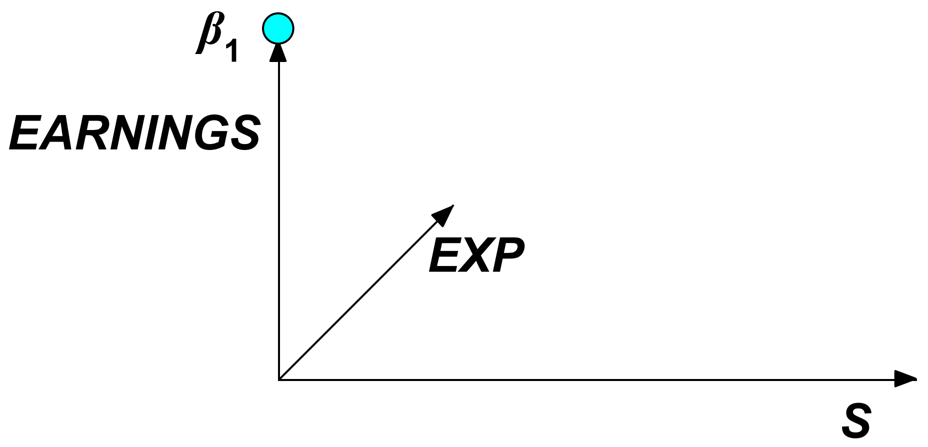
МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

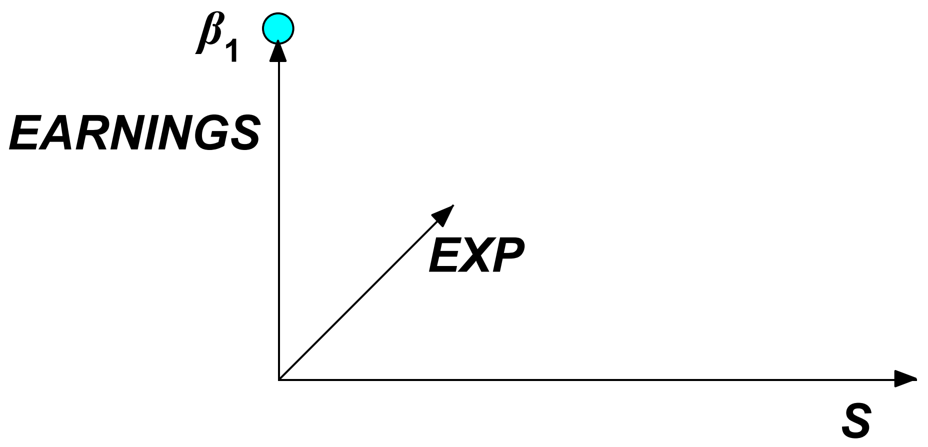
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Эта последовательность обеспечивает геометрическую интерпретацию модели множественной регрессии с двумя объясняющими переменными.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

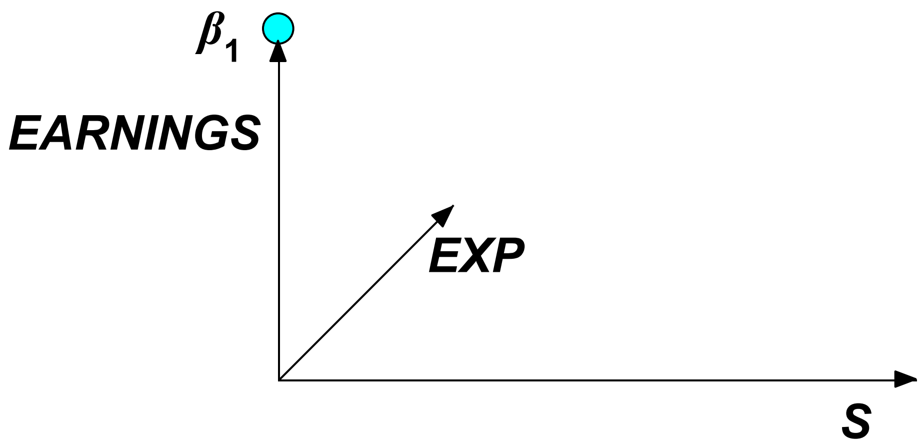
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



В частности, мы рассмотрим модель функции заработка, где почасовой заработок, *EARNINGS*, зависит от количества лет обучения (наивысший оконченный класс), S, и опыта работы в годах, *EXP*.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

