

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

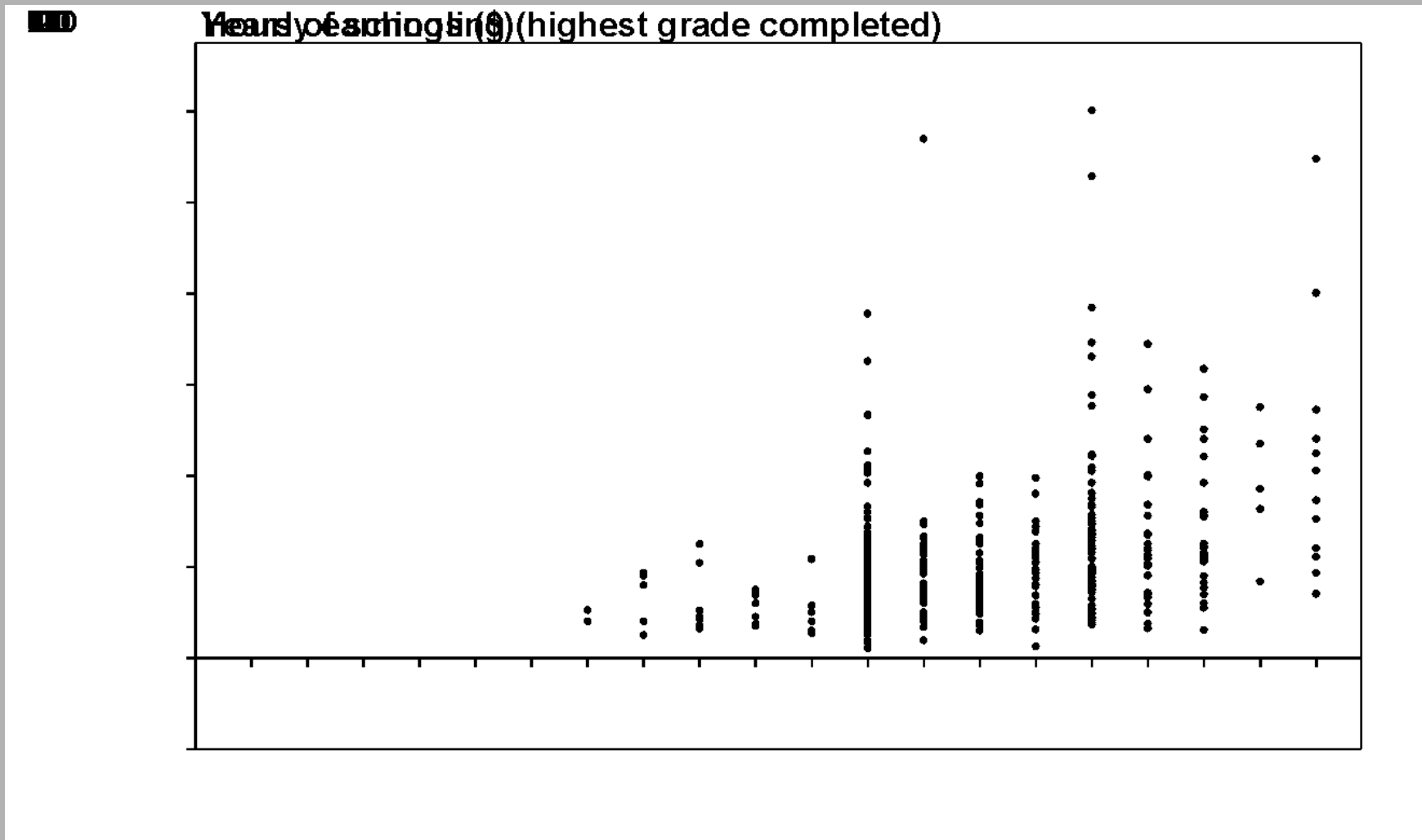
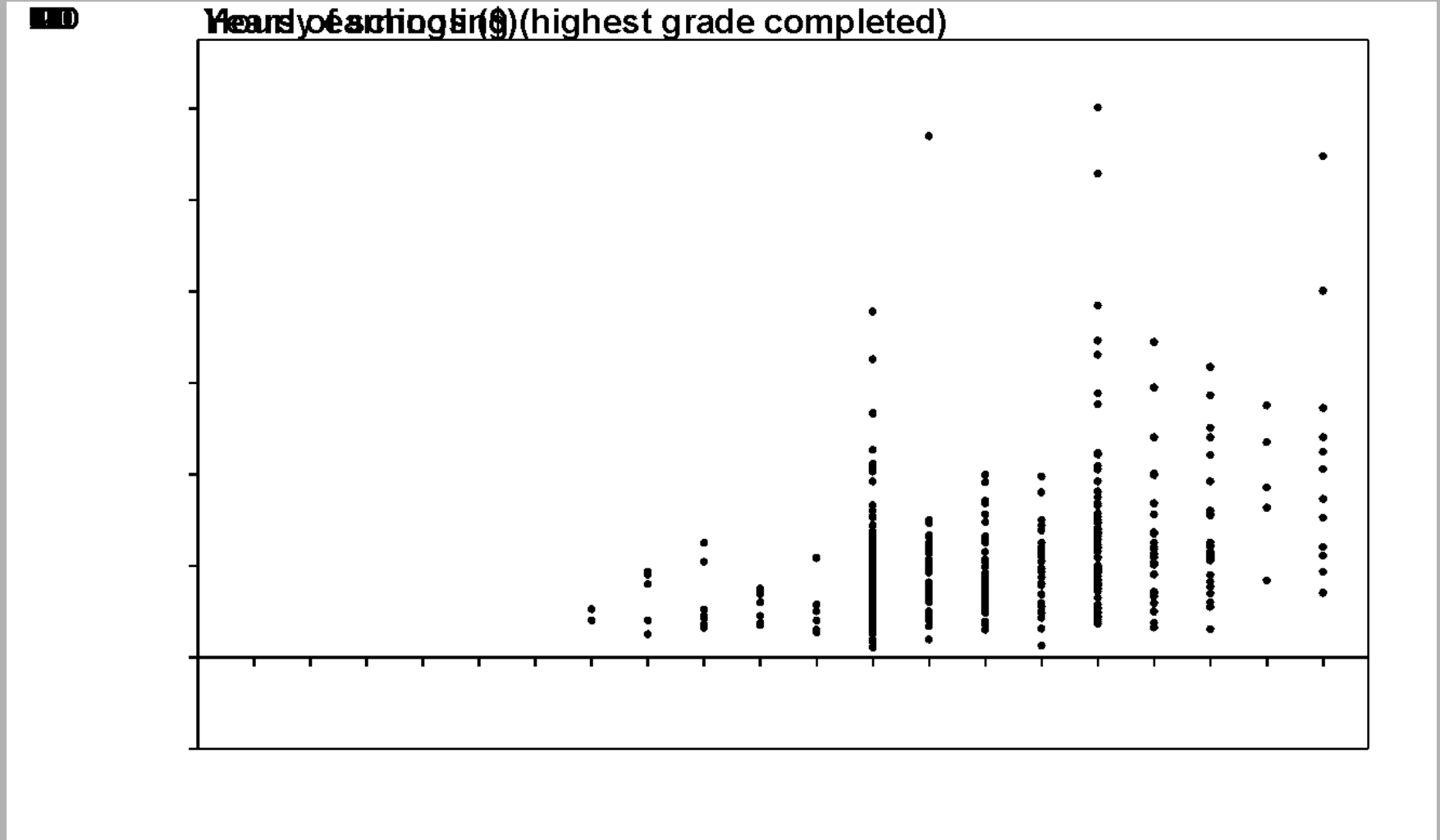


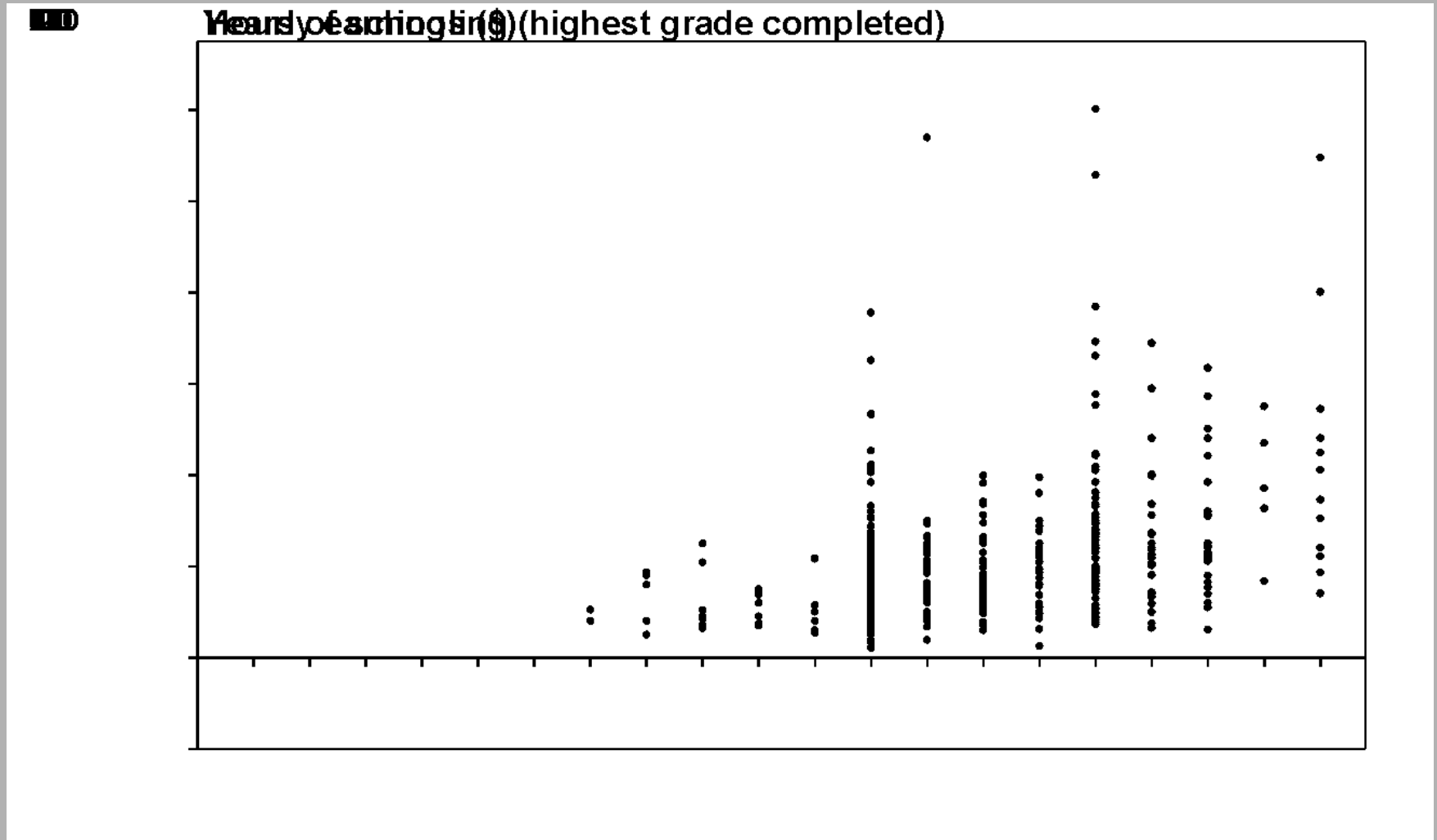
Диаграмма разброса показывает часовую зарплату в 2002 году, составленную по сравнению с годами обучения для выборки из 540 респондентов из Национального опроса молодежи в 1979 г.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ



Оценки 7-12 означают обучение в начальной и средней школах. Оценки 13, 14 и 15 означают завершение одного, двух и трех лет колледжа.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ



