

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{Y}_i = b_1 + b_2 X_i$$

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

Предположим, что единицы измерения Y или X изменены. Как это повлияет на результаты регрессии? Интуитивно мы ожидаем, что ничего существенного не произойдет. Так оно и есть.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{Y}_i = b_1 + b_2 X_i$$

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

Мы продемонстрируем это для оценок коэффициентов регрессии. Начнем с предположения, что истинная и расчетная модели приведены выше.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

Предположим теперь, что единицы измерения Y изменяются, причем новая мера Y^* , является линейной функцией старой.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

Как правило, изменение единицы измерения включает в себя простое мультипликативное масштабирование, например, когда мы конвертируем фунты в граммы.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

Иногда происходит линейное преобразование. Примером может служить конверсия температур от градусов Цельсия до градусов Фаренгейта.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{Y}_i = b_1 + b_2 X_i$$

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

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

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

Показан новый коэффициент наклона b_2^* .

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

$$b_2^* = \frac{\sum (X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})([\lambda_1 + \lambda_2 Y_i] - [\lambda_1 + \lambda_2 \bar{Y}])}{\sum (X_i - \bar{X})^2}$$

Подставим Y^* .

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

$$b_2^* = \frac{\sum (X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})([\lambda_1 + \lambda_2 Y_i] - [\lambda_1 + \lambda_2 \bar{Y}])}{\sum (X_i - \bar{X})^2}$$
$$= \frac{\sum (X_i - \bar{X})(\lambda_2 Y_i - \lambda_2 \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\lambda_2 \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Остается только λ_2 .

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

$$b_2^* = \frac{\sum (X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})([\lambda_1 + \lambda_2 Y_i] - [\lambda_1 + \lambda_2 \bar{Y}])}{\sum (X_i - \bar{X})^2}$$
$$= \frac{\sum (X_i - \bar{X})(\lambda_2 Y_i - \lambda_2 \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\lambda_2 \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \lambda_2 b_2$$

Мы находим, что новый коэффициент наклона равен исходному, умноженному на λ_2 .

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

$$b_2^* = \frac{\sum (X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})([\lambda_1 + \lambda_2 Y_i] - [\lambda_1 + \lambda_2 \bar{Y}])}{\sum (X_i - \bar{X})^2}$$
$$= \frac{\sum (X_i - \bar{X})(\lambda_2 Y_i - \lambda_2 \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\lambda_2 \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \lambda_2 b_2$$

Это логично. Единичное изменение в Y такое же, как изменение λ_2 единиц в Y^* . Согласно уравнению регрессии, единичное изменение X приводит к изменению b_2 единиц в Y , поэтому это должно привести к изменению $\lambda_2 b_2$ единиц в Y^* .

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$Y_i^* = \lambda_1 + \lambda_2 Y_i$$

$$\hat{Y}_i^* = b_1^* + b_2^* X_i$$

$$b_2^* = \frac{\sum (X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})([\lambda_1 + \lambda_2 Y_i] - [\lambda_1 + \lambda_2 \bar{Y}])}{\sum (X_i - \bar{X})^2}$$
$$= \frac{\sum (X_i - \bar{X})(\lambda_2 Y_i - \lambda_2 \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\lambda_2 \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \lambda_2 b_2$$

Эффект изменения единиц измерения X оставлен в качестве упражнения.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{Y}_i = b_1 + b_2 X_i$$

