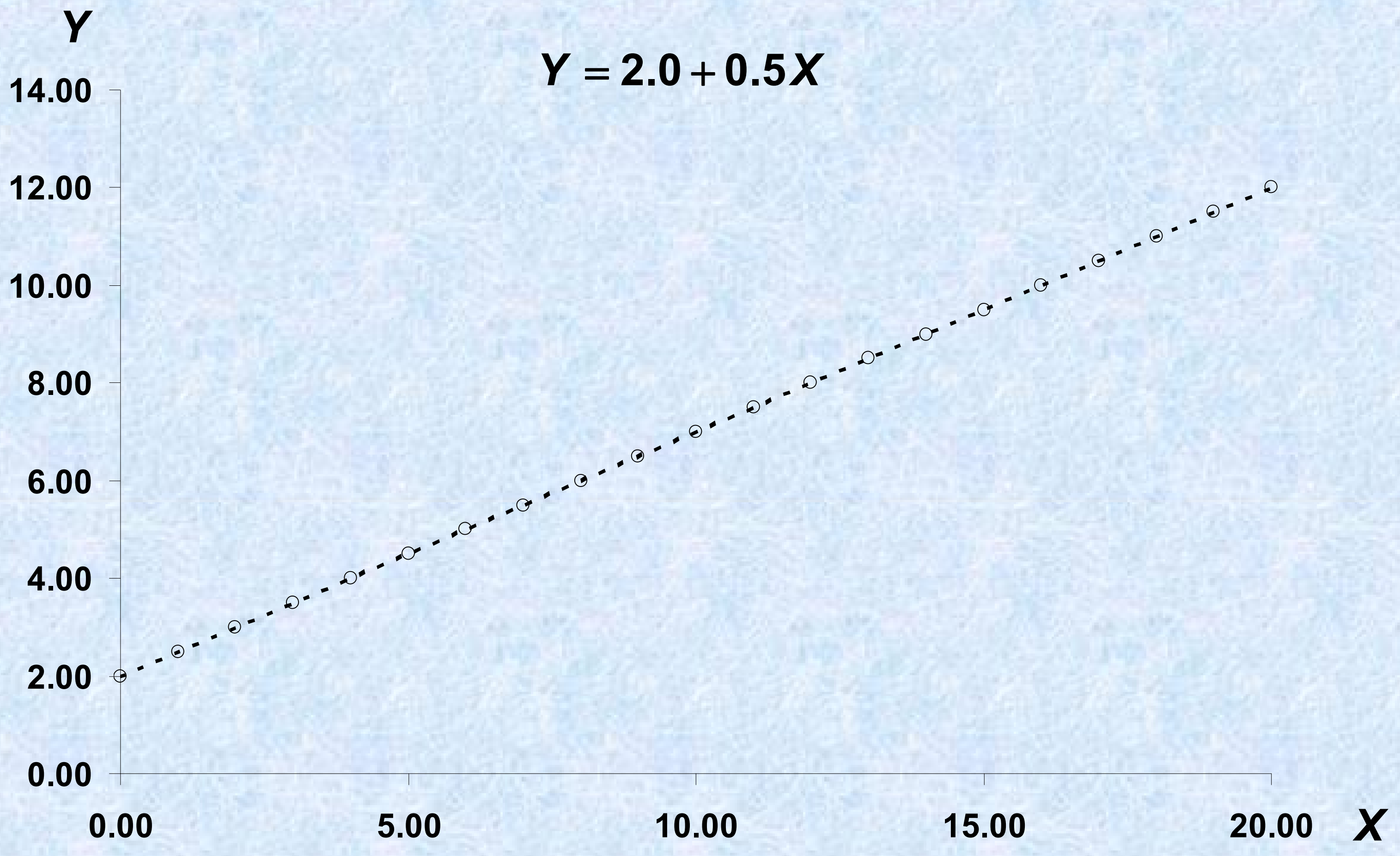
X	$2.0+0.5X$	u	Y	X	$2.0+0.5X$	u	Y
1				11			
2				12			
3				13			
4				14			
5				15			
6				16			
7				17			
8				18			
9				19			
10				20			

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

$$Y = 2.0 + 0.5X + u$$

X	$2.0+0.5X$	u	Y	X	$2.0+0.5X$	u	Y
1	2.5			11	7.5		
2	3.0			12	8.0		
3	3.5			13	8.5		
4	4.0			14	9.0		
5	4.5			15	9.5		
6	5.0			16	10.0		
7	5.5			17	10.5		
8	6.0			18	11.0		
9	6.5			19	11.5		
10	7.0			20	12.0		

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



Это линия теоретической регрессии.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

$$Y = 2.0 + 0.5X + u$$

X	$2.0+0.5X$	u	Y	X	$2.0+0.5X$	u	Y
1	2.5	-0.59		11	7.5	1.59	
2	3.0	-0.24		12	8.0	-0.92	
3	3.5	-0.83		13	8.5	-0.71	
4	4.0	0.03		14	9.0	-0.25	
5	4.5	-0.38		15	9.5	1.69	
6	5.0	-2.19		16	10.0	0.15	
7	5.5	1.03		17	10.5	0.02	
8	6.0	0.24		18	11.0	-0.11	
9	6.5	2.53		19	11.5	-0.91	
10	7.0	-0.13		20	12.0	1.42	

Случайные ошибки сгенерированы с помощью стандартного нормального закона $N(0,1)$.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

$$Y = 2.0 + 0.5X + u$$

X	$2.0+0.5X$	u	Y	X	$2.0+0.5X$	u	Y
1	2.5	-0.59	1.91	11	7.5	1.59	
2	3.0	-0.24		12	8.0	-0.92	
3	3.5	-0.83		13	8.5	-0.71	
4	4.0	0.03		14	9.0	-0.25	
5	4.5	-0.38		15	9.5	1.69	
6	5.0	-2.19		16	10.0	0.15	
7	5.5	1.03		17	10.5	0.02	
8	6.0	0.24		18	11.0	-0.11	
9	6.5	2.53		19	11.5	-0.91	
10	7.0	-0.13		20	12.0	1.42	

«*Наблюдаемые*» значения *Y* отличаются от «*теоретических*», например-1.91 вместо 2.50.

