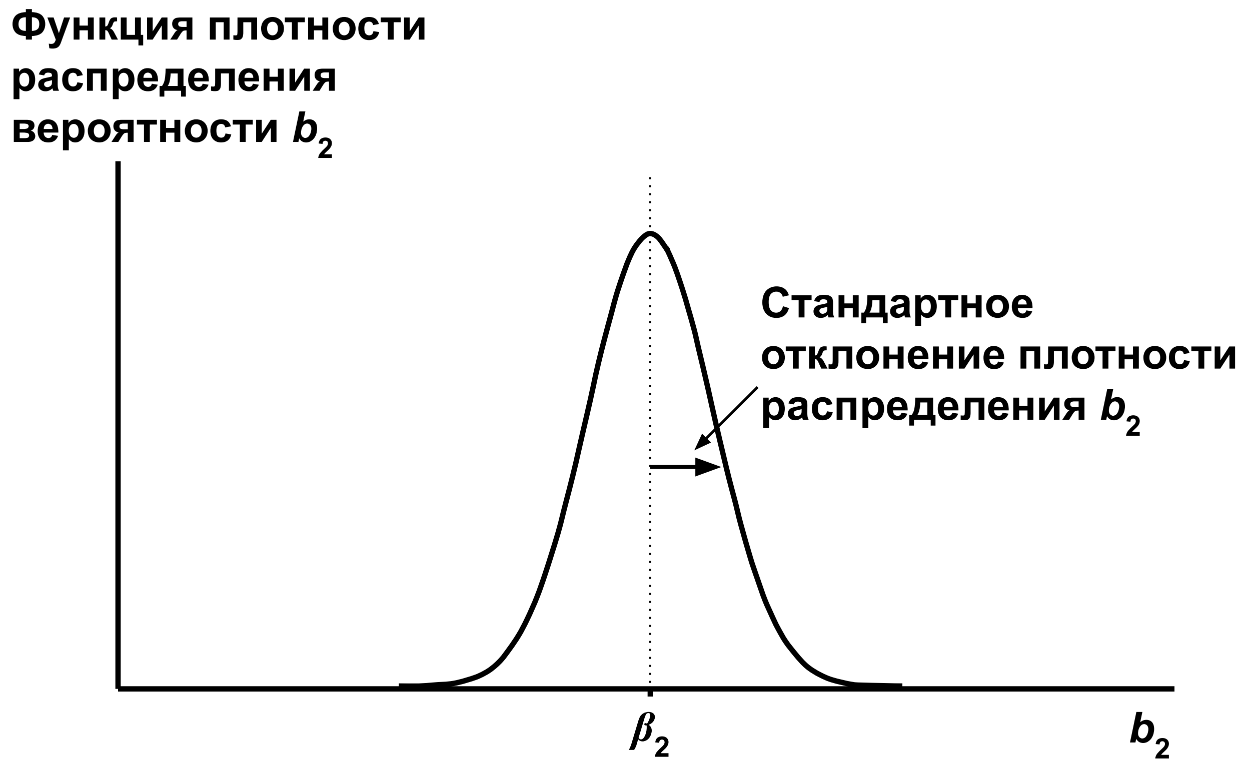
ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

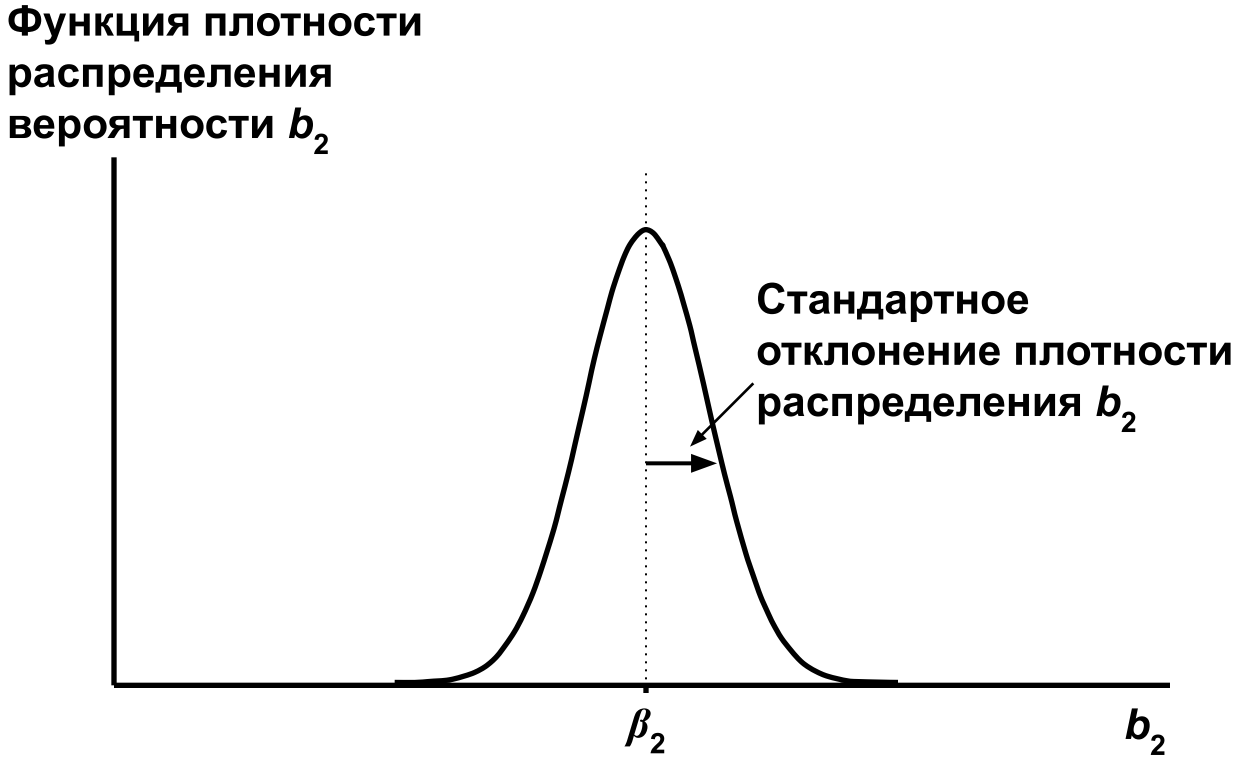
Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$