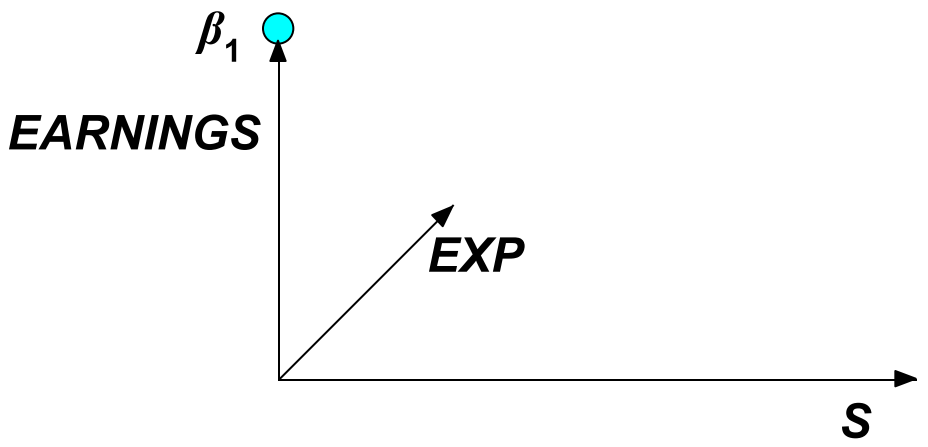
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Модель имеет три измерения, по одному для *EARNINGS*, S и *EXP*. Отправной точкой для исследования определения заработка является константа, β_1.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

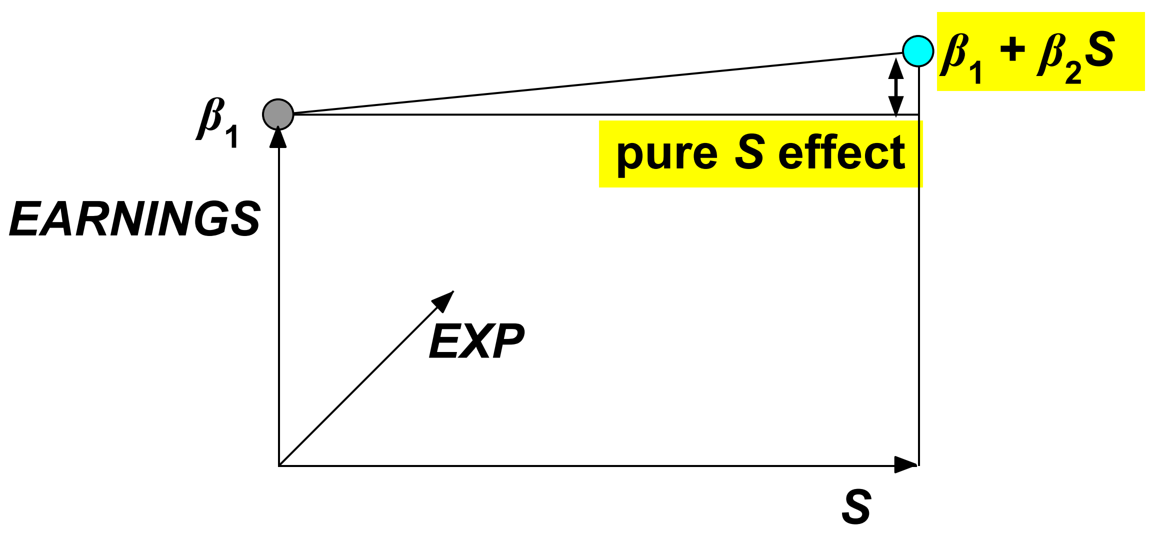
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



В буквальном смысле константа дает заработок *EARNINGS* тем респондентам, у которых нет школьного образования и опыта работы. Однако, не было респондентов с менее чем 6 лет обучения. Следовательно, буквальное толкование β_1 было бы неразумным.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