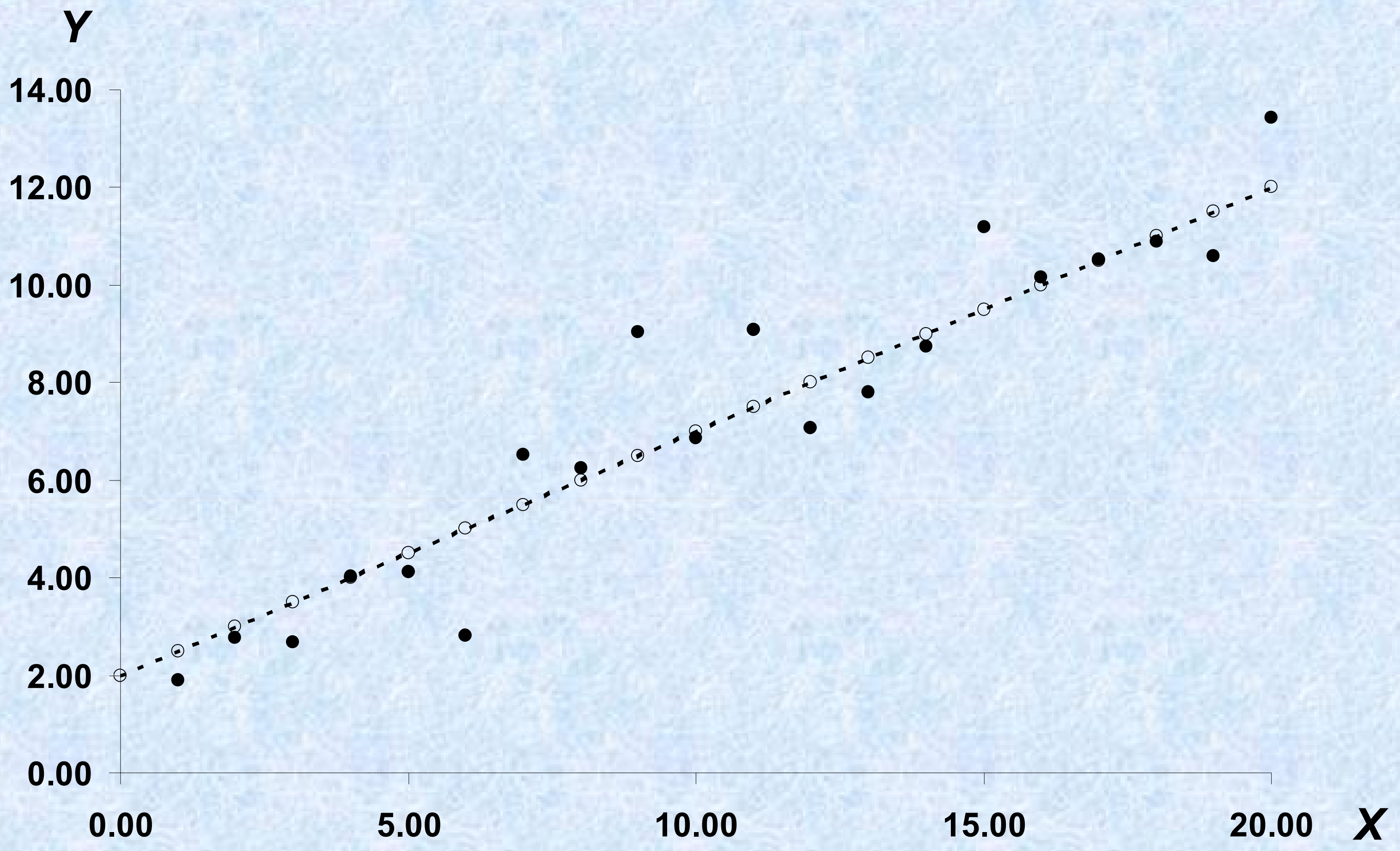
КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

$$Y = 2.0 + 0.5X + u$$

X	$2.0+0.5X$	u	Y	X	$2.0+0.5X$	u	Y
1	2.5	-0.59	1.91	11	7.5	1.59	9.09
2	3.0	-0.24	2.76	12	8.0	-0.92	7.08
3	3.5	-0.83	2.67	13	8.5	-0.71	7.79
4	4.0	0.03	4.03	14	9.0	-0.25	8.75
5	4.5	-0.38	4.12	15	9.5	1.69	11.19
6	5.0	-2.19	2.81	16	10.0	0.15	10.15
7	5.5	1.03	6.53	17	10.5	0.02	10.52
8	6.0	0.24	6.24	18	11.0	-0.11	10.89
9	6.5	2.53	9.03	19	11.5	-0.91	10.59
10	7.0	-0.13	6.87	20	12.0	1.42	13.42

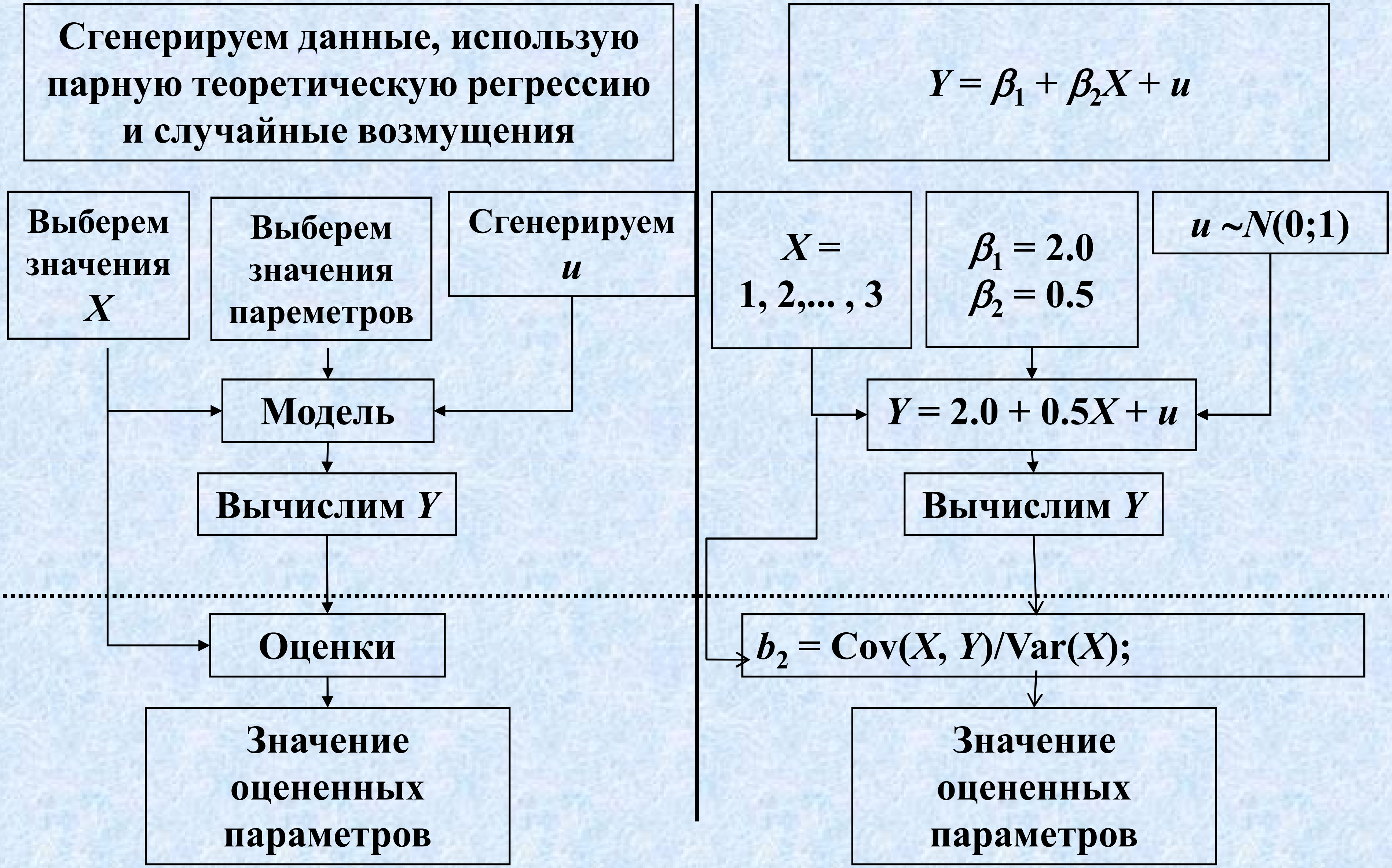
Аналогично для остальных 19 наблюдений.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