Мы видим, что коэффициенты регрессии b_1 и b_2 являются случайными величинами. Они представляют точечные оценки β_1 и β_2, соответственно. Последним следствием мы показываем, что точечные оценки являются несмещенными.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$



В этом же следствии мы увидим, что можем также получить оценки стандартного отклонения распределения. Это даст некоторое представление об их вероятной надежности и послужит основой для проверки гипотез.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2}$$

Выражения (которые не решены) для дисперсий их распределений показаны выше. См. Вставку 2.3 в тексте для доказательства выражения дисперсии b_2.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\}$$

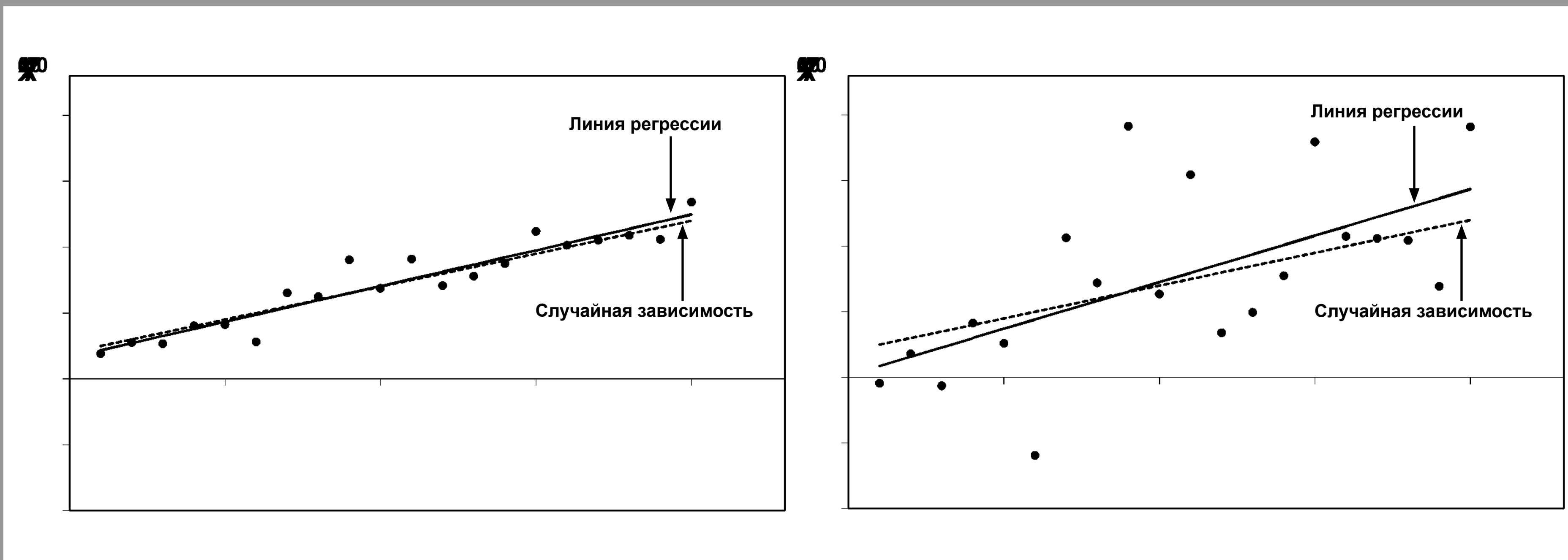
$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2}$$

Мы сосредоточимся на значении выражения для дисперсии b_2. Рассматривая числитель, мы видим, что дисперсия b_2 пропорциональна σ_u^2. Этого и следовало ожидать. Чем больше разброс в модели, тем менее точными будут наши оценки.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$