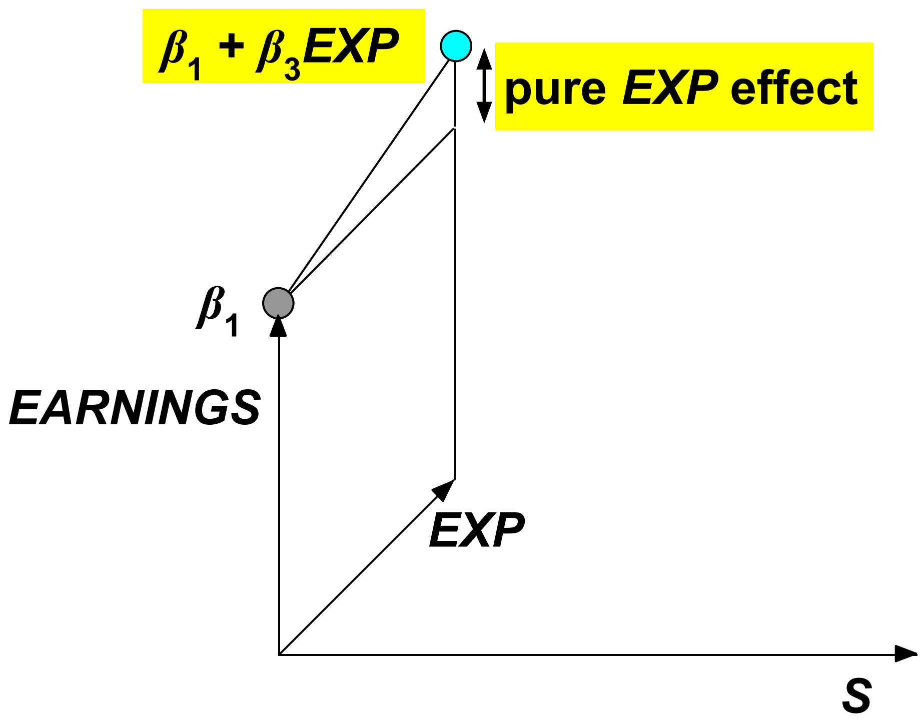
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Следующий член в правой части уравнения дает эффект вариаций в *S*, а увеличение *S* на один год приводит к увеличению *EARNINGS* на β_2 долларов, удерживая постоянную *EXP*.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

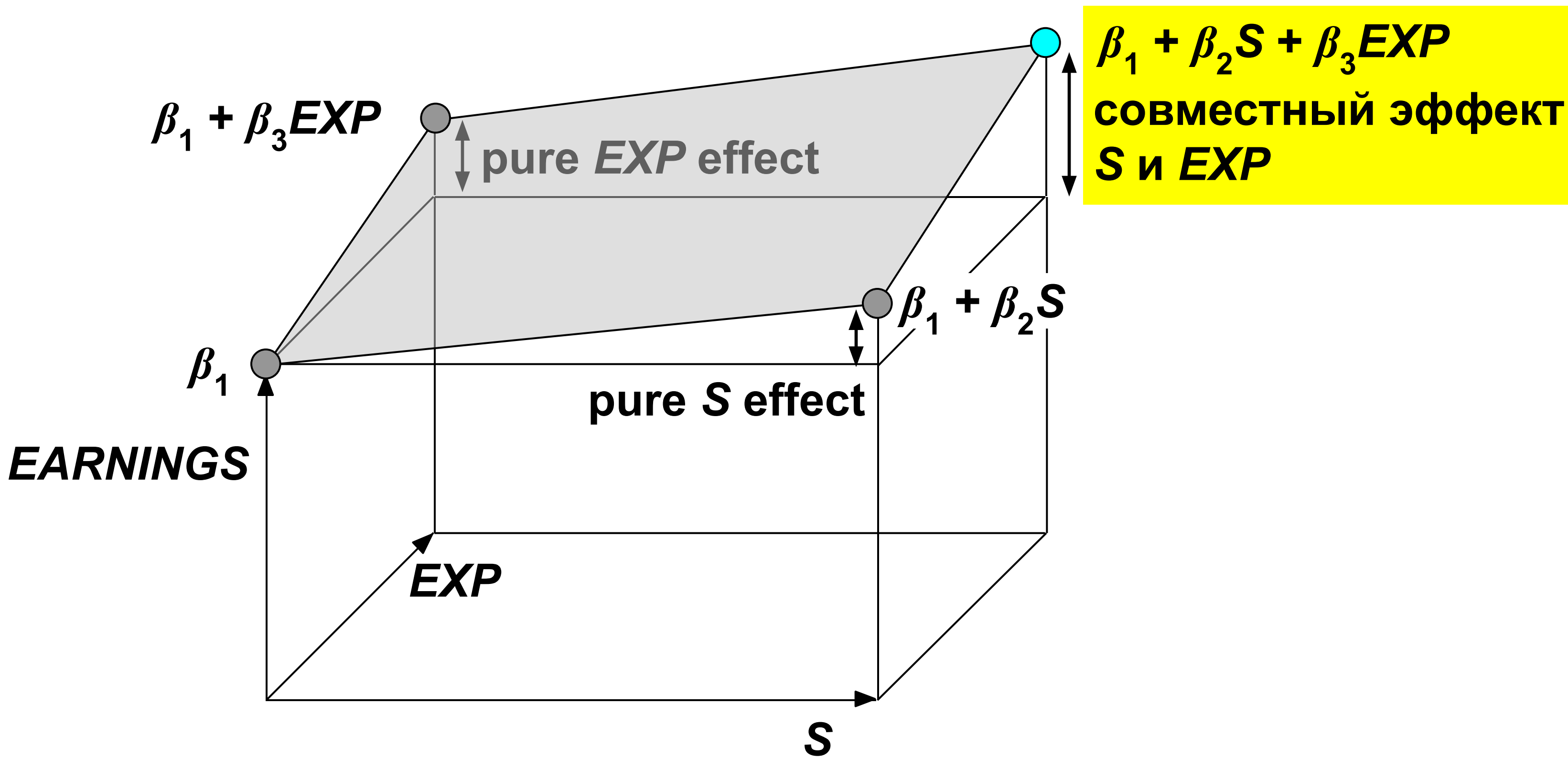
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Аналогичным образом, третий член дает эффект вариаций в *EXP*. Увеличение *EXP* на один год приводит к увеличению прибыли на β_3 долларов, оставляя *S* постоянным.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