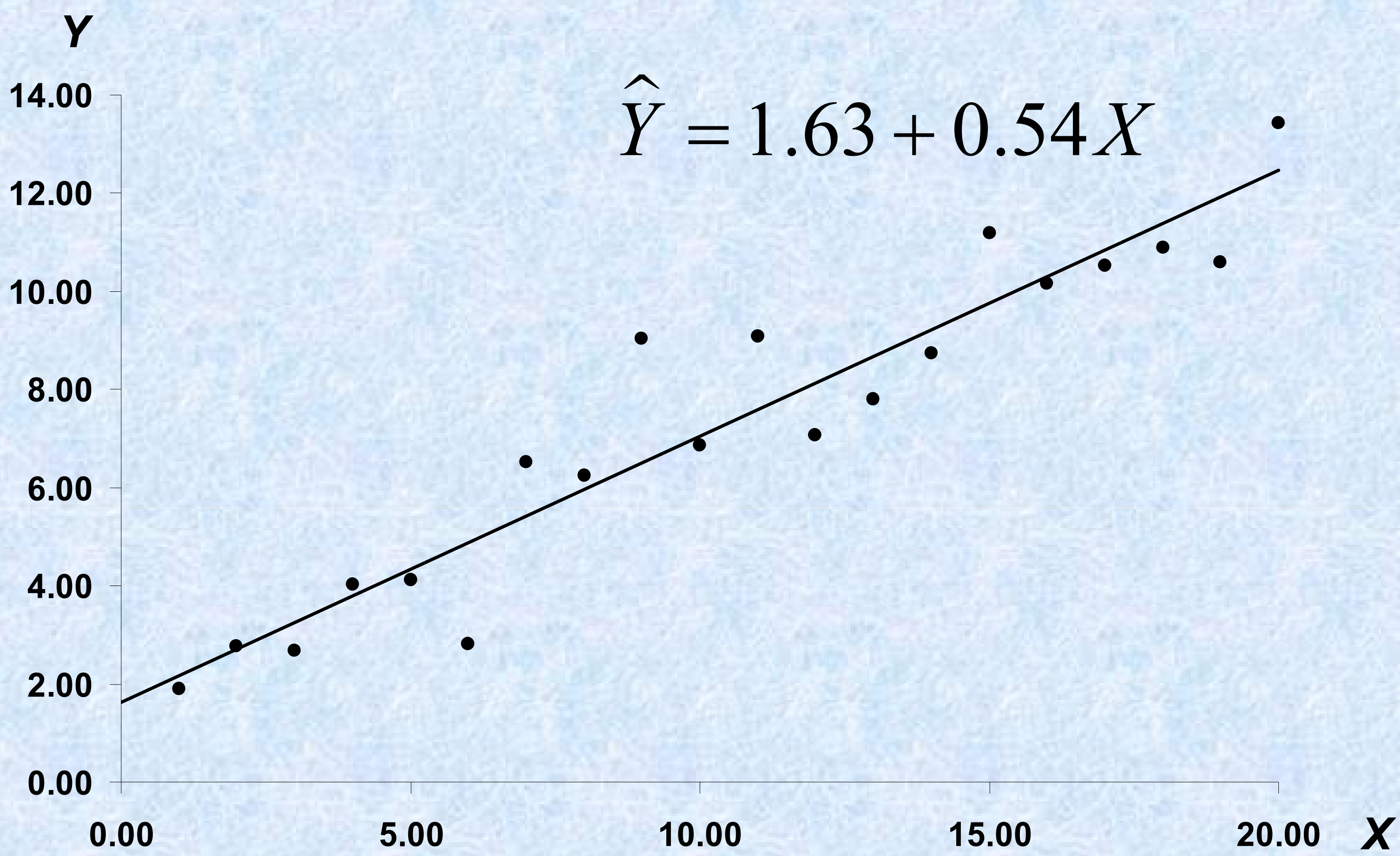
Связь теоретической регрессии и «*наблюдаемых*» данных.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



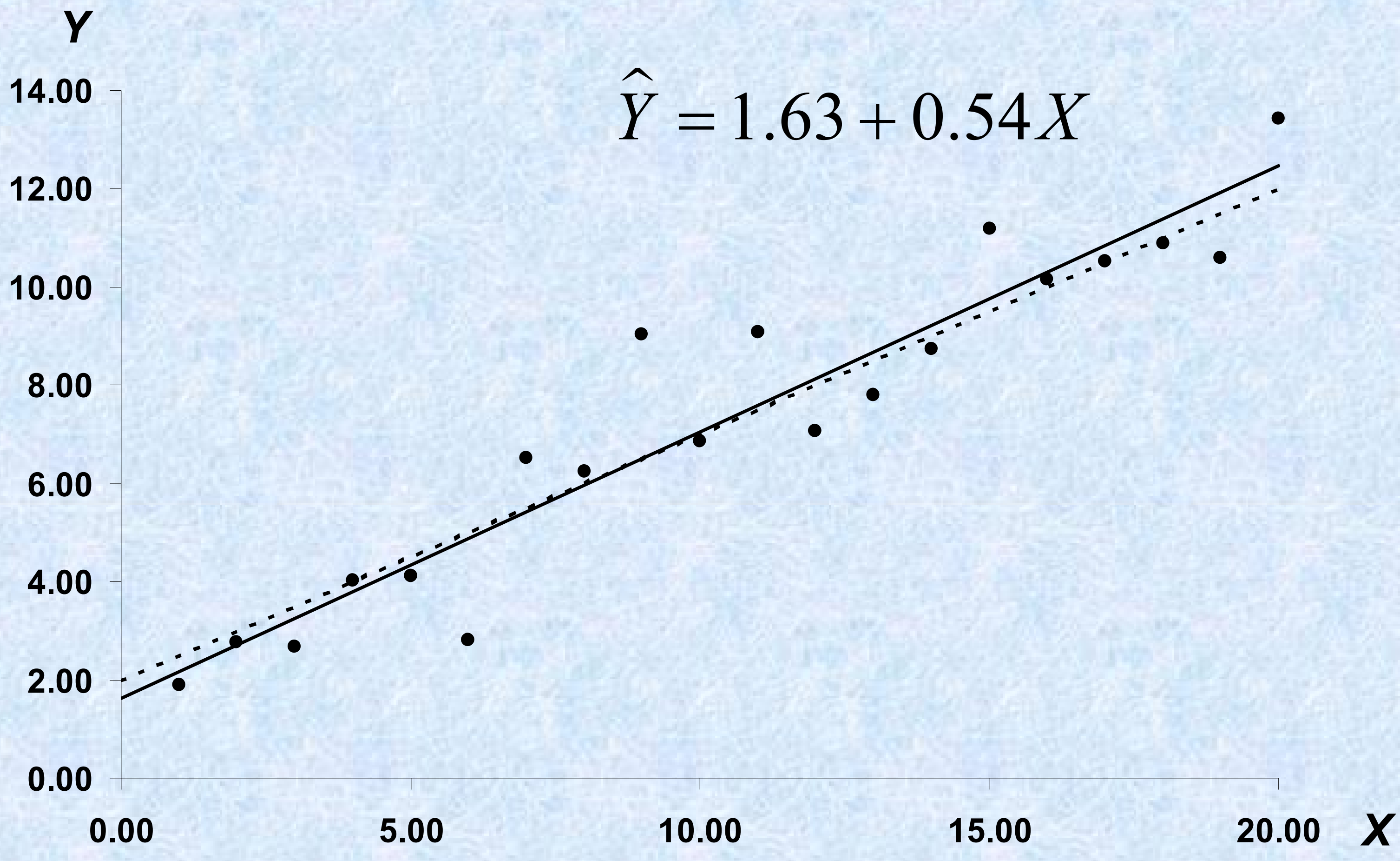
Повторим эксперимент для новых сгенерированных ошибок, и вновь получим МНК оценки коэффициентов.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



«Подогнанная» линия регресии. Обратите внимание, что коэффициенты отличны от «теоретических»