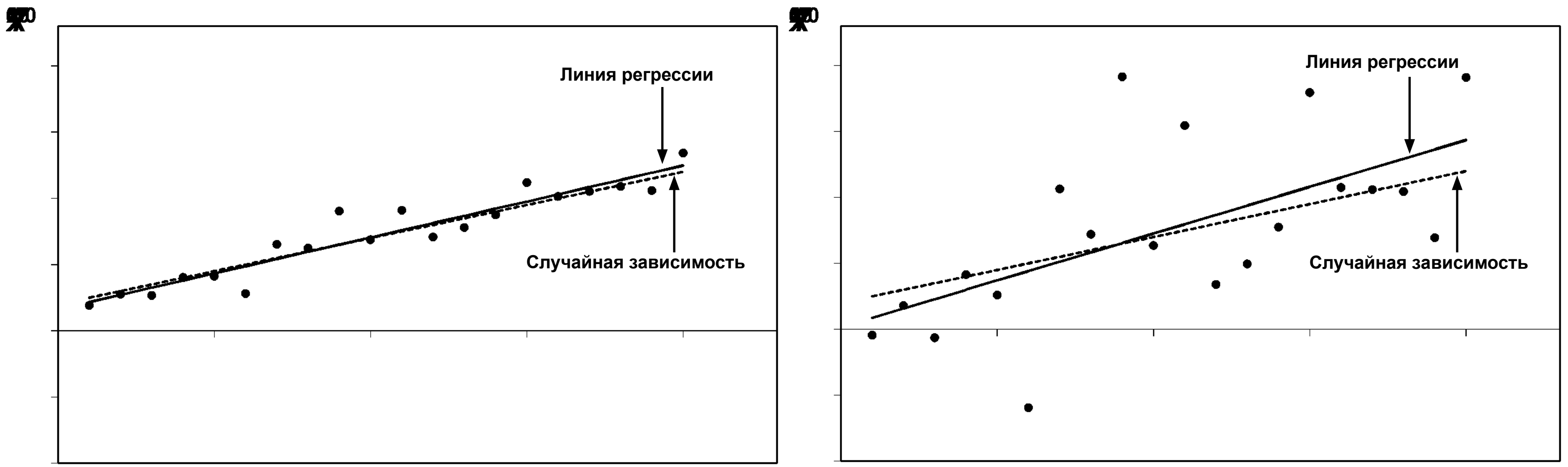


Это показано на диаграммах, представленных выше. Случайная составляющая зависимости, Y = 3.0 + 0.8Xпредставленная пунктирной линейна обеих диаграммах одинакова.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$

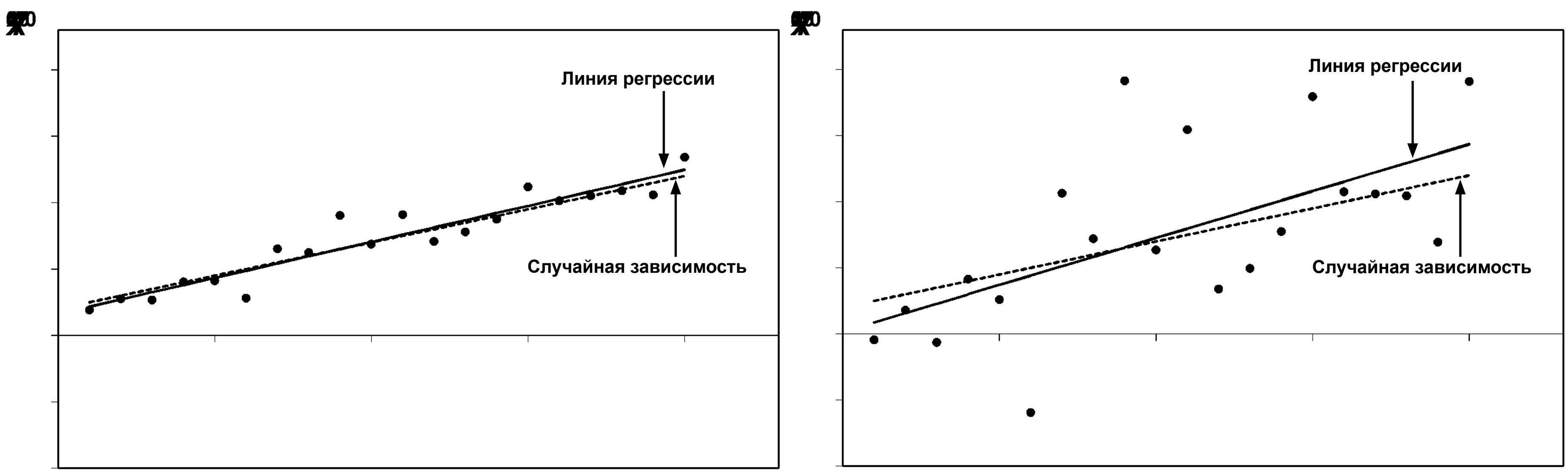


Значения X одинаковы, и одинаковые случайные числа использовались для генерирования значений остаточного члена в 20 наблюдениях.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$



Однако, на правой диаграмме случайные числа умножались в 5 раз. Как следствие, линия регрессии, сплошная линия, намного меньше приближена к линии случайной зависимости.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2}$$

Посмотрим на знаменатель выражения для дисперсии b_2Чем больше сумма квадратов отклонений *X*, тем меньше дисперсия b_2.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2} = \frac{\sigma_u^2}{n \, \mathrm{MSD}(X)}$$

$$\mathrm{MSD}(X) = \frac{1}{n} \sum \left(X_i - \overline{X} \right)^2$$

Однако, значение суммы квадратов отклонений зависит от двух факторов: количества наблюдений и размера отклонений X_i от его выборочного среднего. Для того, чтобы различать их, будет удобно определить среднее квадратическое отклонение *X*, MSD(*X*).

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n\,\mathrm{MSD}(X)}$$

$$\mathrm{MSD}(X) = \frac{1}{n} \sum (X_i - \overline{X})^2$$

Из приведенного выражения видно, что дисперсия b2 обратно пропорциональна *n*, числу наблюдений в выборке, которые управляют MSD(X). Чем больше информации мы имеем, тем точнее будут оценки.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2} = \frac{\sigma_u^2}{n \,\mathrm{MSD}(X)}$$