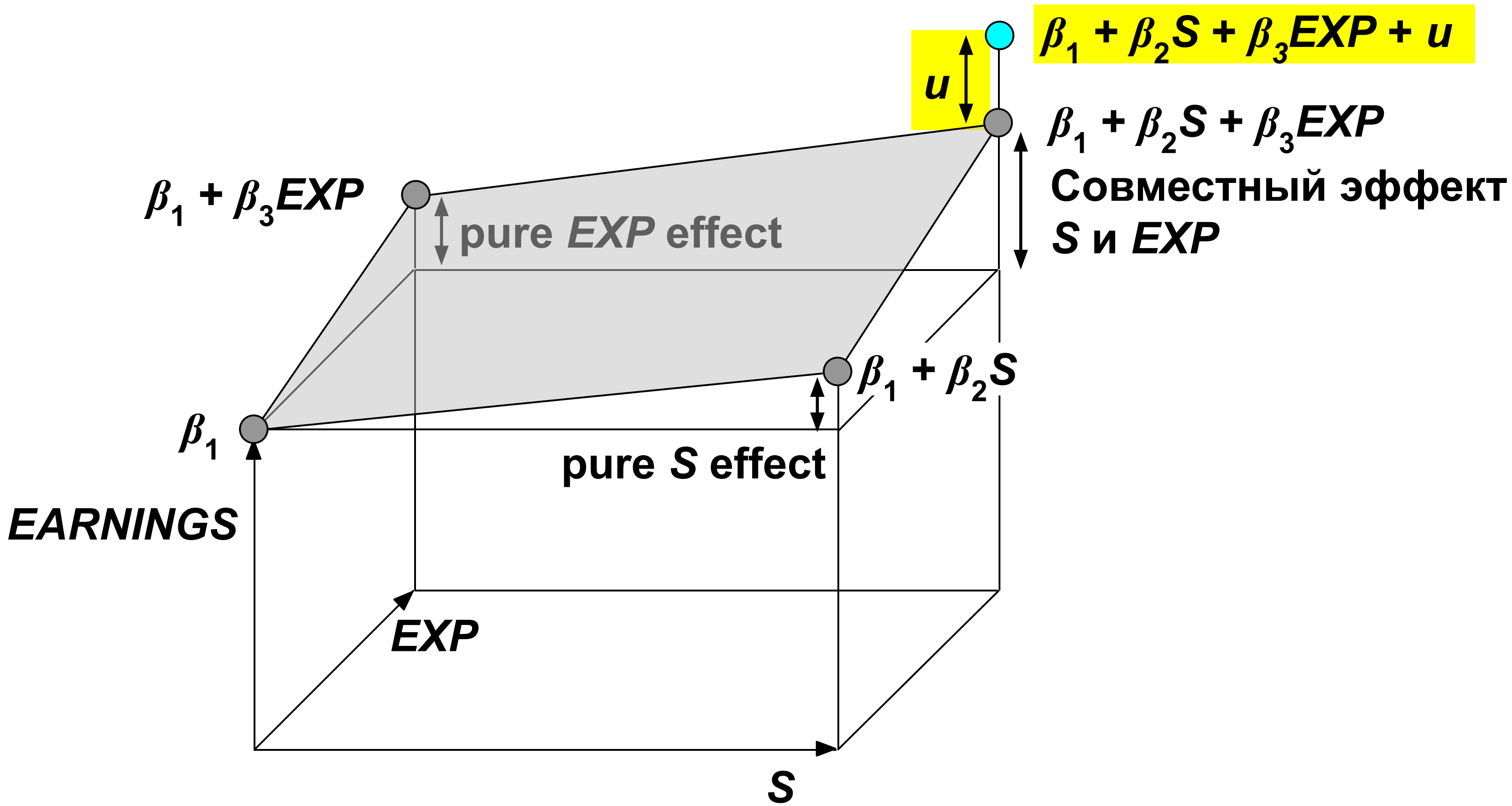
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Различные комбинации S и EXP увеличивают значения *EARNINGS*, которые лежат на плоскости, показанной на диаграмме, определяемой уравнением *EARNINGS* = $\beta_1 + \beta_2 S + \beta_3 EXP$. Это нестохастическая (неслучайная) составляющая модели.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