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



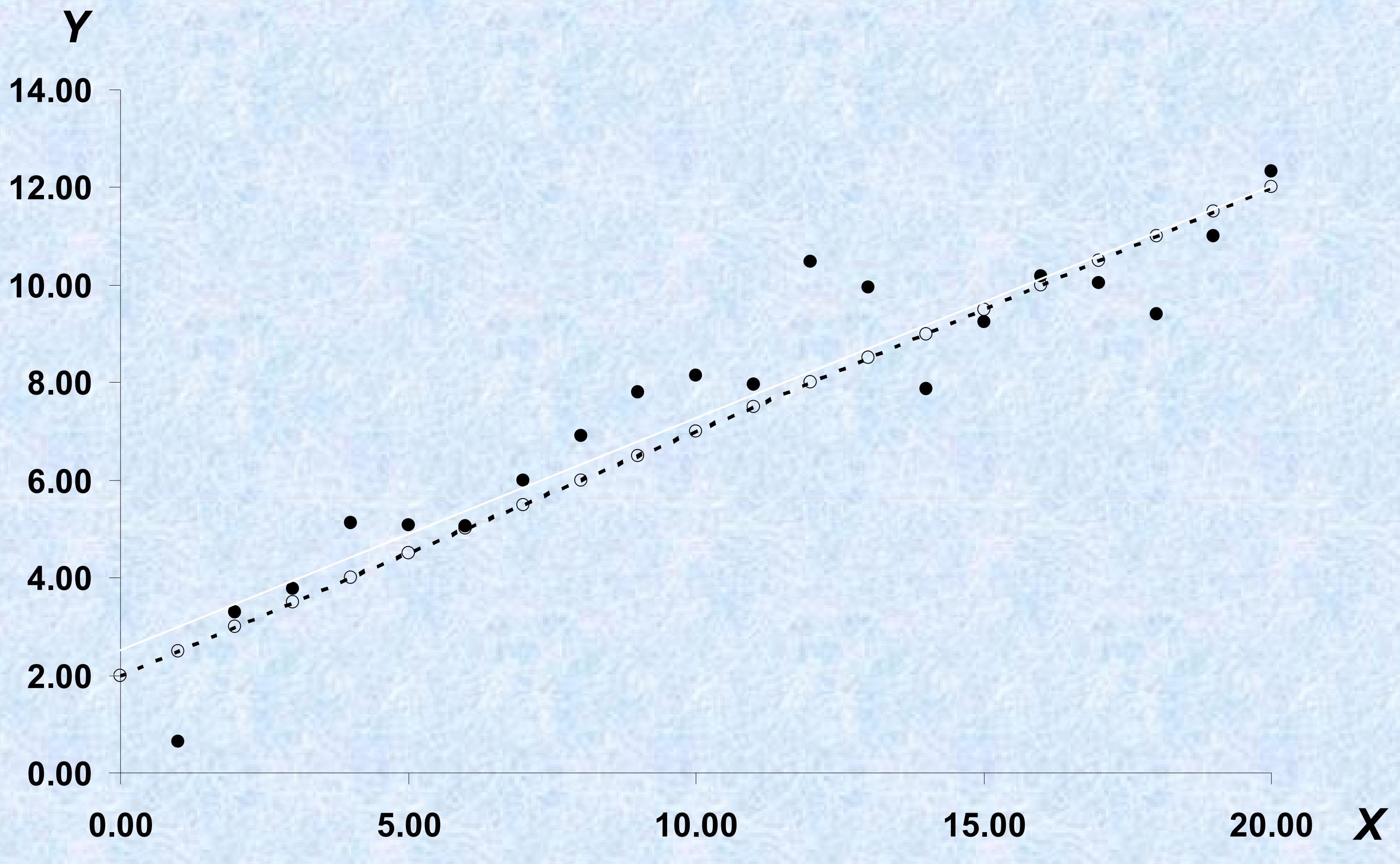
$$\hat{Y} = 1.63 + 0.54X$$

«Подогнанная» и теоретическая линии регрессии.

Обратите внимание, что вычисленные коэффициенты отличны от «теоретических».

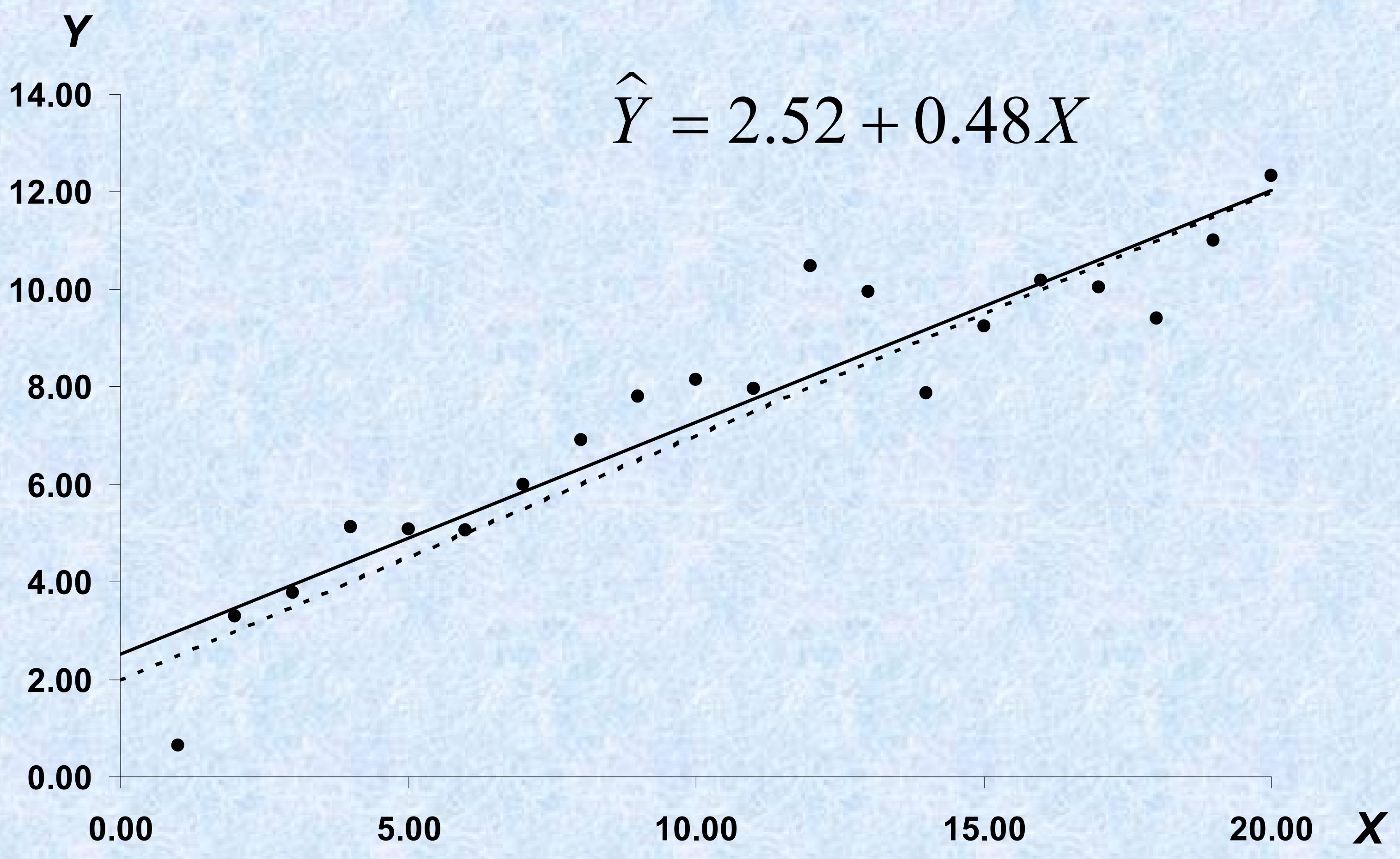
В частности коэффициент наклона переоценен (b2=0.54 вместо β_2=0.50), а свободный член недооценен (b1=1.63 вместо β_1= 2.00)

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



Новый эксперимент Монте-Карло, и новое распределение «*наблюдаемых*» значений, отличное от «*теоретического*».