Оценка 16 означает завершение четырехлетнего колледжа. Дальнейшие года означают годы последипломного образования.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

```
. 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

Это результат регрессии доходов по годам обучения, используется программа Stata.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

```
. 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

Пока мы будем иметь дело только с оценками параметров. Переменные в регрессии перечислены в первом столбце, а во втором столбце приведены их оценки.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

```
. 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	

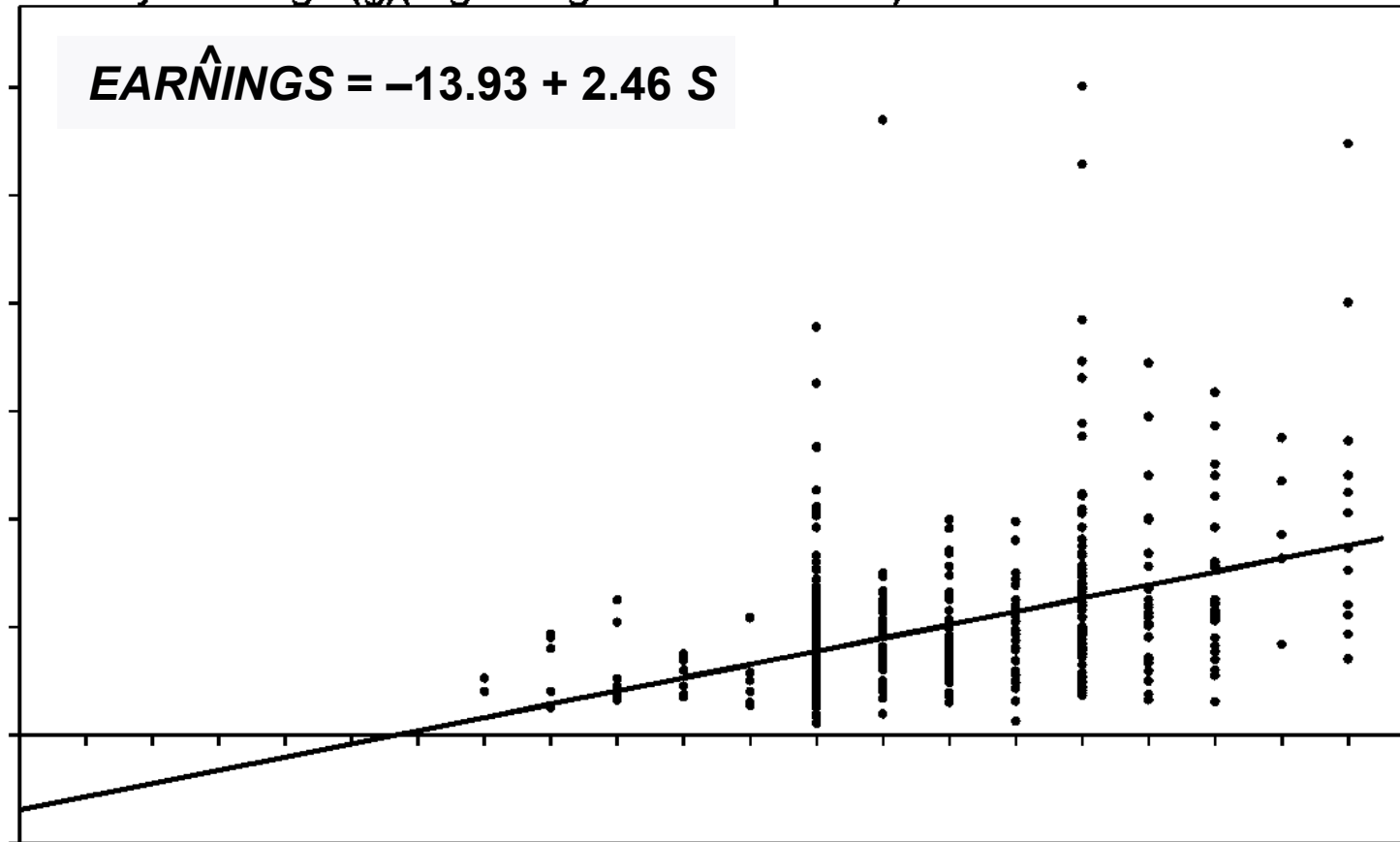
EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

В нашем случае есть только одна переменная S, а ее оценка равна 2.46. Свободный член равен константе. Оценка константы составляет -13,93.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)



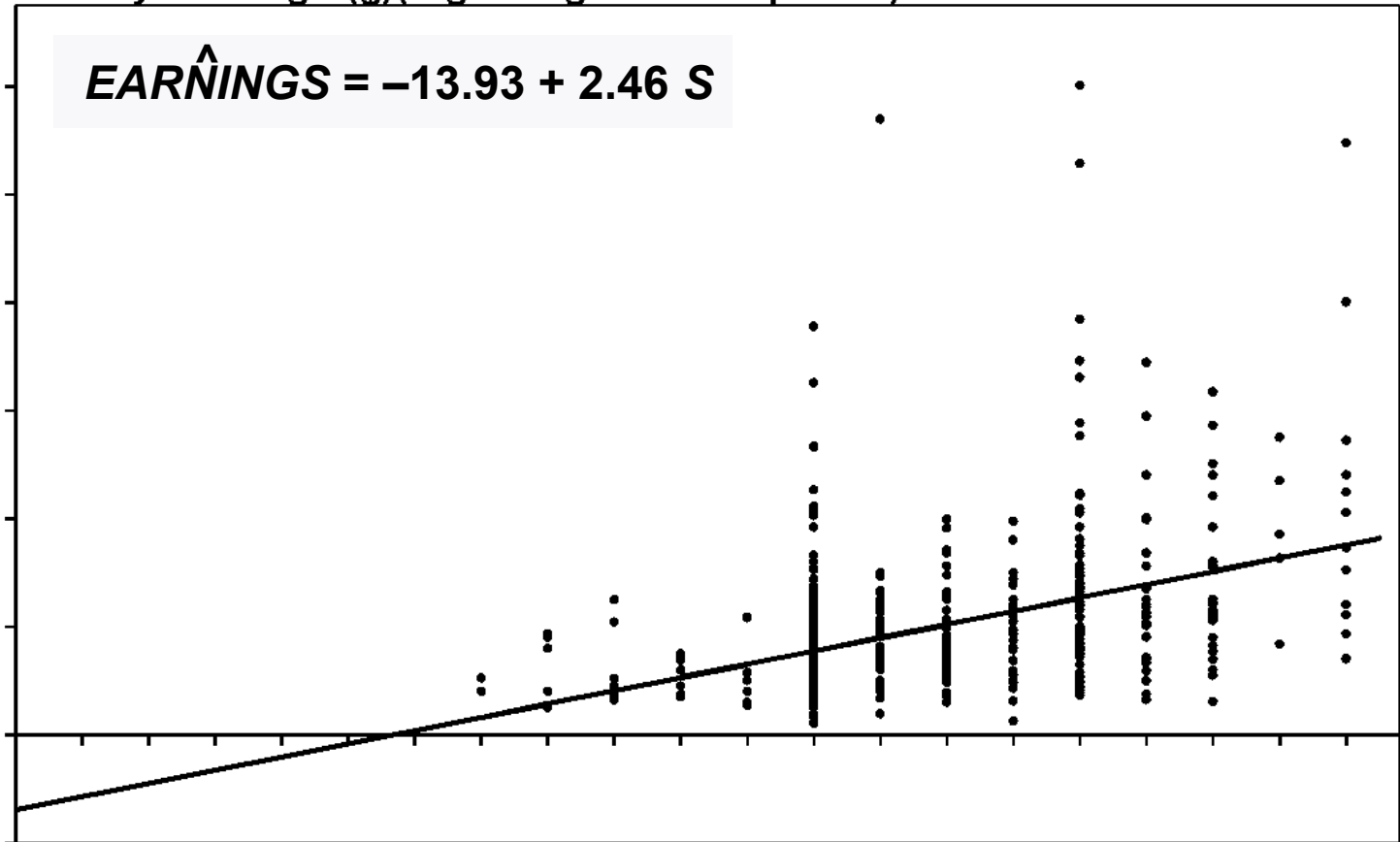
Приведена диаграмма рассеяния, показана линия регрессии.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)

$$\hat{EARNINGS} = -13.93 + 2.46 S$$



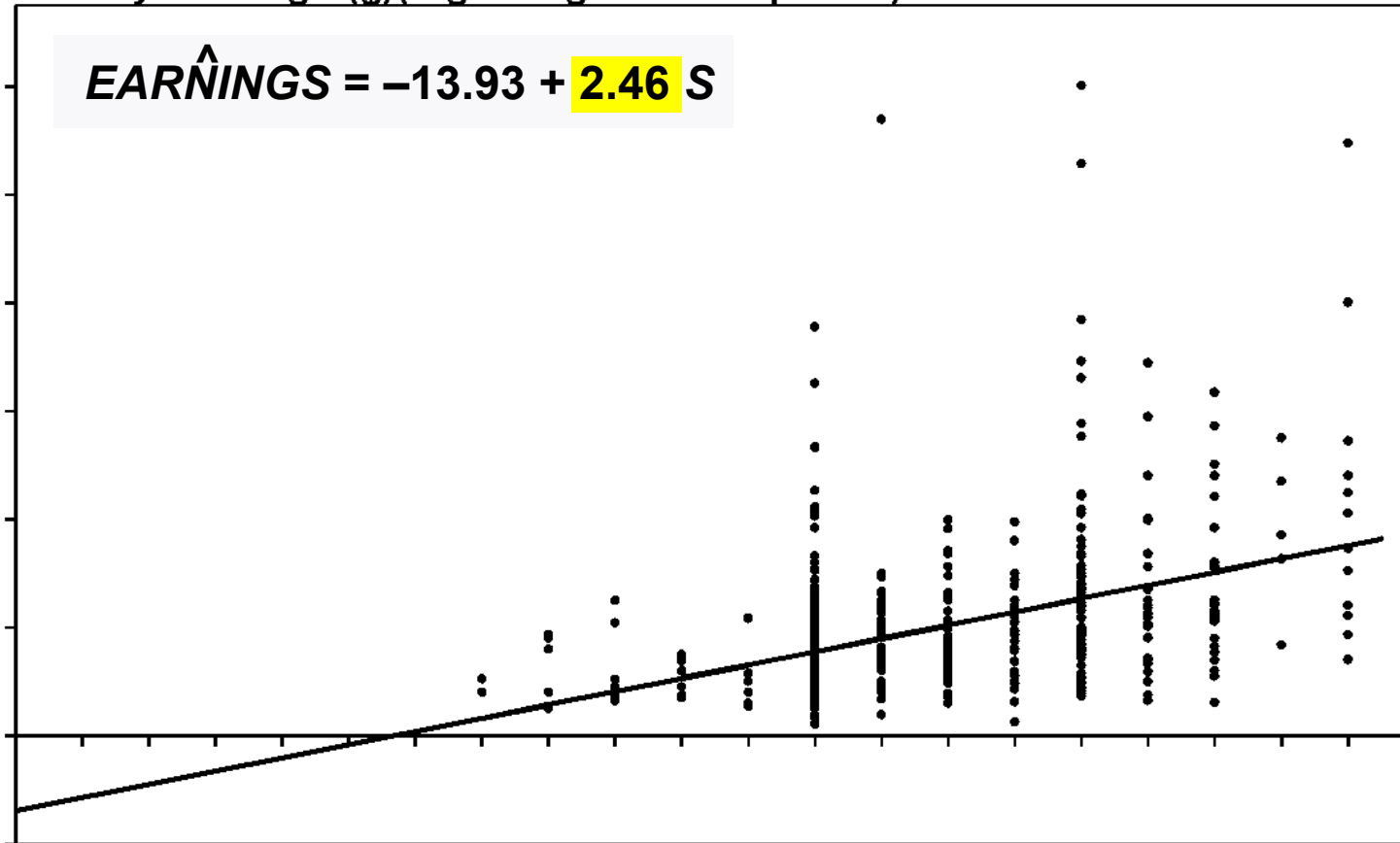
Что на самом деле означают коэффициенты?