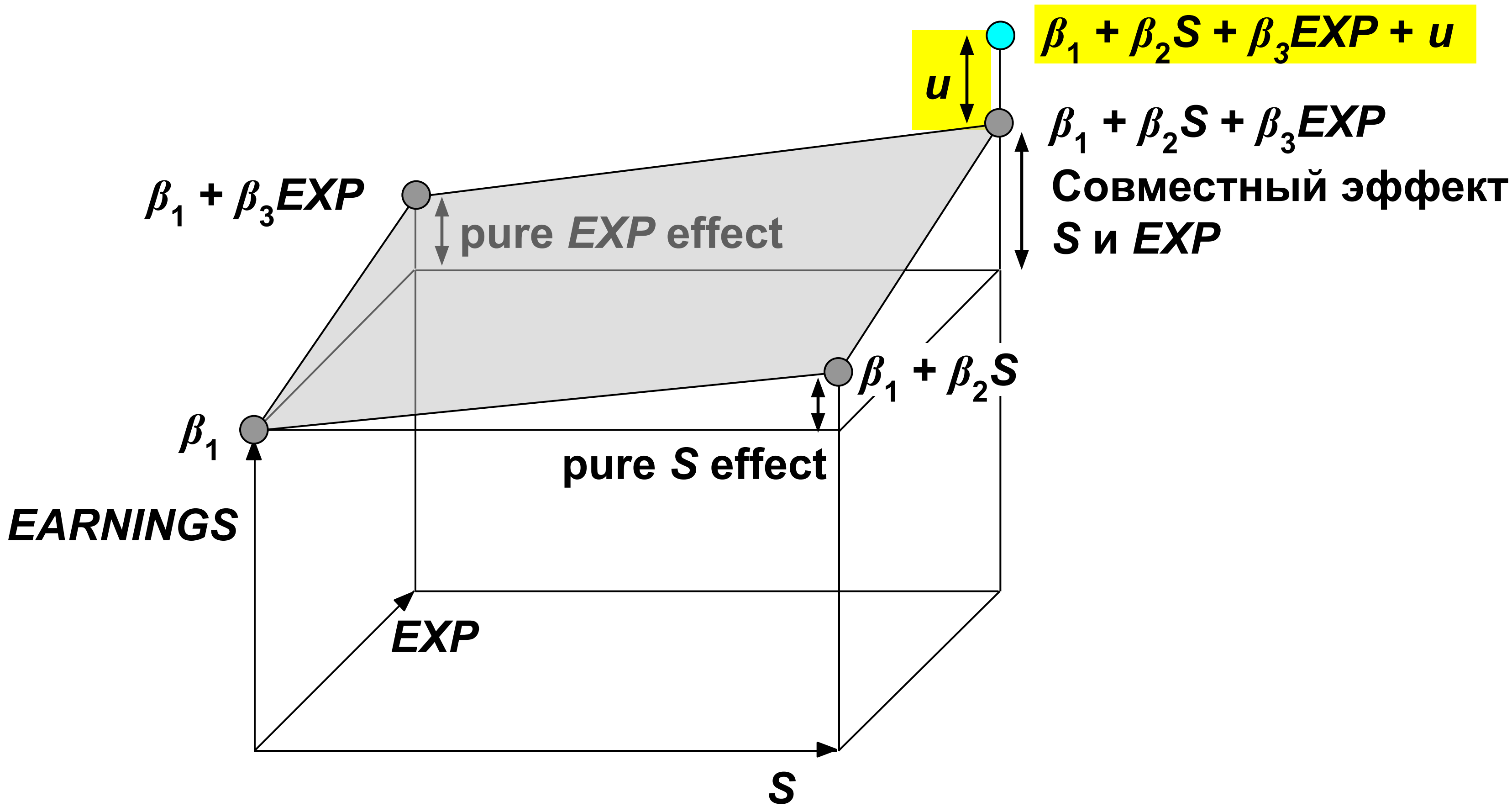
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Последним элементом модели является остаточный член, *u*. Это приводит к тому, что фактические значения *EARNINGS* отклоняются от плоскости. В этом наблюдении *u* имеет положительное значение.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