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



$$\hat{Y} = 2.52 + 0.48X$$

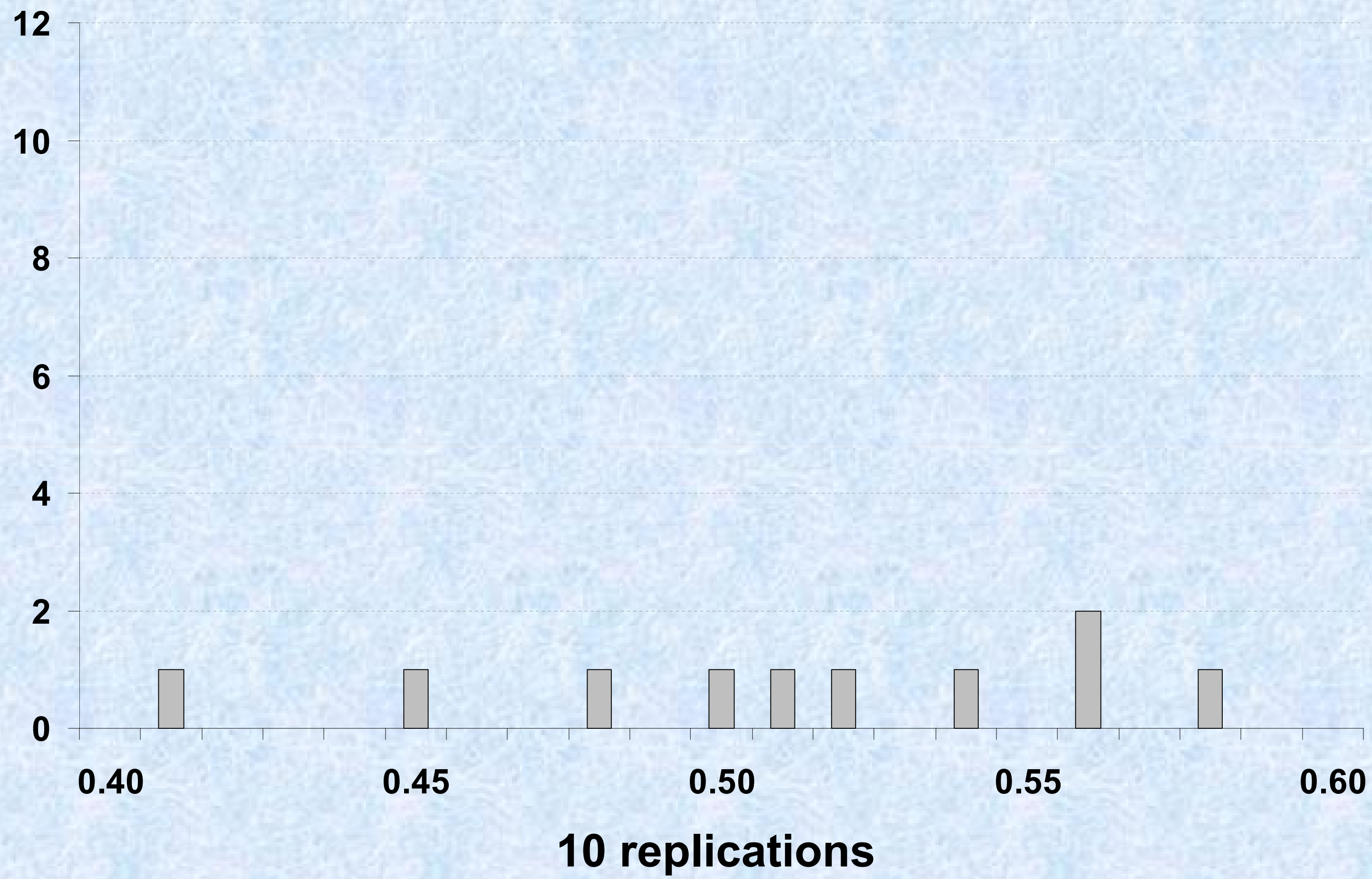
В данном случае коэффициент наклона оказался недооценен, а константа- переоценена.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

номер эксперимента	b_1	b_2
1	1.63	0.54
2	2.52	0.48
3	2.13	0.45
4	2.14	0.50
5	1.71	0.56
6	1.81	0.51
7	1.72	0.56
8	3.18	0.41
9	1.26	0.58
10	1.94	0.52

Таблица распределения оценок коэффициентов по 10 экспериментам Монте-Карло.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



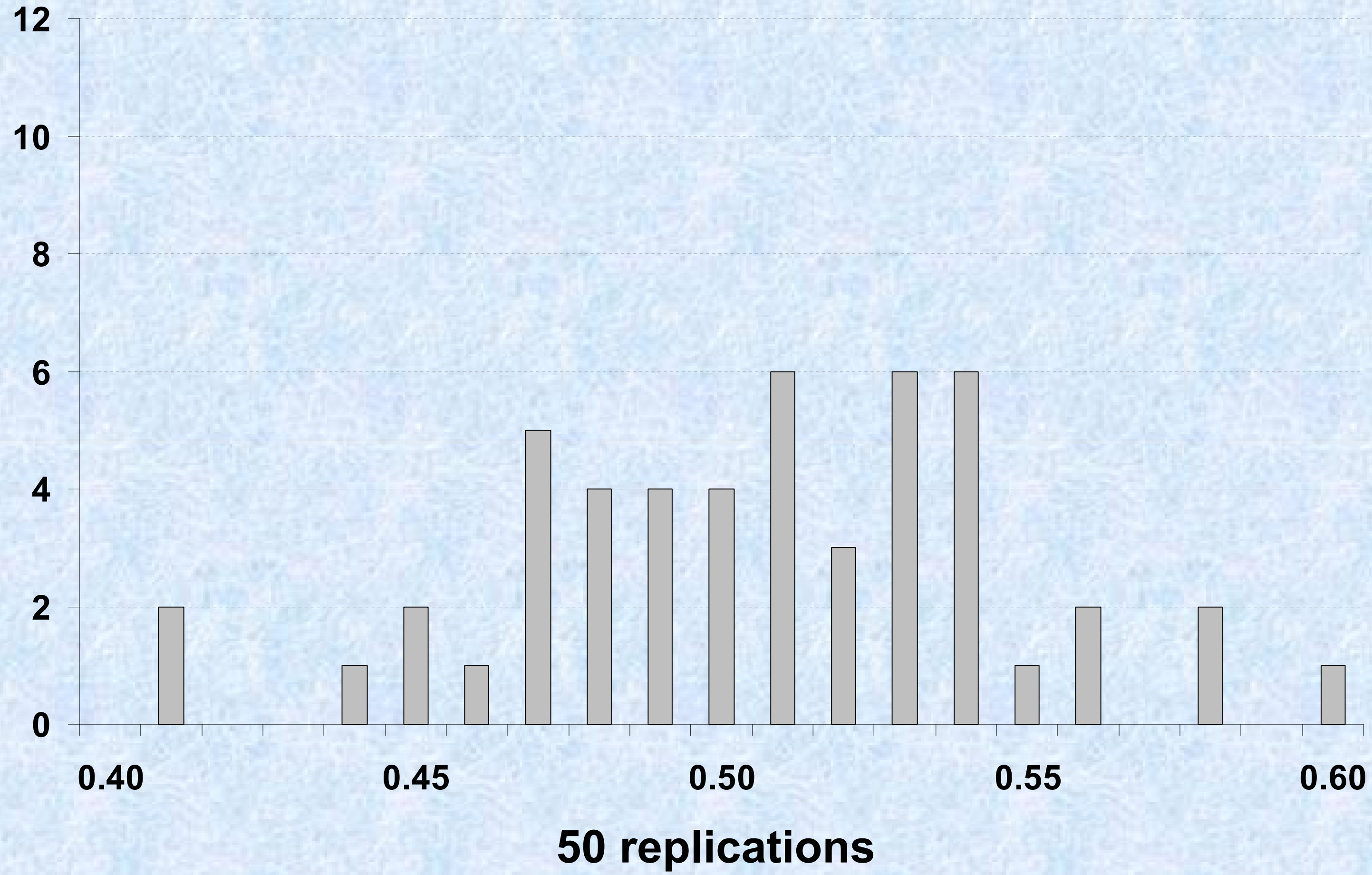
По гистограмме распределения β_2 пока еще ничего не понятно в силу небольшого количества экспериментов.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

1-10	11-20	21-30	31-40	41-50
0.54	0.49	0.54	0.52	0.49
0.48	0.54	0.46	0.47	0.50
0.45	0.49	0.45	0.54	0.48
0.50	0.54	0.50	0.53	0.44
0.56	0.54	0.41	0.51	0.53
0.51	0.52	0.53	0.51	0.48
0.56	0.49	0.53	0.47	0.47
0.41	0.53	0.47	0.55	0.50
0.58	0.60	0.51	0.51	0.53
0.52	0.48	0.47	0.58	0.51

