

Remember that in the real world, we never know or observe the true model

True population model: $wage = -12.72 + 2.27 * educ + u$

1. Repeating the same random sampling of 1000 observations gives different estimates, but if you were to average them up, you would find the true value, because $E(\hat{\beta}) = \beta$

2. Increasing sample size increases the precision of the estimate, because $\text{var}(\hat{\beta}_1) = \frac{\widehat{\text{var}(u)}}{n \text{var}(x)}$

. reg wage educ if _n<401;

Source	SS	df	MS	Number of obs = 400		
Model	7947.97607	1	7947.97607	F(1, 398)	= 69.64	
Residual	45425.1083	398	114.133438	Prob > F	= 0.0000	
Total	53373.0843	399	133.767129	R-squared	= 0.1489	
				Adj R-squared	= 0.1468	
				Root MSE	= 10.683	

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	2.166719	.2596455	8.34	0.000	1.656271	2.677168
_cons	-11.09661	3.579697	-3.10	0.002	-18.13409	-4.059135

. reg wage educ if _n<2001;

Source	SS	df	MS	Number of obs = 2000		
Model	41122.3613	1	41122.3613	F(1, 1998)	= 333.22	
Residual	246572.252	1998	123.409535	Prob > F	= 0.0000	
Total	287694.613	1999	143.919266	R-squared	= 0.1429	
				Adj R-squared	= 0.1425	
				Root MSE	= 11.109	

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	2.203853	.1207308	18.25	0.000	1.967082	2.440625
_cons	-11.93304	1.661577	-7.18	0.000	-15.19164	-8.674433

. reg wage educ if _n<4001;

Source	SS	df	MS	Number of obs = 4000		
Model	85546.3393	1	85546.3393	F(1, 3998)	= 728.91	
Residual	469213.909	3998	117.362158	Prob > F	= 0.0000	
Total	554760.248	3999	138.724743	R-squared	= 0.1542	
				Adj R-squared	= 0.1540	
				Root MSE	= 10.833	

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	2.221041	.0822659	27.00	0.000	2.059754	2.382328
_cons	-12.02081	1.133032	-10.61	0.000	-14.24218	-9.799433

Back to our sample of 2000 workers

```
. sum wage educ;
```

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	2000	13.633	2.0877	9	18
wage	2000	18.34701	11.49495	.7	82.42857

```
. reg wage educ;
```

Source	SS	df	MS	Number of obs =	2000
Model	41922.0349	1	41922.0349	F(1, 1998) =	376.94
Residual	222213.443	1998	111.217939	Prob > F =	0.0000
Total	264135.478	1999	132.133806	R-squared =	0.1587
				Adj R-squared =	0.1583
				Root MSE =	10.546

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	2.193546	.112983	19.41	0.000	1.971969 2.415123
_cons	-11.5576	1.558244	-7.42	0.000	-14.61355 -8.501649

Calculating $\widehat{\text{var}}(\hat{\beta}_1)$

Compute $\widehat{\text{var}}(u) = \frac{SSR}{n-2} = \frac{222213.44}{1998} = 111.22$

Compute $\text{var}(x) = (2.088)^2 = 4.358$

Compute $\widehat{\text{var}}(\hat{\beta}_1) = \frac{\widehat{\text{var}}(u)}{n \text{var}(x)} = \frac{111.22}{2000 * 4.358} = 0.01276$

$\text{st.err.}(\hat{\beta}_1) = \sqrt{0.01276} = 0.1130$