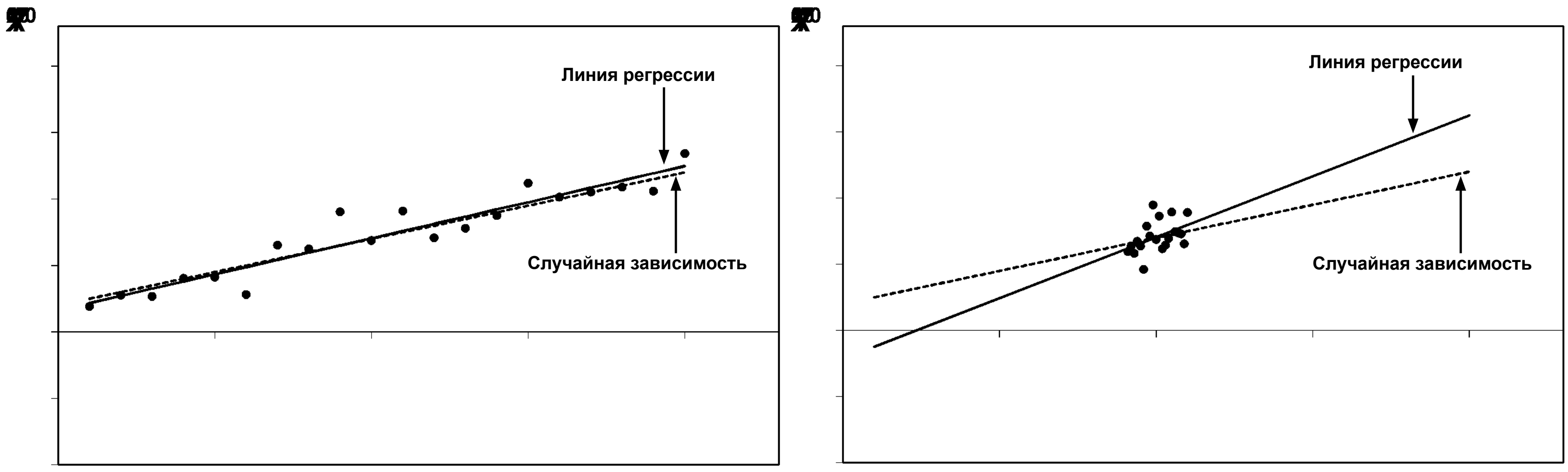
$$\mathrm{MSD}(X) = \frac{1}{n} \sum \left(X_i - \overline{X} \right)^2$$

Третьим следствием выражения является то, что дисперсия обратно пропорциональна среднему квадратическому отклонению X. В чем причина этого?

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$

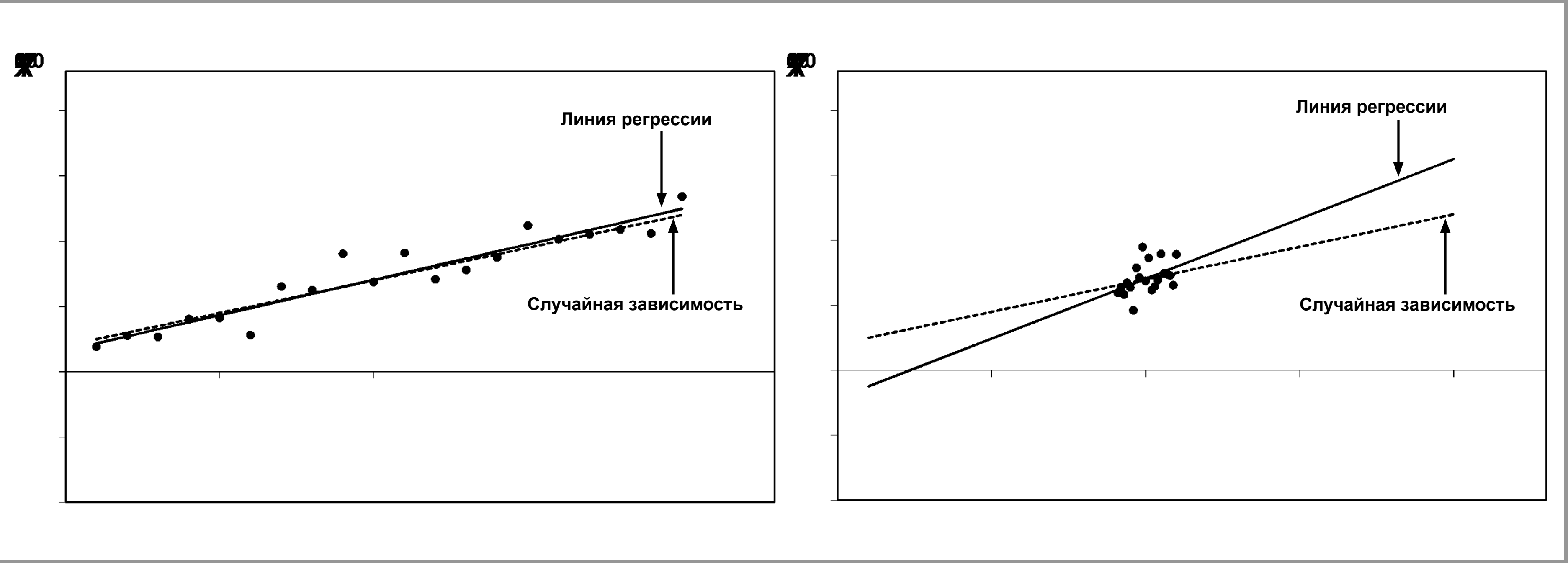


На вышеприведенных диаграммах линия случайной зависимости одинакова и для 20 значений наблюдений в распределении использовались одинаковые случайные числа.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$