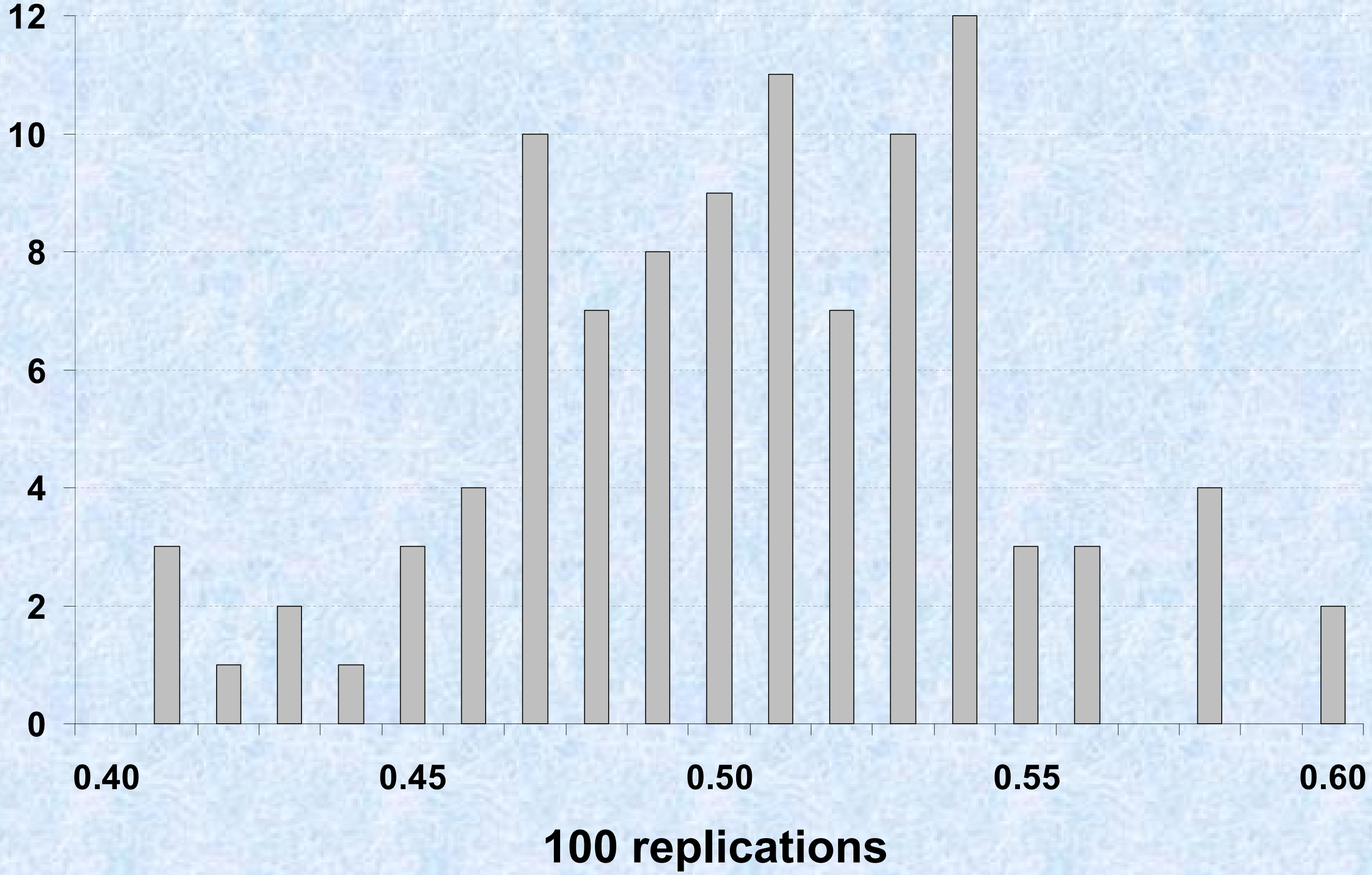
Это уже результаты 50 экспериментов.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



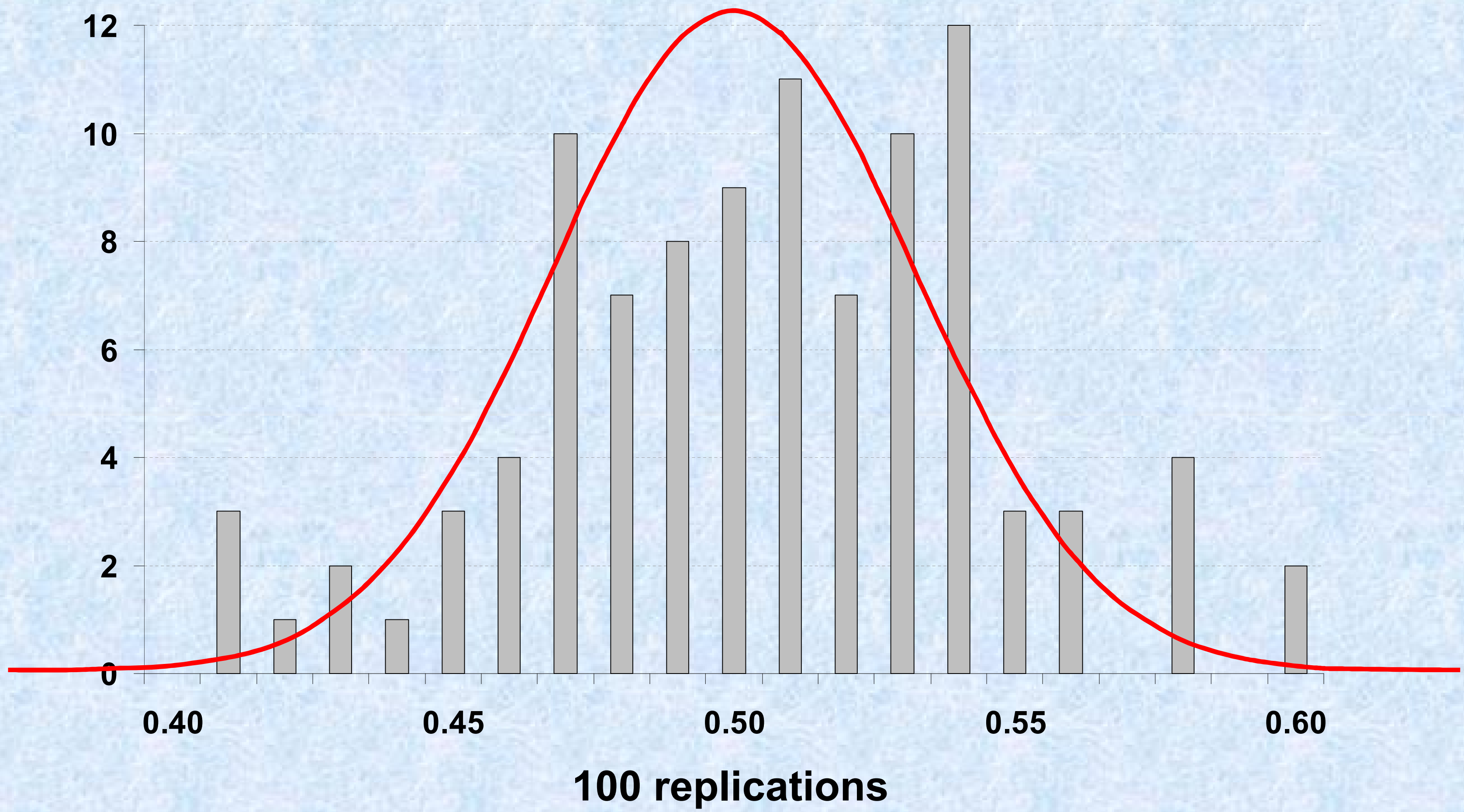
Здесь уже прослеживается центр распределения.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



По 100 экспериментам можно уже заметить симметричность распределения оценки вокруг точки 0,50.