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)

$$\hat{EARNINGS} = -13.93 + 2.46 S$$

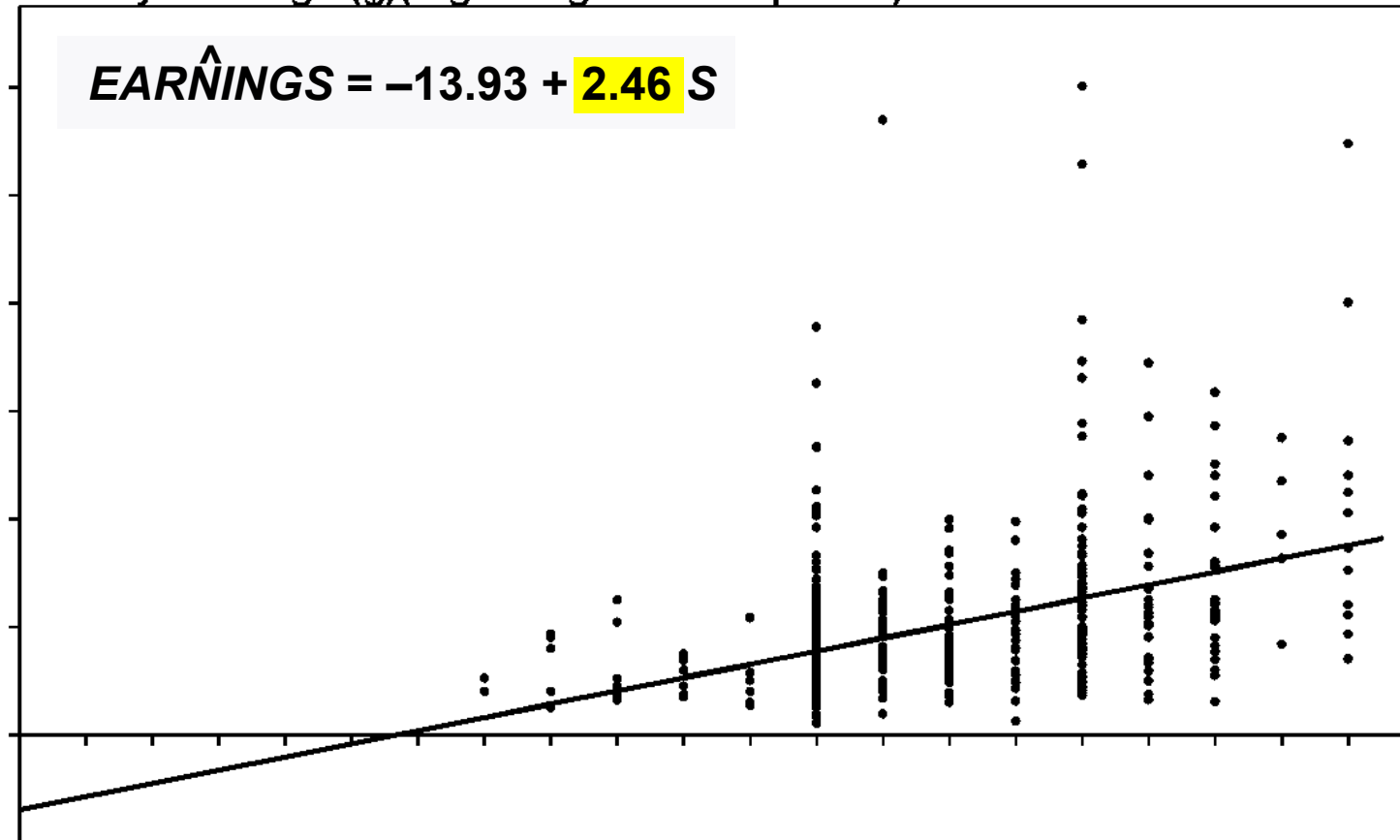


Чтобы ответить на этот вопрос, вы должны обратиться к единицам измерения переменных.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)



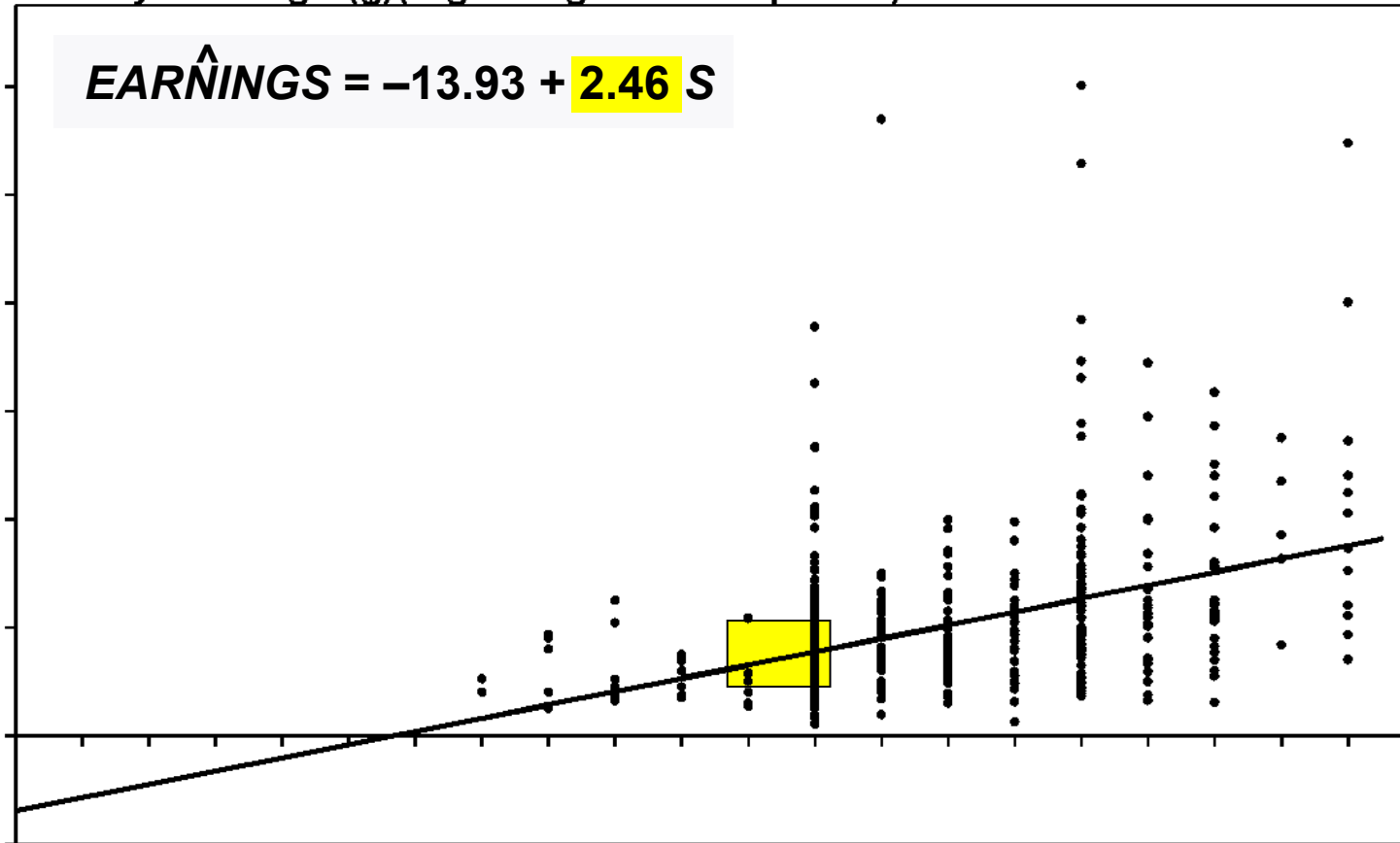
S измеряется в годах, $EARNINGS$ - в долларах в час. Таким образом, коэффициент наклона показывает, что почасовой доход увеличивается на \$ 2,46 за каждый дополнительный год обучения.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