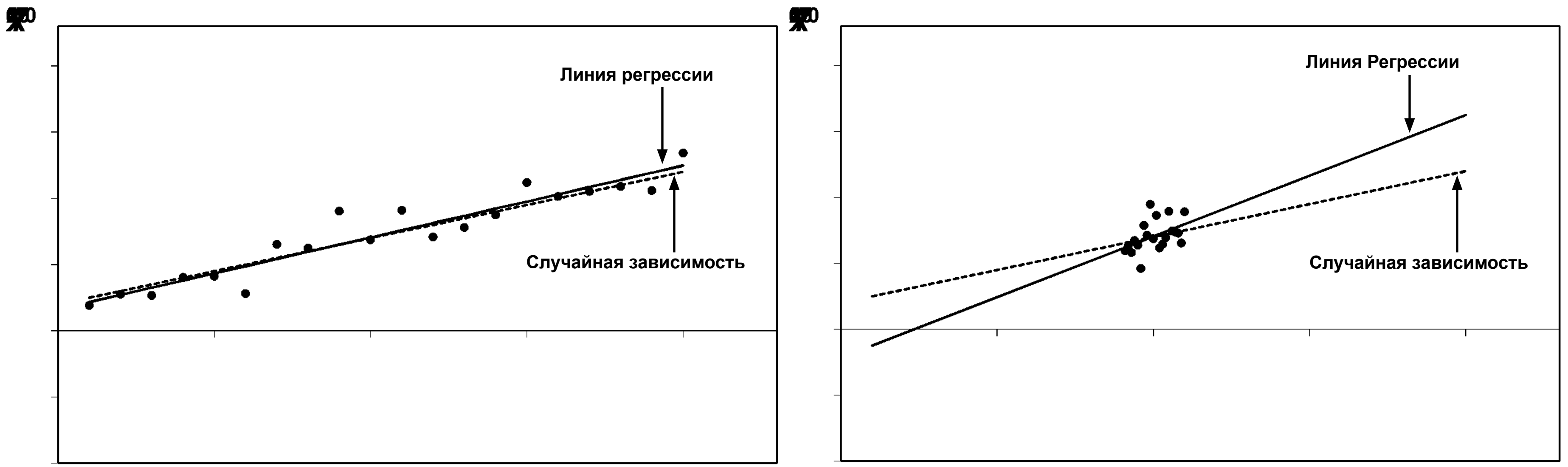


Однако, MSD (X) намного меньше на правой диаграмме, так как значения X намного ближе друг к другу.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

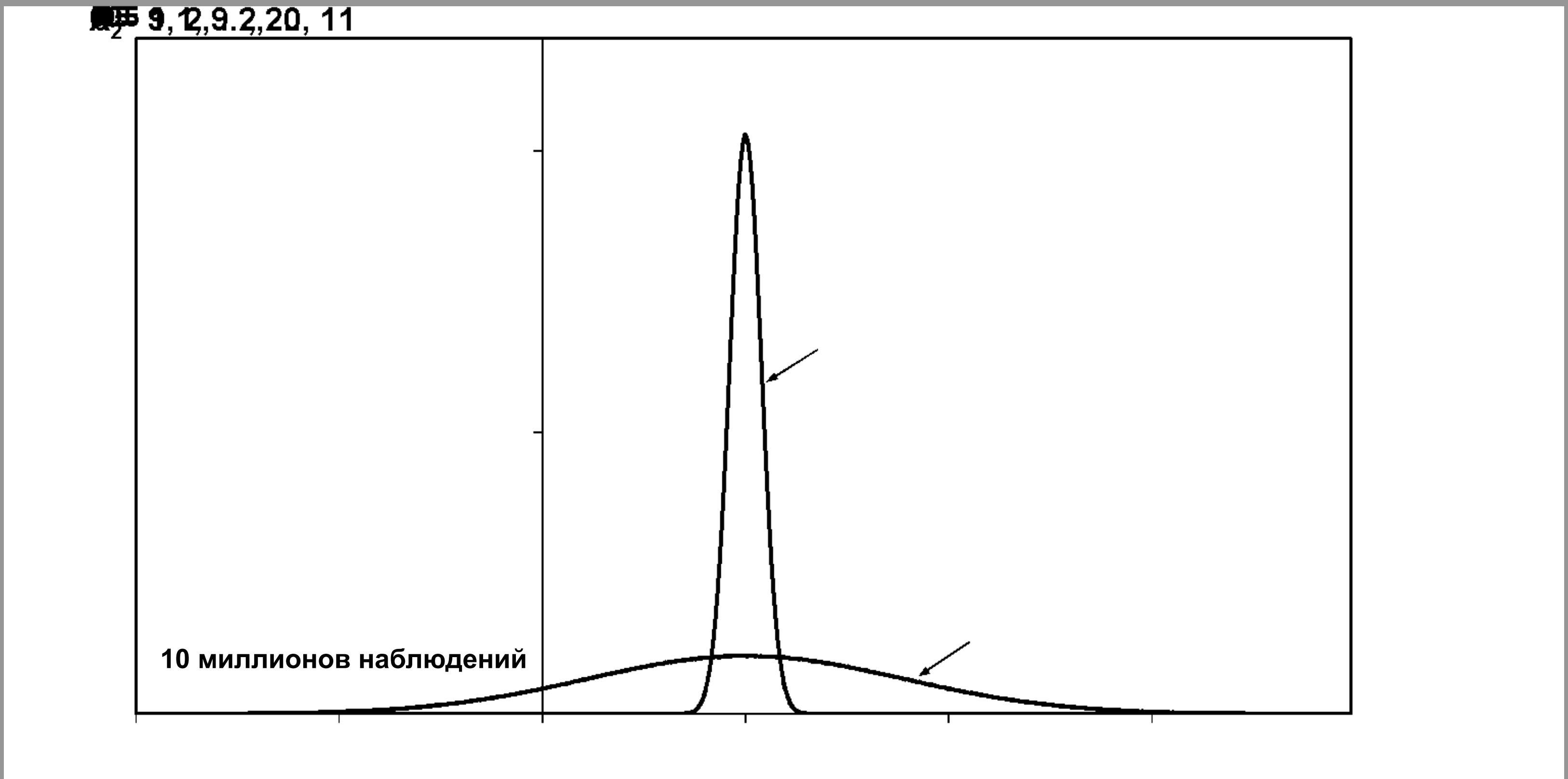
$$Y = 2.0 + 0.5X$$



Следовательно, на этой диаграмме положение линии регрессии более чувствительно к значениям наблюдений распределения, и, как следствие, линия регрессии, вероятно, будет относительно неточной.