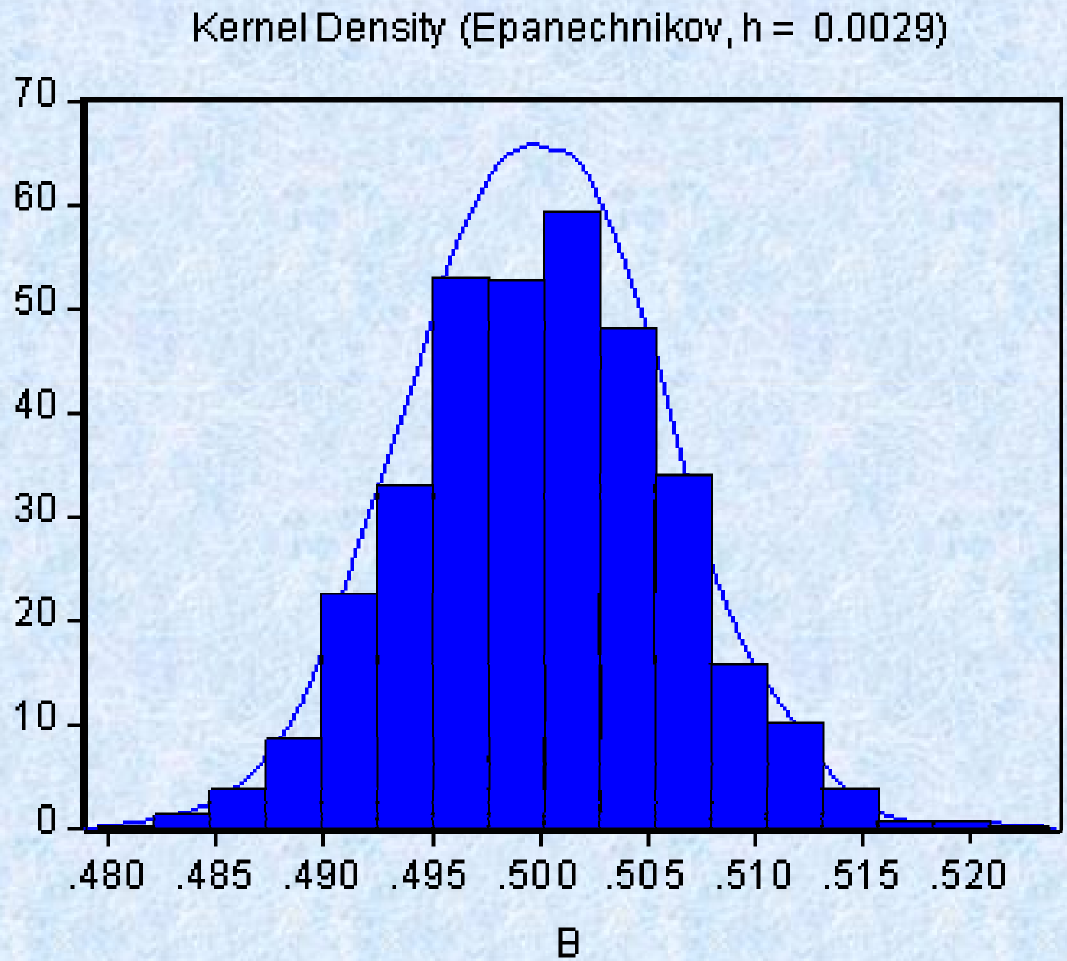
КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



Однако распределение все еще довольно грубое. При 1000 и более экспериментов оно уже будет плотно огибаться кривой плотности нормального закона.

Эта нормальность является следствием нормальности ошибки в наших экспериментах

КОЭФФИЦИЕНТЫ РЕГРЕССИИ КАК СЛУЧАЙНЫЕ ВЕЛИЧИНЫ



1000 экспериментов дают практически идентичное теоретическому распределение.

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Устойчивость по сложению многомерного нормального распределения:

$$\mathbf{Z}_1 \sim \mathrm{N}\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\right), \mathbf{Z}_2 \sim \mathrm{N}\left(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\right),$$

$$\mathbf{X} = \mathbf{Z}_1 + \mathbf{Z}_2 \sim \mathrm{N}\left(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + 2\mathbf{Var}\left(\mathbf{Z}_1, \mathbf{Z}_2\right)\right)$$

Упражнение:

Докажите эту формулу

Аналогии:

Одномерный случай:

$$Z_1 \sim \mathrm{N}\left(\mu_1, \sigma_1^2\right), Z_2 \sim \mathrm{N}\left(\mu_2, \sigma_2^2\right),$$

$$X = Z_1 + Z_2 \sim \mathrm{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2\right)$$

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

<u>Def</u> Маргинальная ф.р. многомерной с.в.-совместная ф.р. выделенных компонент с.в.

Рассмотрим *двухкомпонетную* с.в.:

$$\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right) \in \mathbb{R}^{n_1+n_2} \text{ с ф.р. } F_{\xi}\left(\mathbf{x}\right) = F_{\xi_1,\xi_2}\left(\mathbf{x}_1, \mathbf{x}_2\right)$$

$$F_{\xi_1}\left(\mathbf{x}_1\right) = \mathbf{P}\left(\boldsymbol{\xi}_1 < \mathbf{x}_1\right) = \mathbf{P}\left(\boldsymbol{\xi}_1 < \mathbf{x}_1; \boldsymbol{\xi}_2 \in \mathbb{R}^{n_2}\right) = F_{\xi}\left(\mathbf{x}_1, +\infty\right) =$$

$$= \int_{-\infty}^{\mathbf{x}_1} \int_{\mathbb{R}^{n2}} f_{\xi_1,\xi_2}(\mathbf{z}_1, \mathbf{z}_2)\,\mathbf{dz_1 dz}_2 = \int_{-\infty}^{\mathbf{x}_1} f_{\xi_1}(\mathbf{z}_1)\,\mathbf{dz_1} - \text{маржинальная ф.р.}$$

$$f_{\xi_1}(\mathbf{x}_1) = \int_{\mathbb{R}^{n2}} f_{\xi_1,\xi_2}(\mathbf{x}_1, \mathbf{x}_2)\,\mathbf{dx}_2 \text{ - плотность предельной ф.р.}$$

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Плотность маргинального нормального распределения:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z_1} \\ \mathbf{Z}_2 \end{pmatrix}, \quad \mathbf{Z} \sim \mathrm{N}\left(\boldsymbol{\mu}; \boldsymbol{\Sigma}\right), \quad \boldsymbol{\mu} = \mathrm{E}\,\mathbf{Z} = \begin{pmatrix} \mathrm{E}\,\mathbf{Z}_1 \\ \mathrm{E}\,\mathbf{Z}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} = \mathrm{cov}\left(\mathbf{Z}_1, \mathbf{Z_2}\right), \boldsymbol{\Sigma}_{11} = \mathrm{Var}\left(\mathbf{Z}_1\right), \boldsymbol{\Sigma}_{22} = \mathrm{Var}\left(\mathbf{Z}_2\right)$$

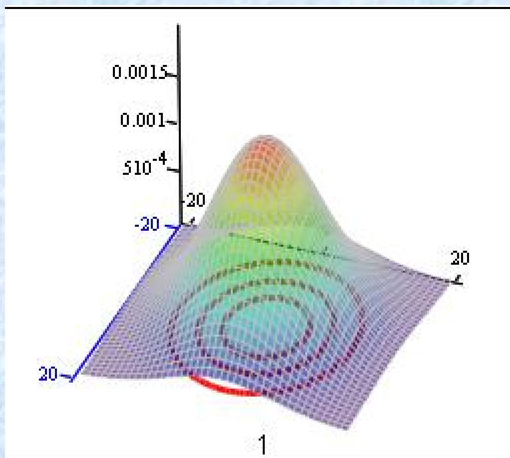
$$\mathbf{Z}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \quad \mathrm{f}_{\mathbf{Z}_1}\left(\mathbf{z}_1\right) = \frac{1}{\left(2\pi\right)^{n_1/2}\left(\det \boldsymbol{\Sigma}_{11}\right)^{1/2}} \cdot \exp\left[-\frac{\left(\mathbf{z}_1 - \boldsymbol{\mu}_1\right)' \boldsymbol{\Sigma}_{11}^{-1}\left(\mathbf{z}_1 - \boldsymbol{\mu}_1\right)}{2}\right]$$