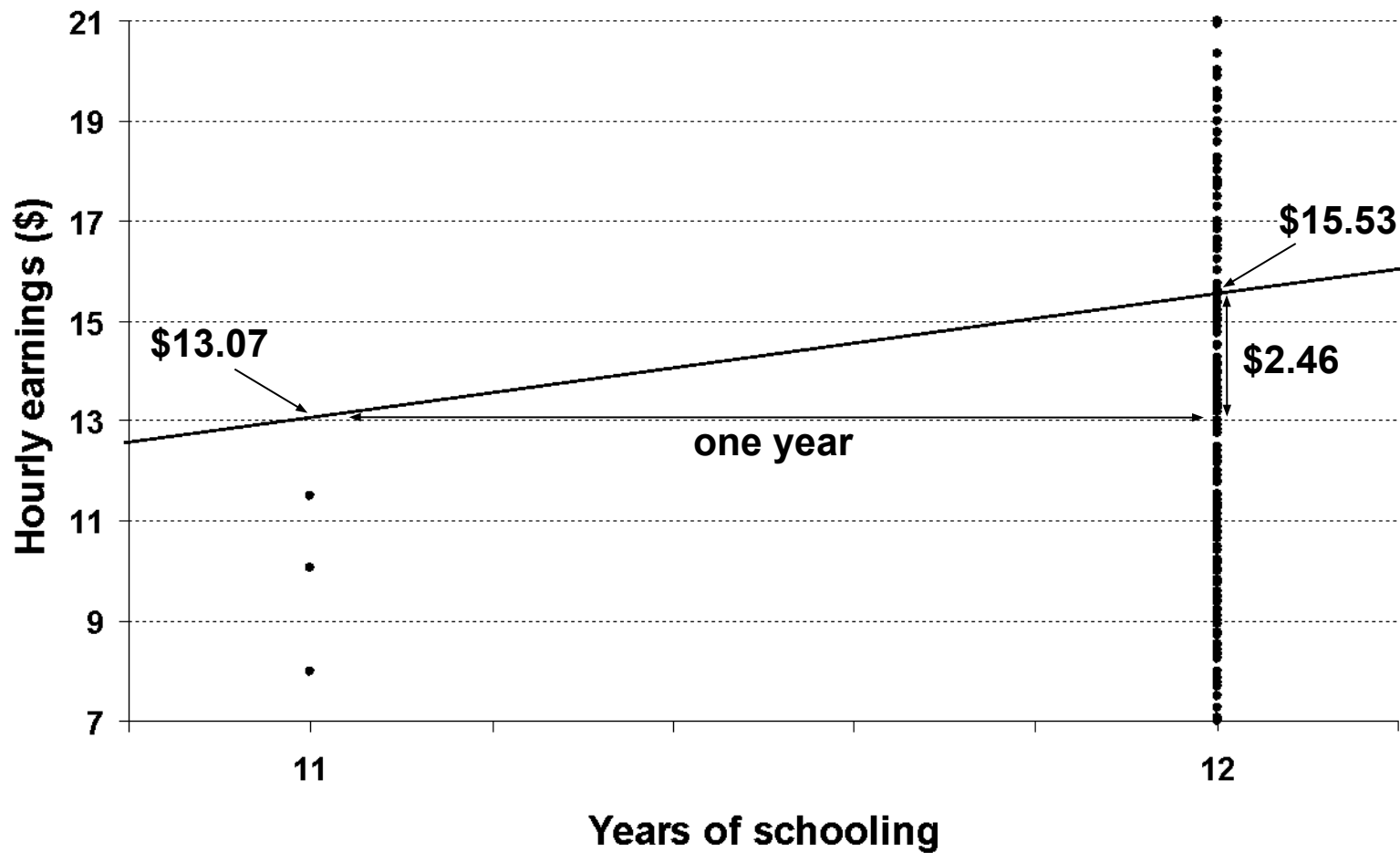
Years of schooling (S) (highest grade completed)

$$\hat{EARNINGS} = -13.93 + 2.46 S$$



Мы рассмотрим геометрическое представление этой интерпретации. Чтобы сделать это, мы увеличим отмеченный раздел диаграммы рассеяния.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ



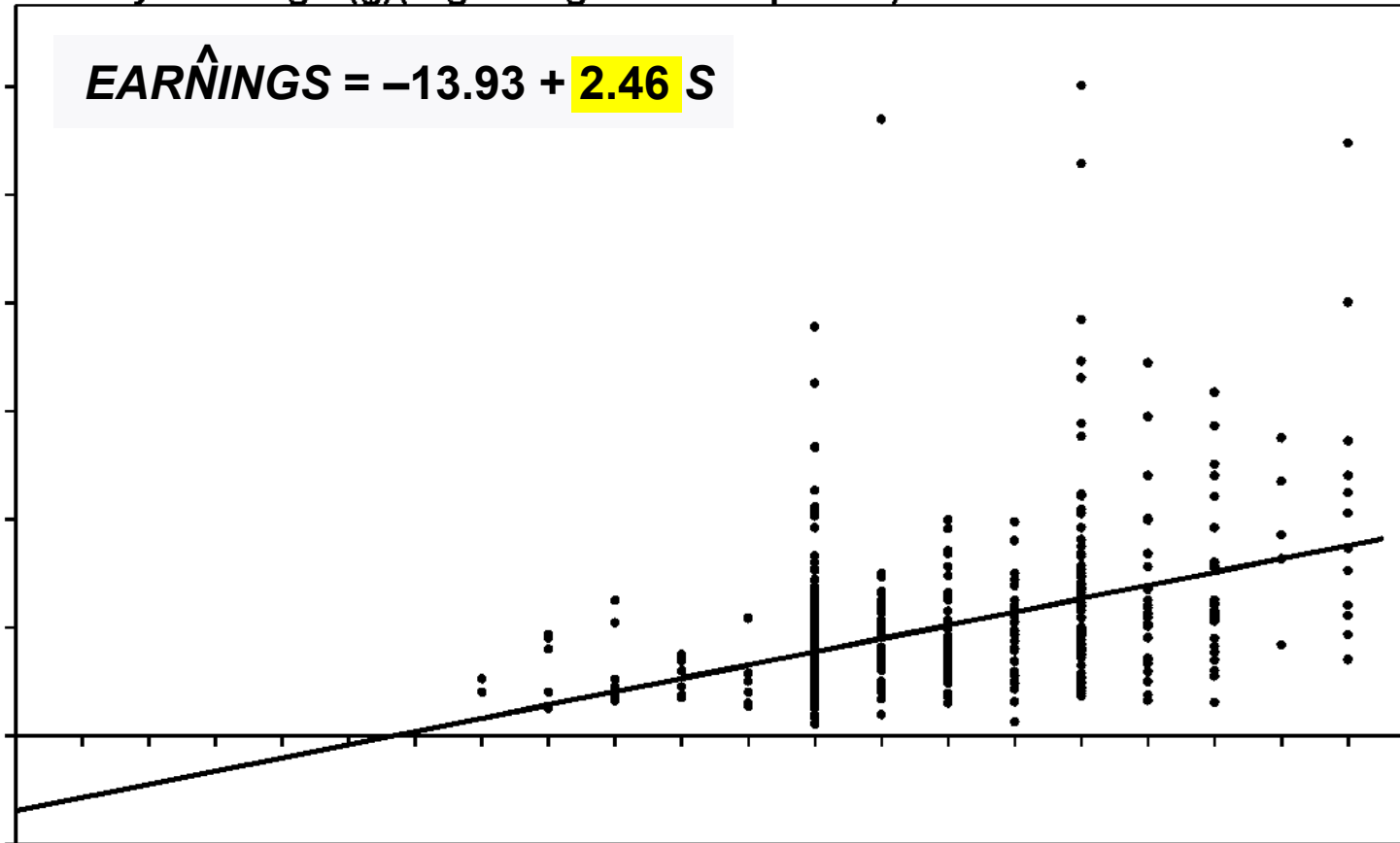
Линия регрессии указывает на то, что завершение 12-го класса вместо 11-го класса увеличит заработок на \$ 2,46, с \$ 13,07 до \$ 15,53.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)