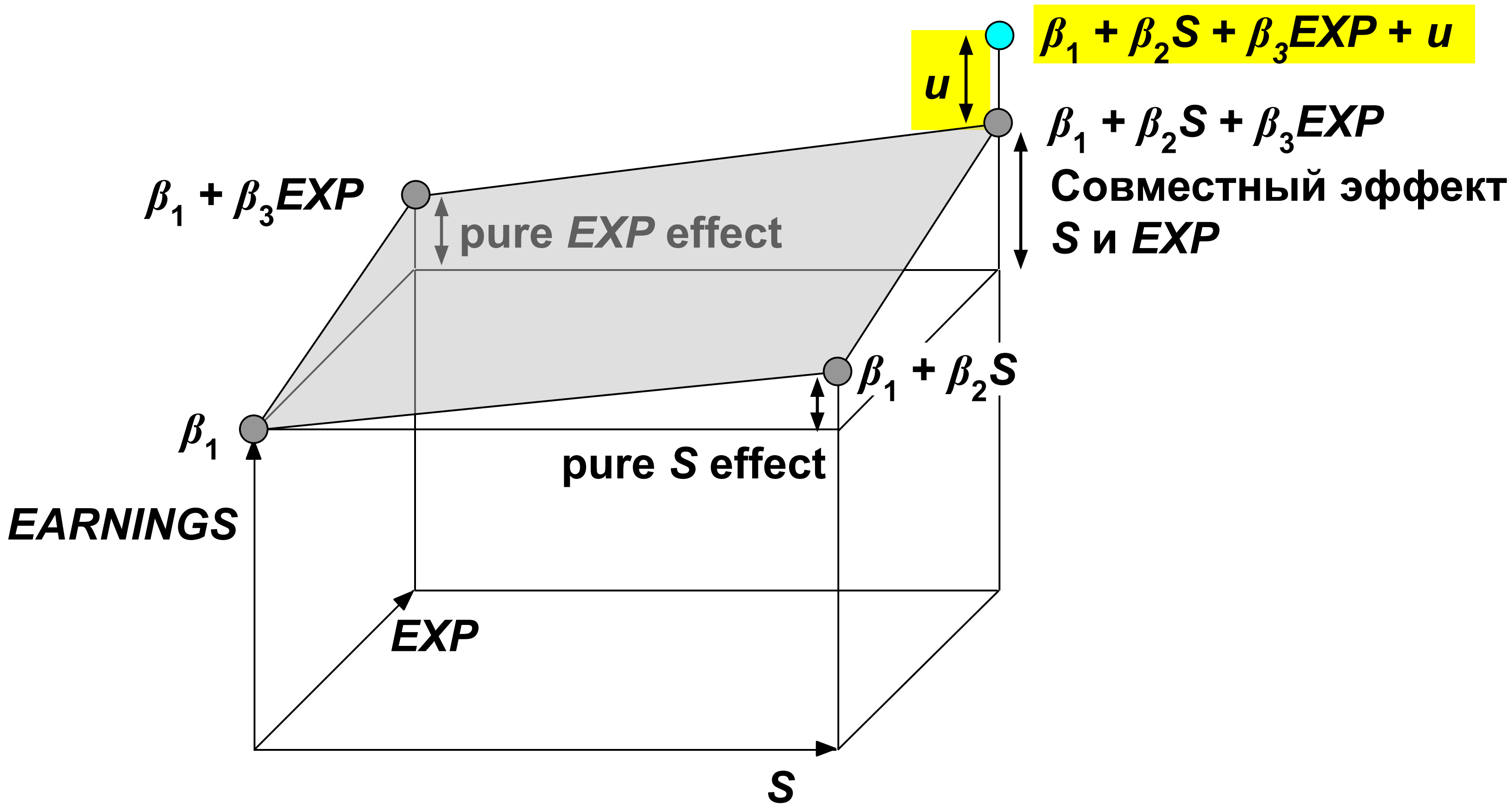
$$EARNINGS = \beta_1 + \beta_2 S + \beta_3 EXP + u$$