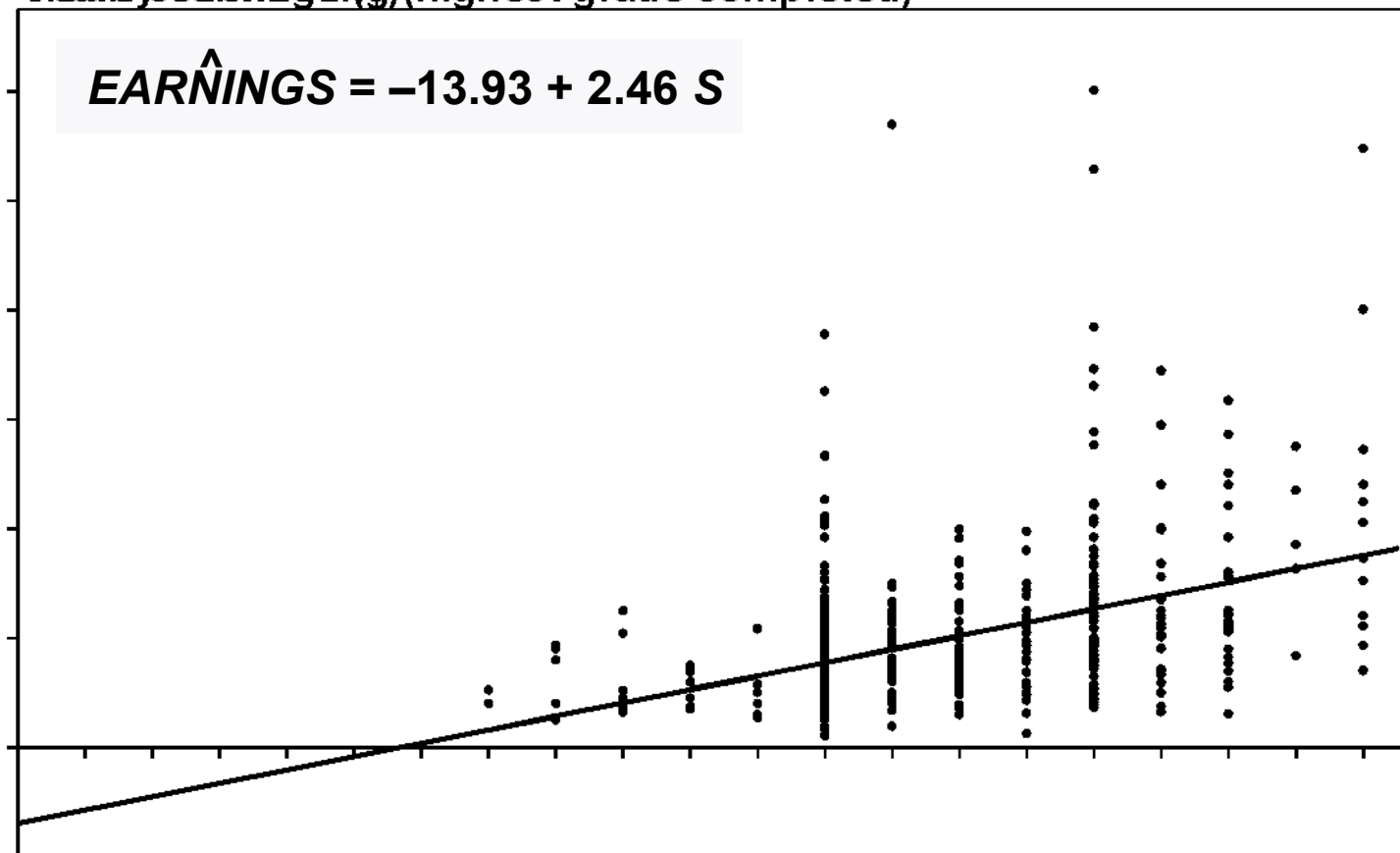
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Мы рассмотрим частный случай изменения единиц измерения X . Часто свободный член в уравнении регрессии не имеет разумной интерпретации, потому что $X = 0$ расположен далеко от диапазона данных.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

IID

Years of schooling (S) (highest grade completed)



Предыдущая презентация является примером, поскольку свободный член является отрицательным.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

```
. reg EARNINGS S
```

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

Вот результат построения регрессии.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

```
. reg EARNINGS S
```

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

Предыдущая презентация.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$X_i^* = X_i - \bar{X}$$

$$\hat{Y}_i = b_1^* + b_2^* X_i^*$$

Можно справиться с этой проблемой, определив X^* как отклонение X от его среднего значения.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$X_i^* = X_i - \bar{X}$$

$$\hat{Y}_i = b_1^* + b_2^* X_i^*$$

$$\begin{aligned}\sum X_i^* &= \sum (X_i - \bar{X}) \\ &= \sum X_i - n\bar{X} \\ &= n\bar{X} - n\bar{X} = 0\end{aligned}$$

$$\bar{X}^* = \frac{1}{n} \sum X_i^* = 0$$

Заметим, что по определению сумма X_i^* равна 0, и, следовательно, среднее значение X^* равно 0.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$X_i^* = X_i - \bar{X}$$

$$\hat{Y}_i = b_1^* + b_2^* X_i^*$$

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

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

$$\bar{X}^* = 0$$

Если мы ищем уравнение регрессии Y от X^* вместо X , то коэффициент наклона не изменяется.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\hat{Y}_i = b_1 + b_2 X_i$$
$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$X_i^* = X_i - \bar{X}$$

$$\hat{Y}_i = b_1^* + b_2^* X_i^*$$

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

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

$$\bar{X}^* = 0$$

$$b_1^* = \bar{Y} - b_2^* \bar{X}^* = \bar{Y}$$

Теперь свободный член представляет собой среднее значение Y при среднем значении X .

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

```
-----  
. sum S
```

Variable	Obs	Mean	Std. Dev.	Min	Max
S	540	13.67222	2.438476	7	20

```
. gen SDEV = S - 13.67
```

Мы находим среднее значение выборки (в Stata мы используем команду «sum»). Мы находим, что среднее составляет 13.67 лет, и мы определяем новую переменную SDEV путем вычитания 13.67 из S.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

```
. reg EARNINGS SDEV
```

Source	SS	df	MS			
Model	19321.5587	1	19321.5587	Number of obs =	540	
Residual	92688.6723	538	172.283778	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
SDEV	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	19.63077	.5648401	34.75	0.000	18.5212	20.74033

Вот результат с использованием SDEV вместо S. Свободный член = 19.63, теперь дает прогнозируемый доход тех, кто имеет среднее образование.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