$$\hat{EARNINGS} = -13.93 + 2.46 S$$

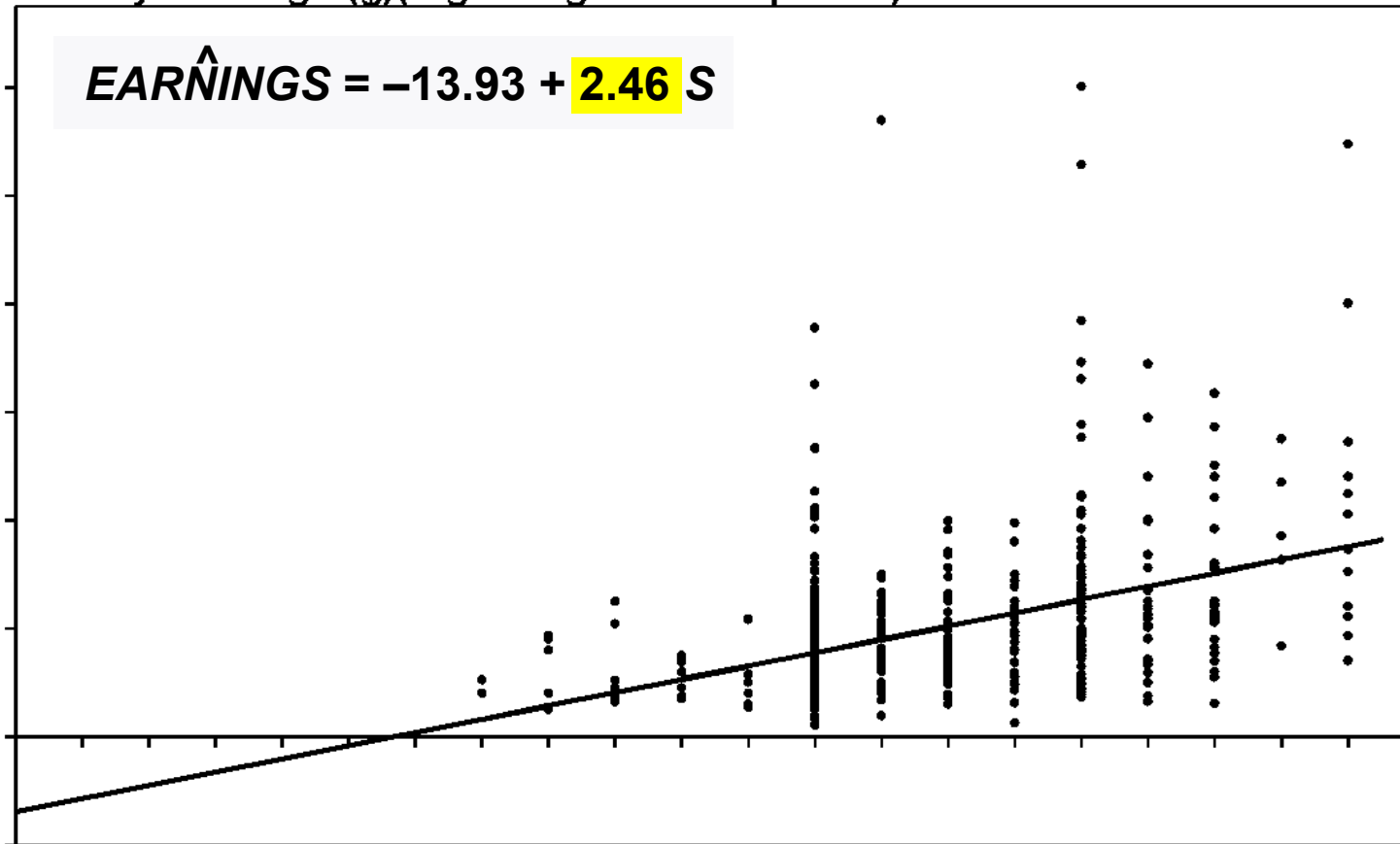


Является ли этот вывод правдоподобным? Если это неправдоподобно, то ваша модель некорректна.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)



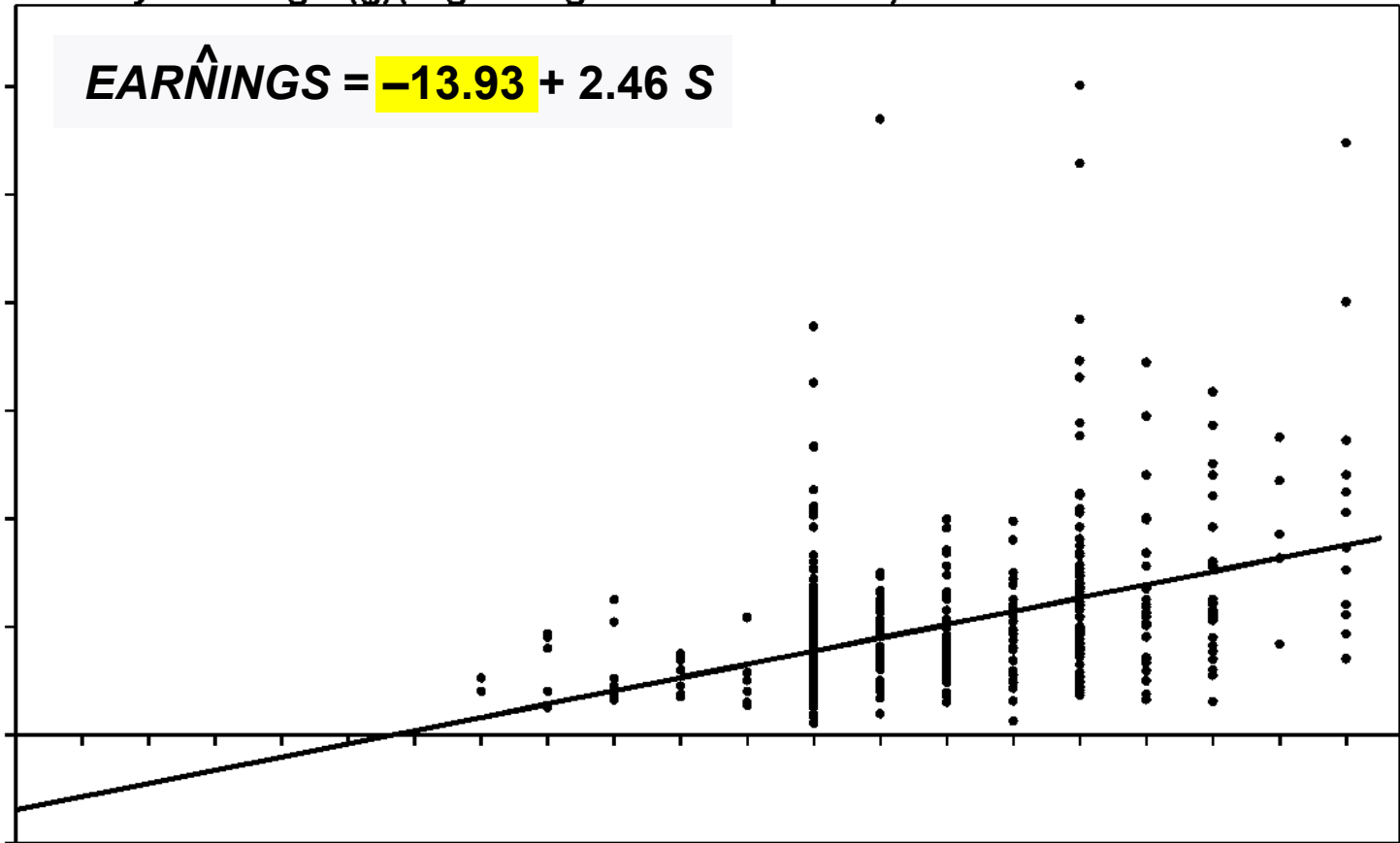
Для низких уровней образования это может быть правдоподобно. Но для высоких уровней это не так.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)

$$\hat{EARNINGS} = -13.93 + 2.46 S$$

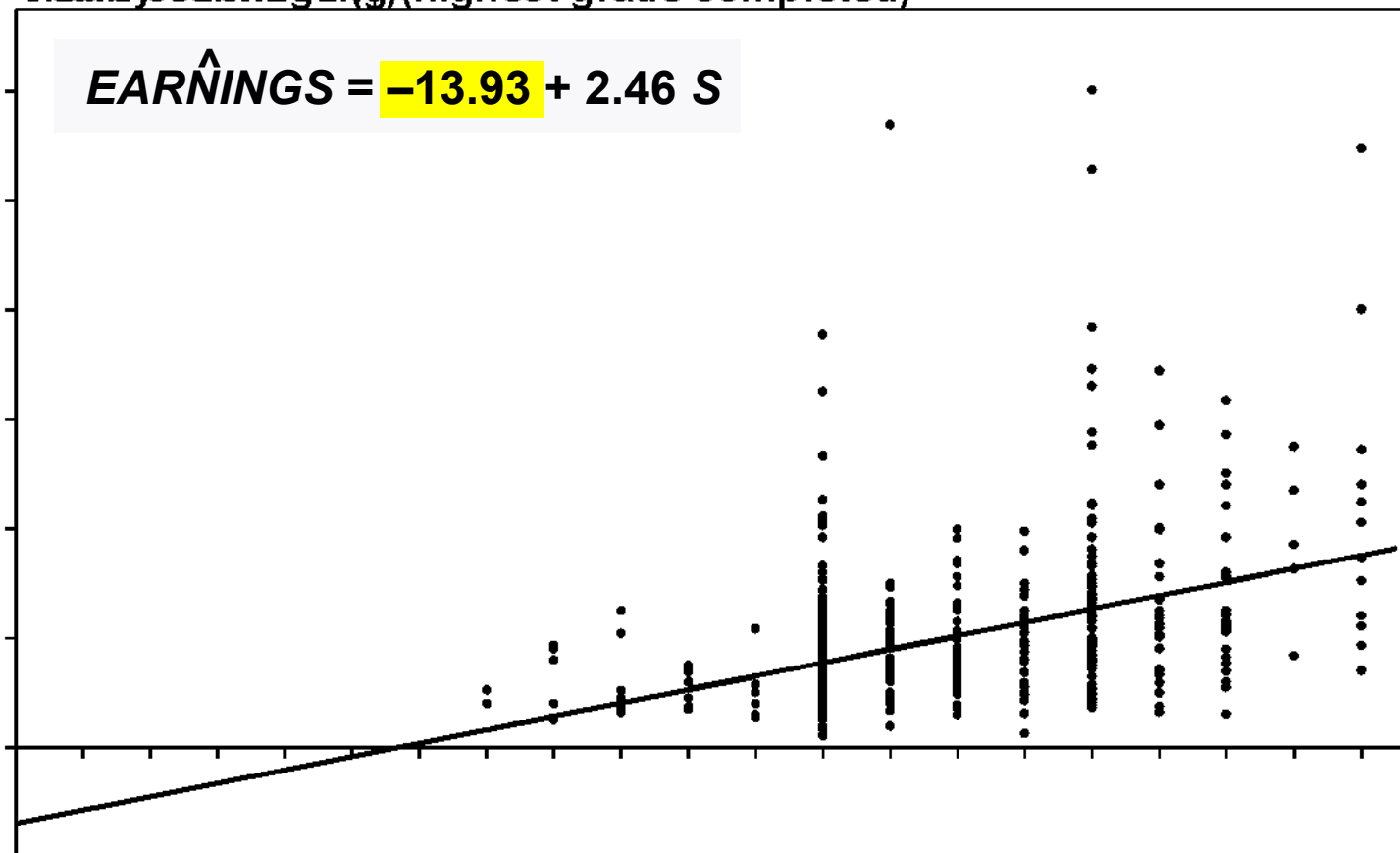


О чем нам говорит свободный член (константа)?