Выборка состоит из ряда наблюдений, созданных таким образом. Обратите внимание, что интерпретация модели не зависит от того, коррелированы ли *S* и *EXP* или нет.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

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



Однако мы предполагаем, что влияние *S* и *EXP* на *EARNINGS* являются добавочными. Влияние разницы в *S* на *EARNINGS* не зависит от величины *EXP* или наоборот.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

Коэффициенты регрессии рассчитываются по тому же принципу наименьших квадратов, что и при простом регрессионном анализе. Установленное значение *Y* в наблюдении *i* зависит от нашего выбора b_1, b_2, и b_3.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель **Подогнанная модель**

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \qquad \hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

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

Остаток e_i в наблюдении *i* - это разница между фактическим и установленным значениями *Y*.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

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

$$RSS = \sum \hat{u}_i^2 = \sum \left(Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i} \right)^2$$

Мы определяем RSS, сумму квадратов остатков, и выбираем b_1, b_2, и b_3 , чтобы минимизировать его.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

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

$$\begin{aligned} RSS &= \sum \hat{u}_i^2 = \sum \left(Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i} \right)^2 \\ &= \sum (Y_i^2 + b_1^2 + b_2^2 X_{2i}^2 + b_3^2 X_{3i}^2 - 2b_1 Y_i - 2b_2 X_{2i} Y_i \\ &\quad - 2b_3 X_{3i} Y_i + 2b_1 b_2 X_{2i} + 2b_1 b_3 X_{3i} + 2b_2 b_3 X_{2i} X_{3i}) \\ &= \sum Y_i^2 + n b_1^2 + b_2^2 \sum X_{2i}^2 + b_3^2 \sum X_{3i}^2 - 2b_1 \sum Y_i \\ &\quad - 2b_2 \sum X_{2i} Y_i - 2b_3 \sum X_{3i} Y_i + 2b_1 b_2 \sum X_{2i} \\ &\quad + 2b_1 b_3 \sum X_{3i} + 2b_2 b_3 \sum X_{2i} X_{3i} \end{aligned}$$

Сначала мы раскладываем RSS как показано сверху.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

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

$$\begin{aligned} RSS = {} & \sum Y_i^2 + n b_1^2 + b_2^2 \sum X_{2i}^2 + b_3^2 \sum X_{3i}^2 - 2b_1 \sum Y_i \\ & - 2b_2 \sum X_{2i} Y_i - 2b_3 \sum X_{3i} Y_i + 2b_1 b_2 \sum X_{2i} \\ & + 2b_1 b_3 \sum X_{3i} + 2b_2 b_3 \sum X_{2i} X_{3i} \end{aligned}$$

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

Затем мы используем условия первого порядка для его минимизации.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

$$\hat{\beta}_2 = \frac{\sum (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}) \sum (X_{3i} - \bar{X}_3)^2 - \sum (X_{3i} - \bar{X}_3)(Y_i - \bar{Y}) \sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)}{\sum (X_{2i} - \bar{X}_2)^2 \sum (X_{3i} - \bar{X}_3)^2 - \left(\sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)\right)^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

Таким образом, мы получаем три уравнения на три неизвестных. Решая эти уравнения, получаем выражения для конкретных значений, удовлетворяющих критерию МНК. (Выражение для $\hat{\beta}_2$ совпадает с выражением для $\hat{\beta}_3$, с индексами 2 и 3, которые поменяны местами.)

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

$$\hat{\beta}_2 = \frac{\sum (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}) \sum (X_{3i} - \bar{X}_3)^2 - \sum (X_{3i} - \bar{X}_3)(Y_i - \bar{Y}) \sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)}{\sum (X_{2i} - \bar{X}_2)^2 \sum (X_{3i} - \bar{X}_3)^2 - \left(\sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) \right)^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

Выражение для $\hat{\beta}_1$ является прямым продолжением выражения для него в простом регрессионном анализе.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

$$\hat{\beta}_2 = \frac{\sum\left(X_{2i} - \bar{X}_2\right)\left(Y_i - \bar{Y}\right)\sum\left(X_{3i} - \bar{X}_3\right)^2 - \sum\left(X_{3i} - \bar{X}_3\right)\left(Y_i - \bar{Y}\right)\sum\left(X_{2i} - \bar{X}_2\right)\left(X_{3i} - \bar{X}_3\right)}{\sum\left(X_{2i} - \bar{X}_2\right)^2\sum\left(X_{3i} - \bar{X}_3\right)^2 - \left(\sum\left(X_{2i} - \bar{X}_2\right)\left(X_{3i} - \bar{X}_3\right)\right)^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