В двумерном случае $\quad \xi_x \sim N\left(\mu_x, \sigma_x^2\right)$

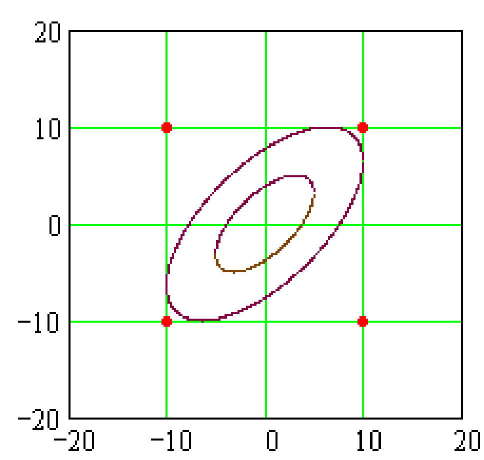
МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Нормальный закон (двумерный случай)

$$\mathrm{f}_{\xi}\left(z_1, z_2\right) = \frac{1}{2\pi\sqrt{\left(1-\rho^2\right)\sigma_1^2\sigma_2^2}} \cdot \exp\left[-\frac{\begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}' \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}}{2}\right]$$



f,f



2

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Условное распределение компонент нормального вектора:

$$\mathbf{Z}_1 \mid \left(\mathbf{Z}_2 = \mathbf{z}_2\right) \sim N(\boldsymbol{\mu}_1 + \mathbf{B}' \left(\mathbf{Z}_1 - \boldsymbol{\mu}_1\right), \boldsymbol{\Sigma}_{11} - \mathbf{B}'\boldsymbol{\Sigma}_{22}\mathbf{B}), \mathbf{B} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}$$

Rem1 Если Σ**=diag**, то компоненты с.в. **Z** независимы, и имеют распределения как маргинальное при независимости компонент.

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

УСЛОВНАЯ Ф.Р. ДВУМЕРНОЙ НОРМАЛЬНОЙ С.В

$$\boldsymbol{\xi} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\xi} = (\xi_x, \xi_y) \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}\right)$$

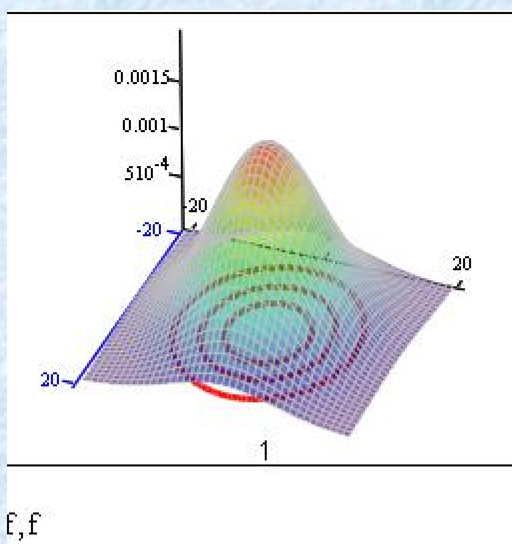
$$\xi_x \mid \xi = y_y \sim N\left(\mu_x + \rho\sigma_x \frac{y - \mu_y}{\sigma_y}, \sigma_x^2 \left(1 - \rho^2\right)\right)$$

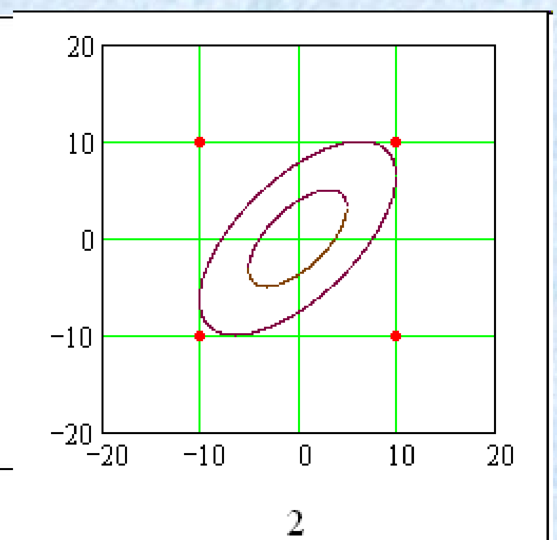
Заметьте, что дисперсия условного распределения одной компоненты не зависит от второй компоненты- свойство гомоскедастичности нормального закона!!!

МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Нормальный закон (двумерный случай)

$$\mathrm{f}_\xi\left(z_1, z_2\right) = \frac{1}{2\pi\sqrt{\left(1-\rho^2\right)\sigma_1^2\sigma_2^2}} \cdot \exp\left[-\frac{\begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}' \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}}{2}\right]$$





МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Условное распределение компонент нормального вектора:

$$\mathbf{Z}_1 \,|\, \left(\mathbf{Z}_2 = \mathbf{z_2}\right) \sim N(\boldsymbol{\mu}_1 + \mathbf{B}'\left(\mathbf{Z}_1 - \boldsymbol{\mu}_1\right), \boldsymbol{\Sigma}_{11} - \mathbf{B}'\boldsymbol{\Sigma}_{22}\mathbf{B}),\ \mathbf{B} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}$$

Rem2 Матрица ковариаций условного распределения *оказывается* *независимой* от компонент **Z2**, поэтому (в частности) $\mathrm{D}\left[\mathbf{Z}_1 \,|\, \left(\mathbf{Z}_2 = \mathbf{z_2}\right)\right]$ не зависит от **Z2**

Это важно, поскольку при построении регрессии **Y** на **X** наилучший предиктор $\mathrm{E}\left(\mathbf{Y} \,|\, \mathbf{X}\right)$, а ошибка предсказания (**MSPE** – mean square predictor error): $\mathrm{E}\left[\mathbf{Y} - \mathrm{E}\left(\mathbf{Y} \,|\, \mathbf{X}\right)\right]^2 = \mathrm{D}\left(\mathbf{Y} \,|\, \mathbf{X}\right)$, т.е. мы получаем, что в нашем случае она не зависит от **X-** это хорошее свойство оценки – ***гомоскедастичность***

гомоскедастичность (хорошая регрессия) гетероскедастичность (плохая регрессия)

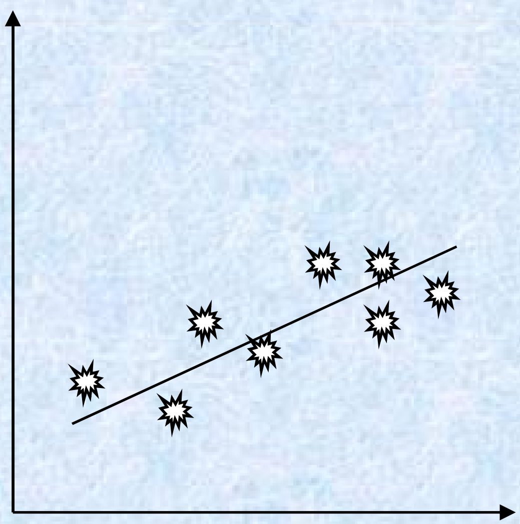
МНОГОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Условное распределение компонент нормального вектора:

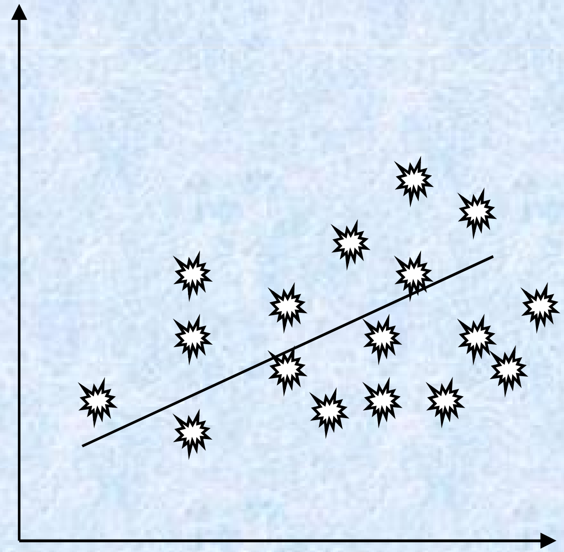
гомоскедастичность

(хорошая регрессия)



гетероскедастичность

(плохая регрессия)



Конец лекции