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)



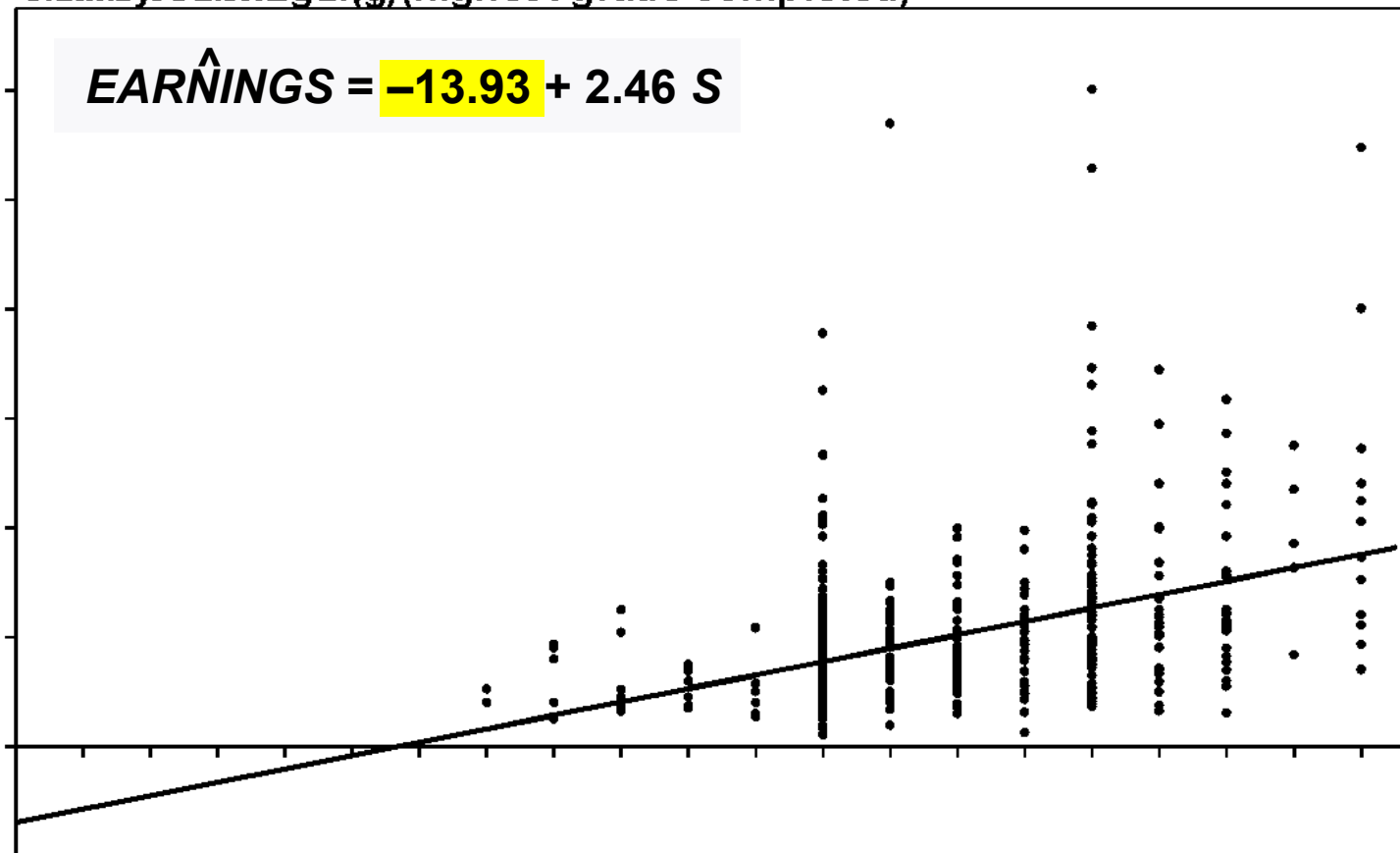
Константа указывает на то, что человек, не имеющий образования, должен будет заплатить 13,93 долл. в час, чтобы ему разрешили работать.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)

$$\hat{EARNINGS} = -13.93 + 2.46 S$$

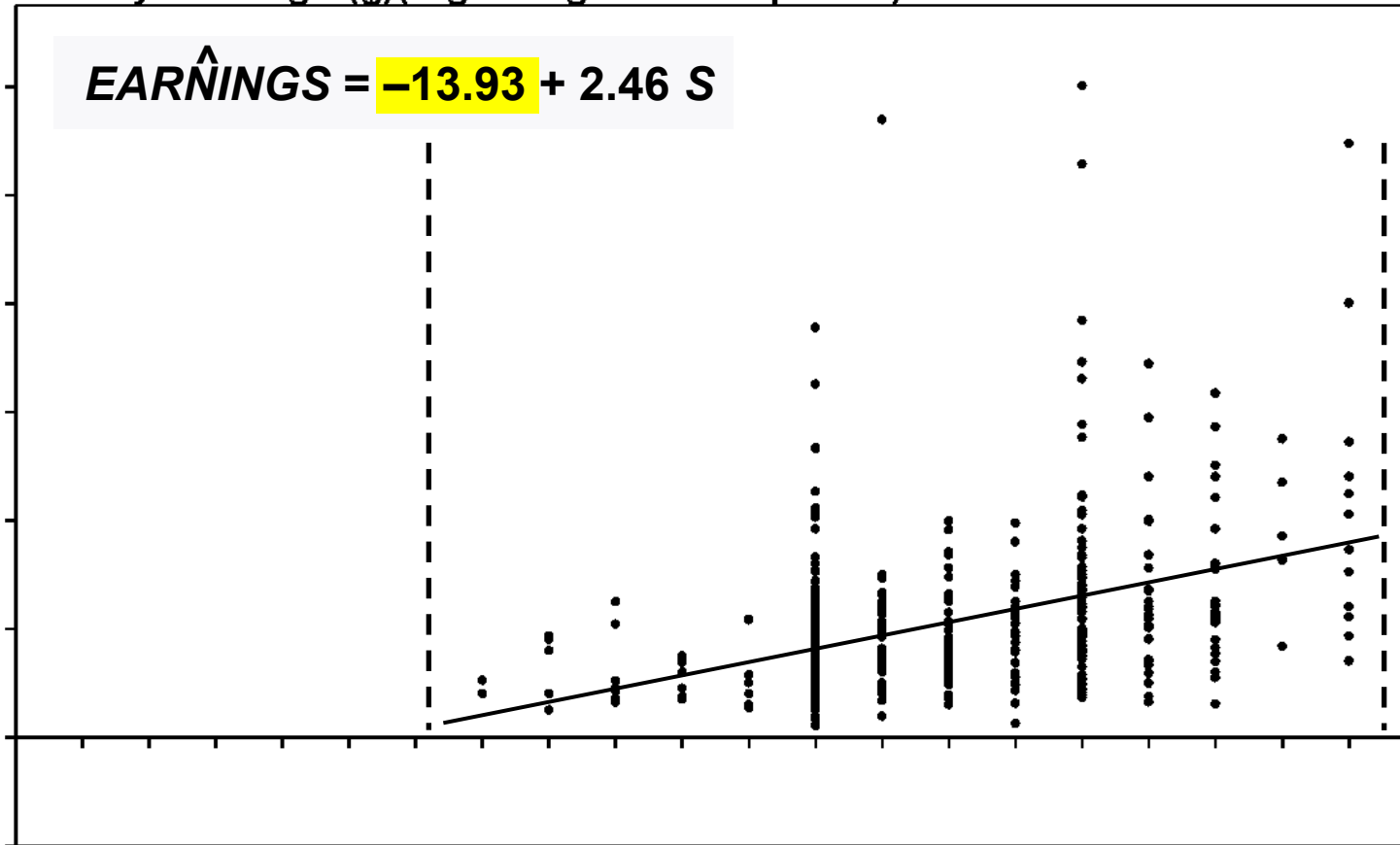


Это не имеет никакого смысла. В прежние времена ремесленники могли потребовать первоначальный взнос, когда принимали ученика, но интерпретацию отрицательного платежа невозможно представить.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling S (highest grade completed)

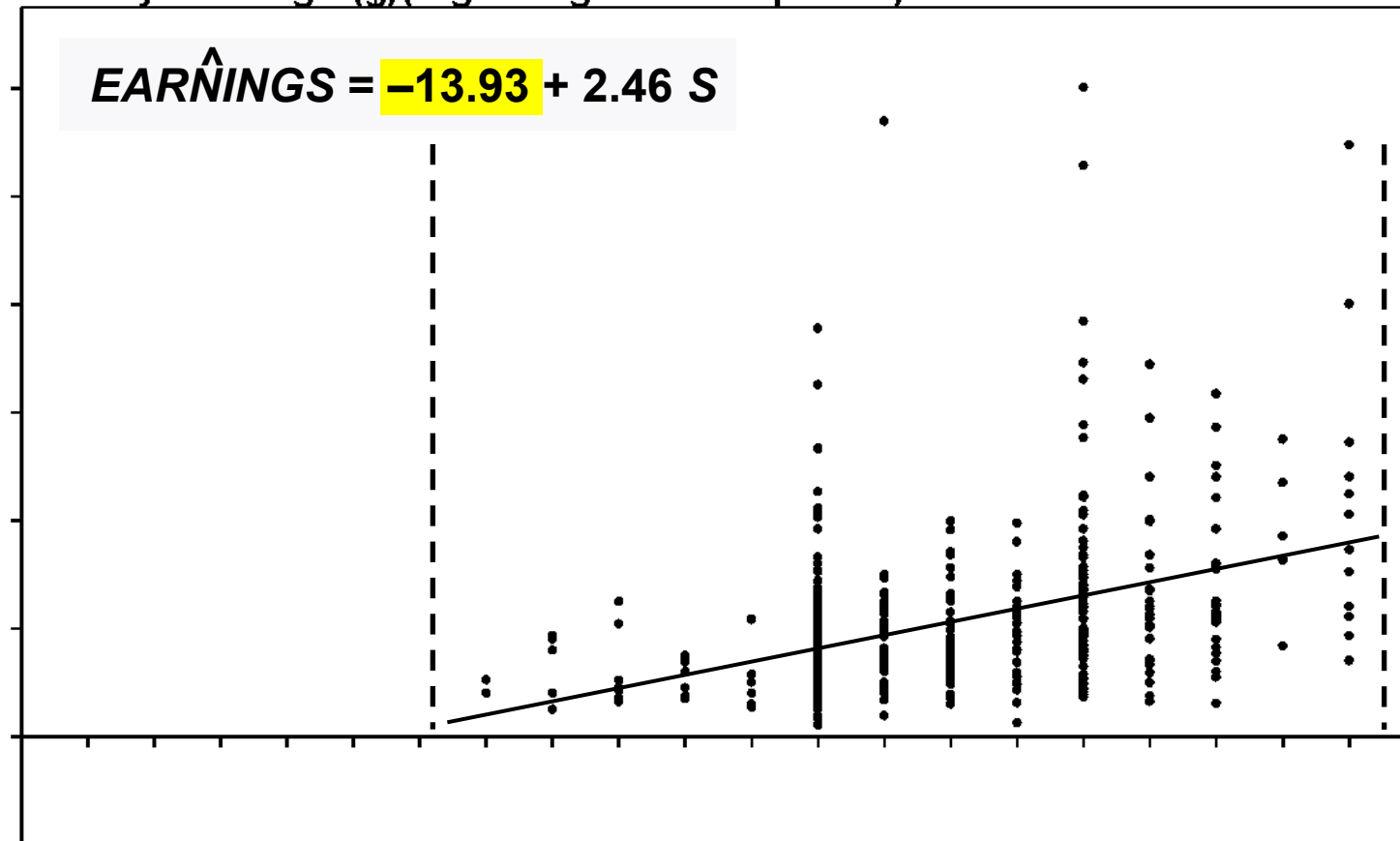