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$

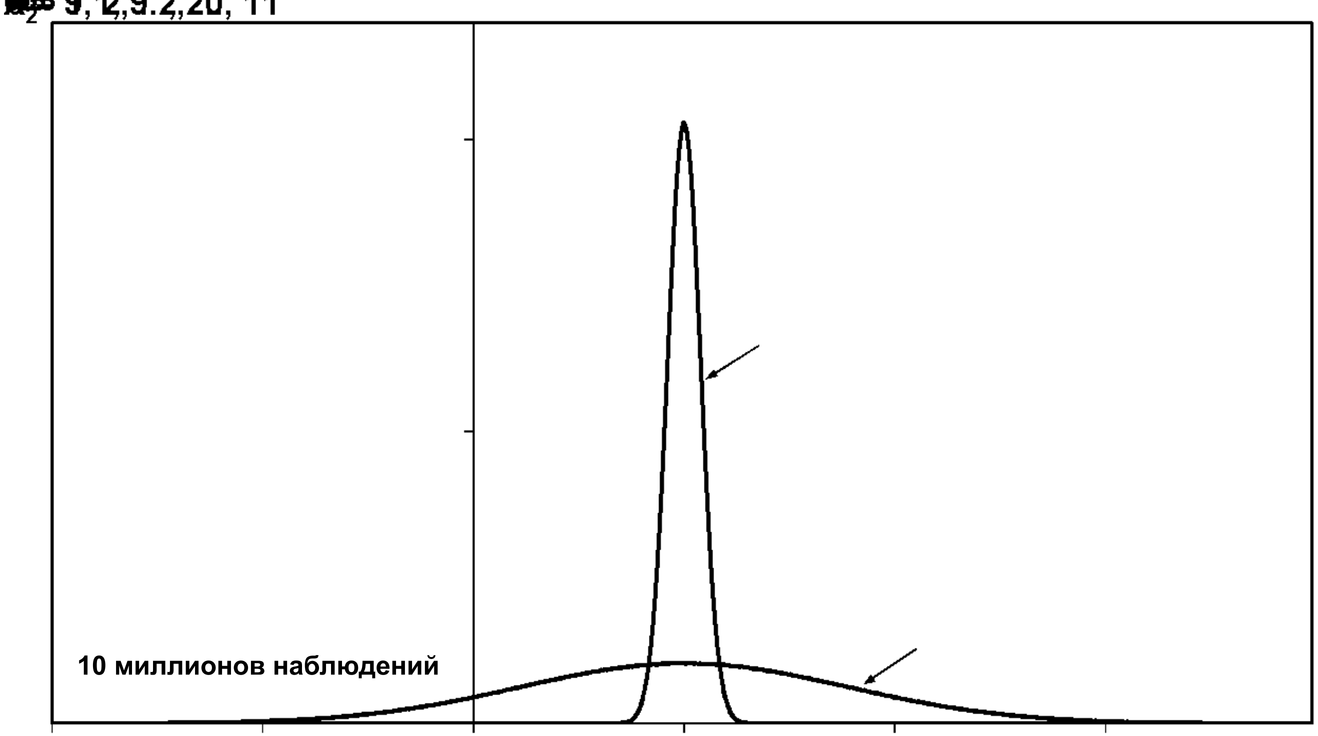


На рисунке показаны распределения оценок b_2 для X = 1, 2, ..., 20 и X = 9.1, 9.2, ..., 11 при моделировании с 10 миллионами наблюдений.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$Y = 2.0 + 0.5X$$



Это подтверждает, что распределение оценок, полученных с высокой дисперсией X, имеет гораздо меньшее отклонение, чем распределение с низкой дисперсией X.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n \,\mathrm{MSD}(X)}$$

Конечно, как видно из выражений дисперсии, отношение MSD (X) к дисперсии u важнее, чем ее абсолютное значение.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2} = \frac{\sigma_u^2}{n \, \mathrm{MSD}(X)}$$

Мы не можем рассчитать теоретические дисперсии именно потому, что не знаем дисперсии остаточного члена. Однако, мы можем получить оценку σ_u^2 из остатков.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n\,\mathrm{MSD}(X)}$$

Очевидно, что разброс остатков относительно линии регрессии будет отражать неизвестный разброс u относительно линии $Y_i = \beta_1 + b_2 X_i$ хотя в общем остаток и случайный член ни в одном из наблюдений не равны друг другу.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n \, \mathrm{MSD}(X)}$$

$$\mathrm{MSD}(e) = \frac{1}{n} \sum (e_i - \overline{e})^2 = \frac{1}{n} \sum e_i^2$$

Одной из мер разброса остатков является их средняя квадратическая ошибка, MSD(*e*), которая определяется формулой, указанной на слайде. (Помните, что среднее значение остатков OLS равно нулю). Интуитивно это должно приводить к дисперсии *u*.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n \,\mathrm{MSD}(X)}$$

$$\mathrm{MSD}(e) = \frac{1}{n} \sum (e_i - \overline{e})^2 = \frac{1}{n} \sum e_i^2$$

Прежде чем пойти дальше, задайте себе следующий вопрос: какая прямая вероятнее будет ближе к точкам, представляющим собой выборку наблюдений по X и Y, истинная прямая $Y = \beta_1 + \beta_2 X$ или линия регрессии $\hat{Y} = b_1 + b_2 X$?