. reg EARNINGS SDEV

Source	SS	df	MS	Number of obs = 540		
Model	19321.5587	1	19321.5587	F(1, 538)	=	112.15
Residual	92688.6723	538	172.283778	Prob > F	=	0.0000
-----				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]	
SDEV	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	19.63077	.5648401	34.75	0.000	18.5212	20.74033

. reg EARNINGS S

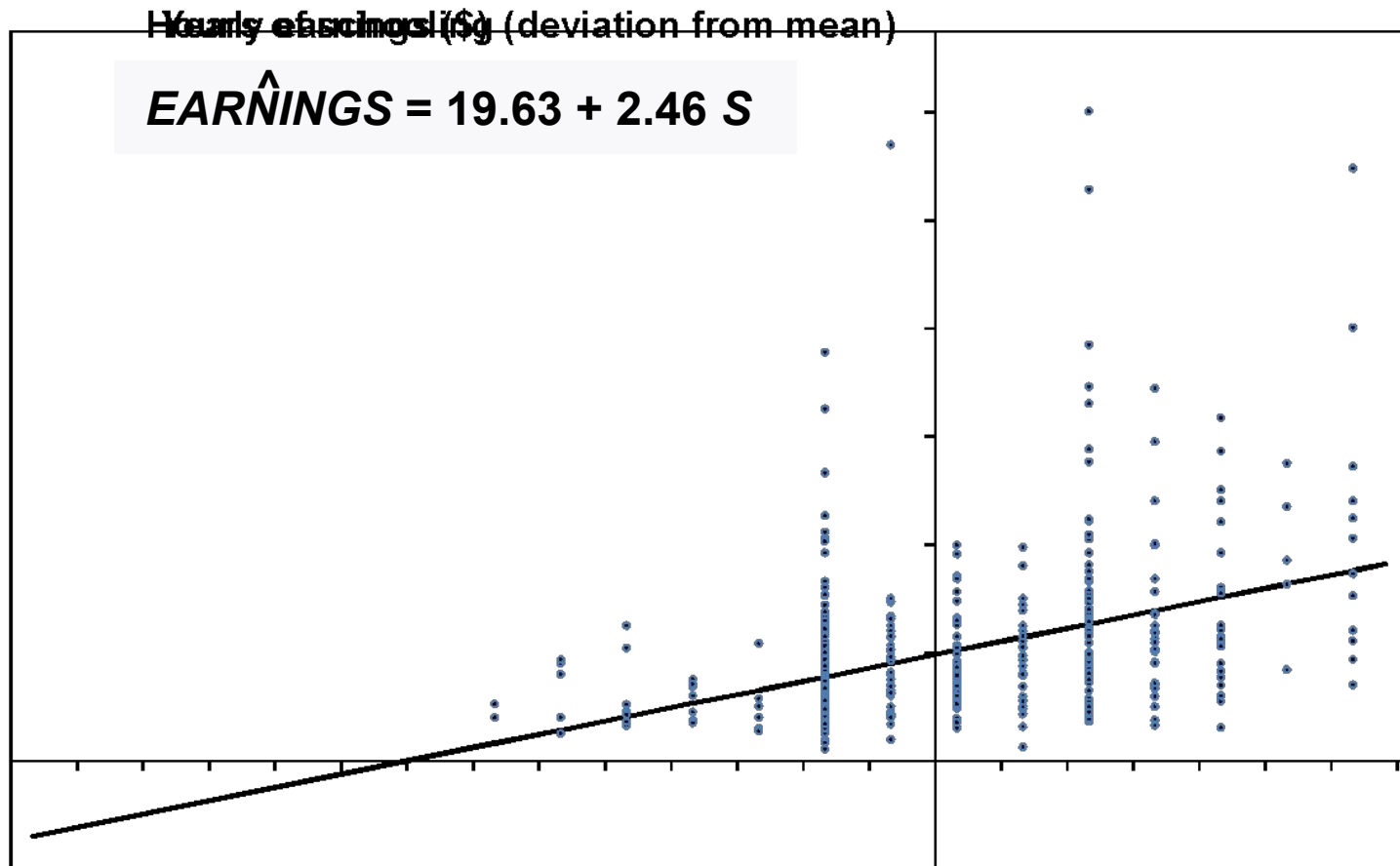
Source	SS	df	MS	Number of obs = 540		
Model	19321.5589	1	19321.5589	F(1, 538)	=	112.15
Residual	92688.6722	538	172.283777	Prob > F	=	0.0000
-----				R-squared	=	0.1725
-----				Adj R-squared	=	0.1710
Total	112010.231	539	207.811189	Root MSE	=	13.126

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

Сравнивая новый результат с оригиналом, мы видим, что помимо стандартной ошибки и t статистики свободного члена ничего больше не изменилось.

ИЗМЕНЕНИЯ В ЕДИНИЦАХ ИЗМЕРЕНИЯ

11)



Единственным эффектом изменения S является перемещение вертикальной оси до точки, которая раньше была 13.67. Как следствие, свободный член становится 19.63. В остальном линия регрессии не изменяется