Решение проблемы заключается в том, чтобы ограничить интерпретацию диапазоном выборочных данных и отказаться от экстраполяции на том основании, что у нас нет данных вне диапазона данных.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

IID

Years of schooling S (highest grade completed)



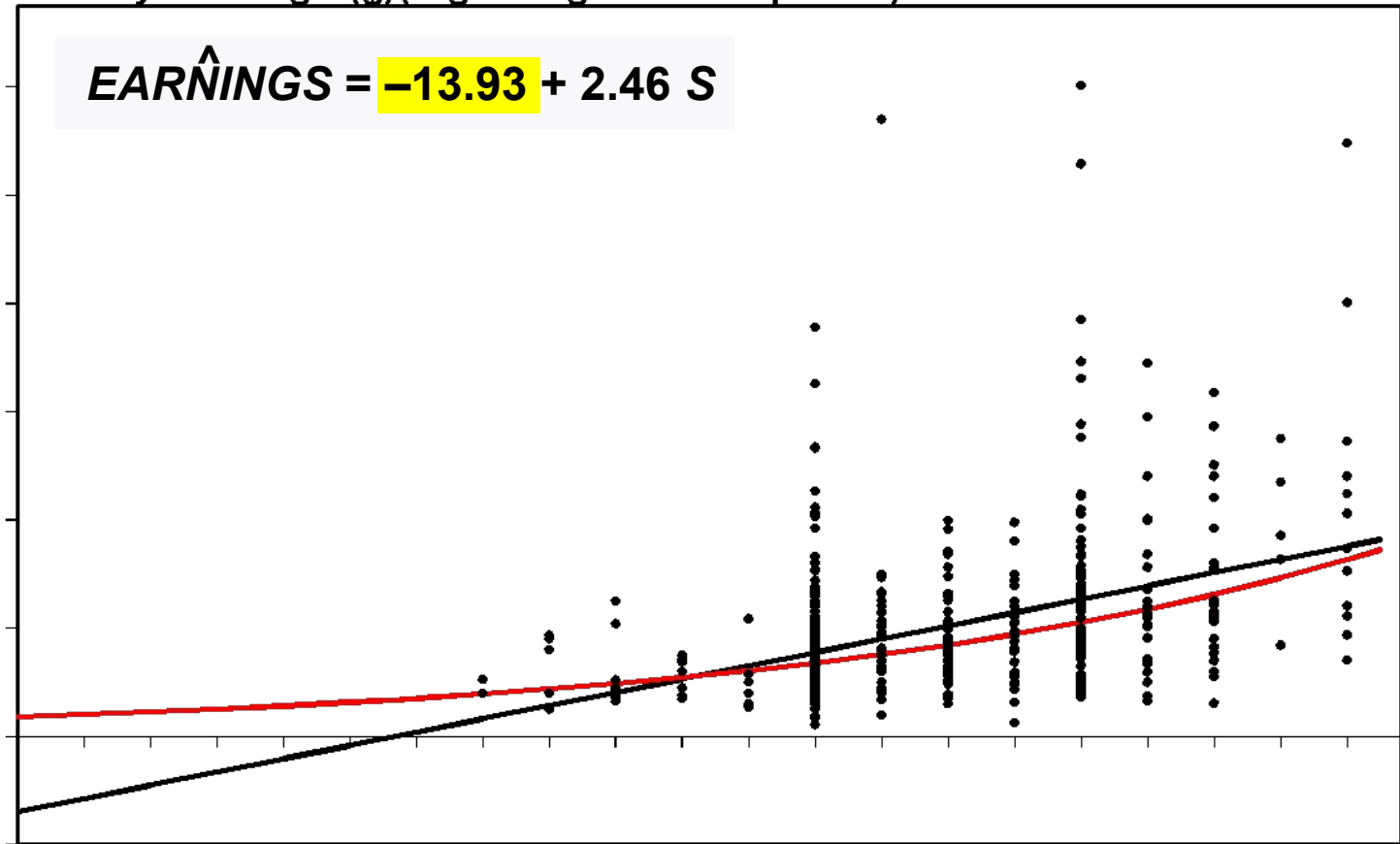
С помощью этого объяснения единственная функция постоянного члена состоит в том, чтобы вы могли нарисовать линию регрессии на правильной высоте на диаграмме рассеяния. Он не имеет никакого значения.

ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

110

Years of schooling (S) (highest grade completed)

$$EARNINGS = -13.93 + 2.46 S$$



Другое решение состоит в том, что истинное соотношение нелинейно и что мы изучаем его с помощью линейной регрессии. Далее мы перейдем к нелинейным моделям.