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\}$$

$$\sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n\,\mathrm{MSD}(X)}$$

$$\mathrm{MSD}(e) = \frac{1}{n} \sum (e_i - \overline{e})^2 = \frac{1}{n} \sum e_i^2$$

Ответ – линия регрессии, так как по определению она строится таким образом, чтобы свести к минимуму сумму квадратов расстояний между ней и значениями наблюдениями.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2} = \frac{\sigma_u^2}{n \, \mathrm{MSD}(X)}$$

$$\mathrm{MSD}(e) = \frac{1}{n} \sum \left(e_i - \overline{e} \right)^2 = \frac{1}{n} \sum e_i^2$$

Следовательно, разброс остатков у нее меньше, чем разброс значений *u*, а MSD(e) имеет тенденцию занижать оценку σ_u^2.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X} \right)^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum \left(X_i - \overline{X} \right)^2} = \frac{\sigma_u^2}{n \,\mathrm{MSD}(X)}$$

$$\mathrm{MSD}(e) = \frac{1}{n} \sum \left(e_i - \overline{e} \right)^2 = \frac{1}{n} \sum e_i^2$$

$$E\left(\mathrm{MSD}(e) \right) = \frac{n-2}{n} \sigma_u^2$$

Действительно, можно показать, что математическое ожидание MSD(e), если имеется всего одна независимая переменная, находится выражением приведенным выше.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n \, \mathrm{MSD}(X)}$$

$$s_u^2 = \frac{n}{n-2} \mathrm{MSD}(e) = \frac{n}{n-2} \frac{1}{n} \sum e_i^2 = \frac{1}{n-2} \sum e_i^2$$

Однако отсюда следует, что мы можем получить несмещенную оценку σ_u^2, умножив MSD(e) на n / (n – 2). Обозначим это s_u^2.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n \,\mathrm{MSD}(X)}$$

$$s_u^2 = \frac{n}{n-2} \mathrm{MSD}(e) = \frac{n}{n-2} \frac{1}{n} \sum e_i^2 = \frac{1}{n-2} \sum e_i^2$$

$$\text{с.о.}(b_1) = s_u \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}} \qquad \text{с.о.}(b_2) = \sqrt{\frac{s_u^2}{\sum (X_i - \overline{X})^2}}$$

Затем мы можем получить оценки стандартных отклонений распределений b_1 и b_2, подставив s_u^2 для σ_u^2 в выражениях дисперсии и взяв квадратные корни.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$

$$\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right\} \qquad \sigma_{b_2}^2 = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2} = \frac{\sigma_u^2}{n\,\mathrm{MSD}(X)}$$

$$s_u^2 = \frac{n}{n-2}\mathrm{MSD}(e) = \frac{n}{n-2}\frac{1}{n}\sum e_i^2 = \frac{1}{n-2}\sum e_i^2$$

$$\mathrm{s.e.}(b_1) = s_u \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}} \qquad \mathrm{s.e.}(b_2) = \sqrt{\frac{s_u^2}{\sum (X_i - \overline{X})^2}}$$

Они описываются как стандартные ошибки b_1 и b_2, «оценки среднеквадратических отклонений» являются более полными.

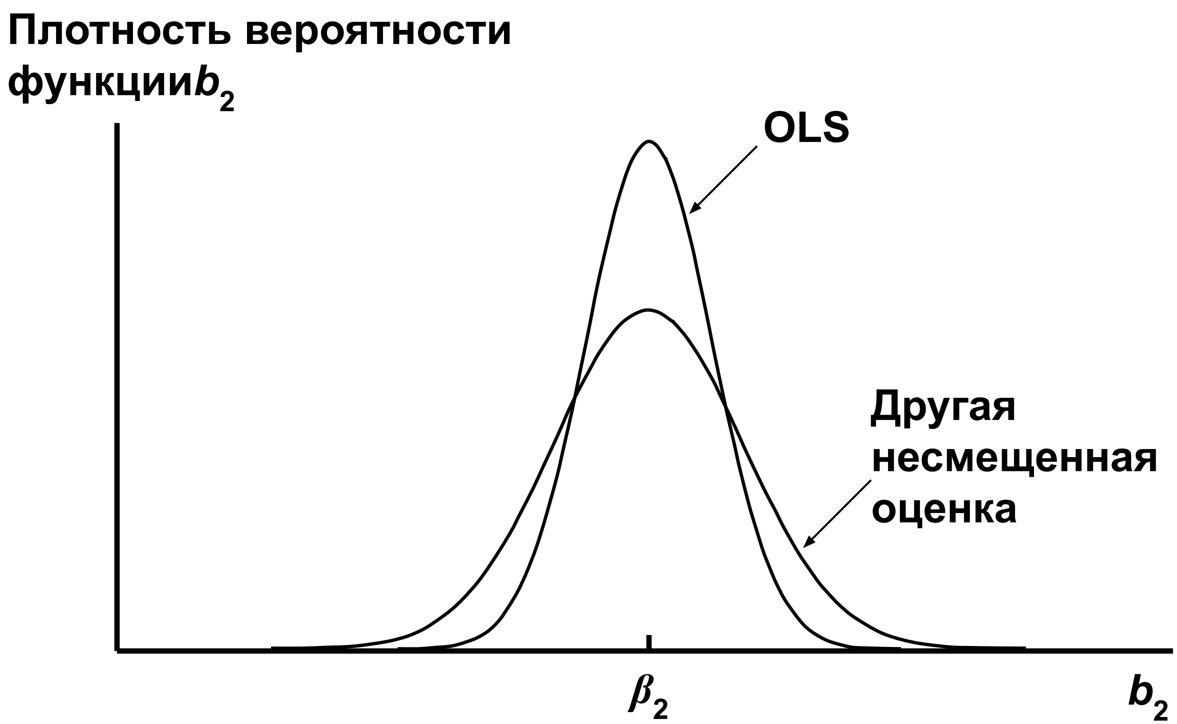
ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

```
. reg EARNINGS S

      Source |       SS       df       MS              Number of obs =     540
-------------+------------------------------           F(  1,   538) =  112.15
       Model |  19321.5589     1  19321.5589           Prob > F      =  0.0000
    Residual |  92688.6722   538  172.283777           R-squared     =  0.1725
-------------+------------------------------           Adj R-squared =  0.1710
       Total |  112010.231   539  207.811189           Root MSE      =  13.126

------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   2.455321   .2318512    10.59   0.000     1.999876    2.910765
       _cons |  -13.93347   3.219851    -4.33   0.000    -20.25849   -7.608444
------------------------------------------------------------------------------
```

