

## In-Class Exercise - Interpretation

Consider a dataset on births to women in the United States. Two variables of interest are infant birth weight in ounces (**bwght**), and the average number of cigarettes the mother smoked per day during pregnancy (**cigs**). The following simple regression was estimated using data on 1,388 births.

```
. regress bwght cigs
```

Source	SS	df	MS	Number of obs	=	1,388
Model	13060.4194	1	13060.4194	F(1, 1386)	=	32.24
Residual	561551.3	1,386	405.159668	Prob > F	=	0.0000
				R-squared	=	0.0227
				Adj R-squared	=	0.0220
Total	574611.72	1,387	414.283864	Root MSE	=	20.129

  

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cigs	-.5137721	.0904909	-5.68	0.000	-.6912861	-.3362581
_cons	119.7719	.5723407	209.27	0.000	118.6492	120.8946

These results can also be written in the following way:

$$\widehat{bwght} = 119.77 - 0.514cigs$$

1. What is the dependent variable? What is the independent variable?
2. Write, in words, what the interpretation of 0.514 is.
3. What is the predicted birth weight among mothers who do not smoke? What about when  $cigs = 20$  (one pack per day)? Comment on the difference.
4. Consider Prof. Beam, whose mother “cut back” to 10 cigarettes per day (it was the 80s) and was born weighing 9lb, 15 oz. What is her residual?
5. Find  $R^2$  in the raw regression output. What does it tell us?

6. Are any least squares assumptions likely to be violated? Explain.
  
7. Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.