Однако выражения для коэффициентов наклона значительно сложнее, чем для коэффициента наклона в простом регрессионном анализе.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

Истинная модель

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

Подогнанная модель

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

$$\hat{\beta}_2 = \frac{\sum (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}) \sum (X_{3i} - \bar{X}_3)^2 - \sum (X_{3i} - \bar{X}_3)(Y_i - \bar{Y}) \sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3)}{\sum (X_{2i} - \bar{X}_2)^2 \sum (X_{3i} - \bar{X}_3)^2 - \left(\sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) \right)^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

Для общего случая, когда существует много объясняющих переменных, обычная алгебра непригодна. Необходимо перейти на матричную алгебру.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

```
. reg EARNINGS S EXP

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =   35.24
       Model |  8735.42401     2   4367.712            Prob > F      =  0.0000
    Residual |  61593.5422   497  123.930668           R-squared     =  0.1242
-------------+------------------------------           Adj R-squared =  0.1207
       Total |  70328.9662   499  140.939812           Root MSE      =  11.132
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.877563   .2237434     8.39   0.000     1.437964    2.317163
         EXP |   .9833436   .2098457     4.69   0.000     .5710495    1.395638
       _cons |  -14.66833   4.288375    -3.42   0.001    -23.09391   -6.242752
----------------------------------------------------------------------------
```

$$\widehat{EARNINGS} = -14.67 + 1.88S + 0.98EXP$$

Вот результат регрессии для уравнения заработной платы с использованием Data Set 21.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

```
. reg EARNINGS S EXP

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =   35.24
       Model |  8735.42401     2   4367.712            Prob > F      =  0.0000
    Residual |  61593.5422   497  123.930668           R-squared     =  0.1242
-------------+------------------------------           Adj R-squared =  0.1207
       Total |  70328.9662   499  140.939812           Root MSE      =  11.132
------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   1.877563   .2237434     8.39   0.000     1.437964    2.317163
         EXP |   .9833436   .2098457     4.69   0.000     .5710495    1.395638
       _cons |  -14.66833   4.288375    -3.42   0.001    -23.09391   -6.242752
------------------------------------------------------------------------------
```

$$\widehat{EARNINGS} = -14.67 + 1.88S + 0.98EXP$$

Это означает, что почасовая заработная плата увеличится на \$1.88 за каждый дополнительный год школьного обучения и на \$0.98 за каждый дополнительный год опыта работы.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

```
. reg EARNINGS S EXP

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =    500
-------------+------------------------------           F(  2,   497) =  35.24
       Model |  8735.42401     2    4367.712           Prob > F      = 0.0000
    Residual |  61593.5422   497  123.930668           R-squared     = 0.1242
-------------+------------------------------           Adj R-squared = 0.1207
       Total |  70328.9662   499  140.939812           Root MSE      = 11.132
----------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   1.877563   .2237434     8.39   0.000     1.437964    2.317163
         EXP |   .9833436   .2098457     4.69   0.000     .5710495    1.395638
       _cons |  -14.66833   4.288375    -3.42   0.001    -23.09391   -6.242752
----------------------------------------------------------------------------
```

$$\widehat{EARNINGS} = -14.67 + 1.88S + 0.98EXP$$

В буквальном смысле, константа показывает, что человек, который не имеет образования или опыта работы, будет иметь почасовой заработок, равный -14.67 $.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ: ПРИМЕР

```
. reg EARNINGS S EXP

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =   35.24
       Model |  8735.42401     2   4367.712            Prob > F      =  0.0000
    Residual |  61593.5422   497  123.930668           R-squared     =  0.1242
-------------+------------------------------           Adj R-squared =  0.1207
       Total |  70328.9662   499  140.939812           Root MSE      =  11.132
------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   1.877563   .2237434     8.39   0.000     1.437964    2.317163
         EXP |   .9833436   .2098457     4.69   0.000     .5710495    1.395638
       _cons |  -14.66833   4.288375    -3.42   0.001    -23.09391   -6.242752
------------------------------------------------------------------------------
```

$$\widehat{EARNINGS} = -14.67 + 1.88S + 0.98EXP$$

Очевидно, что это невозможно. Наименьшее значение *S* в выборке составило 8. Мы получили бессмысленную оценку, поскольку экстраполировали слишком далеко от диапазона данных.