Стандартные ошибки коэффициентов всегда появляются как часть результата регрессии. Здесь представлена регрессия почасовых заработков в годы обучения, которые обсуждались на предыдущих слайдах. Стандартные ошибки появляются в столбце справа от коэффициентов.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

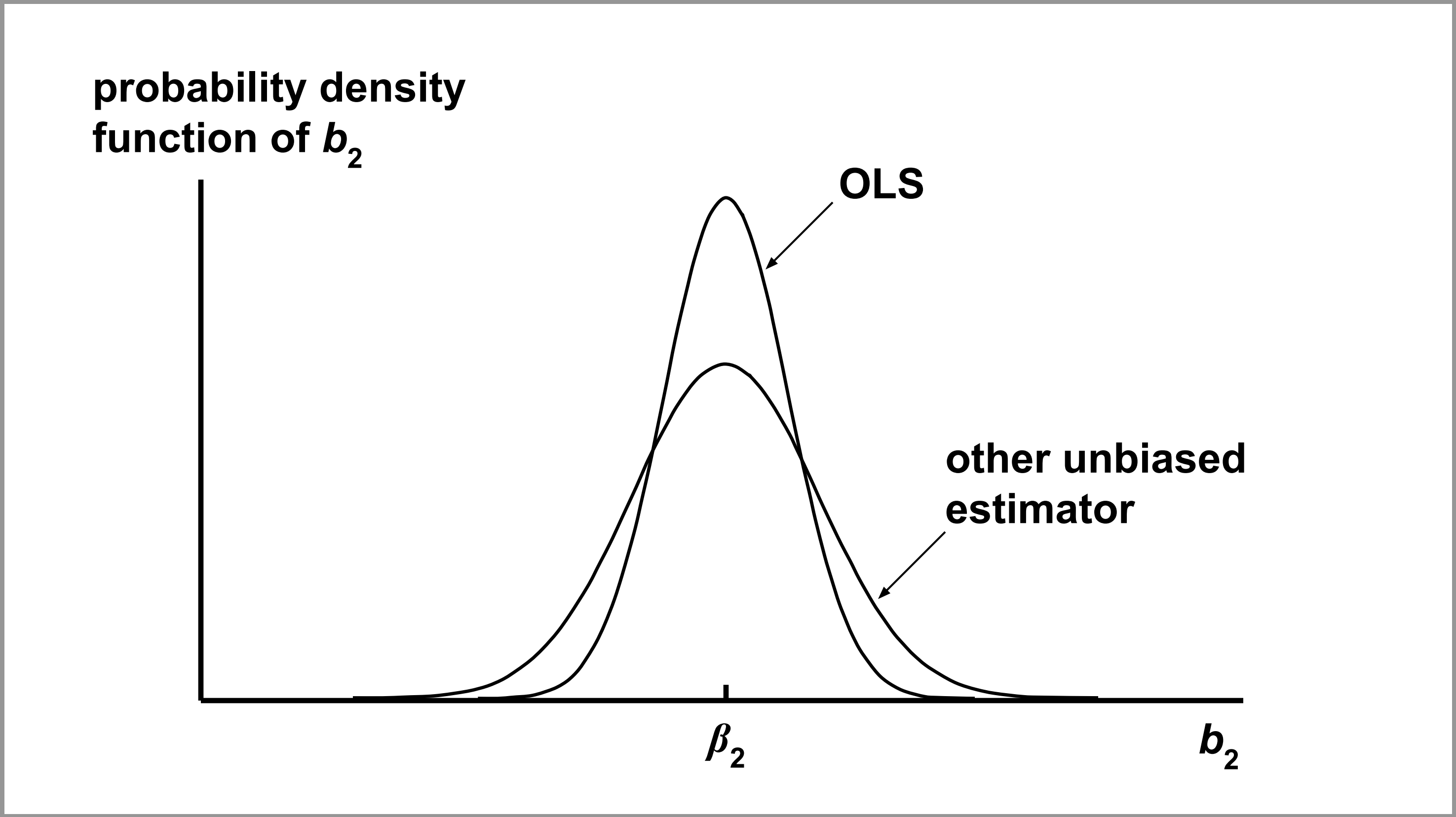
Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$



Теорема Гаусса-Маркова утверждает : при условии, что допущения модели регрессии действительны, оценки OLS являются BLUE: лучшая (наиболее эффективная) линейная (функция значений *Y*) несмещенных оценок параметров.

ТОЧНОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Простая регрессионная модель: $Y = \beta_1 + \beta_2 X + u$



Доказательство теоремы не сложное, но не является высокоприоритетным, и мы будем считать его надежным. См. Раздел 2.7 текста для доказательства простой модели регрессии.