Econometrics: Interpreting Regression Coefficients (Logs & Dummies)

Ryan Safner

Fall 2017

How we interpret the coefficients in regression models will depend on how the dependent (y) and independent (x) variables are measured. In general, there tend to be three types of variables used in econometrics: continuous variables, the natural log (ln) of continuous variables, and dummy variables. In the examples below, we will consider models with three different independent variables:

- X_{1i}: a continuous variable
- $ln(X_{2i})$: the natural log of a continuous variable
- X_{3i}: a dummy variable that equals 1 (if yes) or 0 (if no)

Below are three different OLS models. In each case, we keep the right hand side variables are the same, but as a demonstration, we change the dependent variable (Y) of interest to show the difference when we measure it as a continuous variable, the natural log of a continuous variable, or a dummy variable:

- Y_{1i}: a continuous variable
- $ln(Y_{2i})$: the natural log of a continuous variable
- Y_{3i}: a dummy variable that equals 1 (if yes) or 0 (if no)

Model 1

$$Y_{1i} = \beta_0 + \beta_1 X_{1i} + \beta_2 ln(X_{2i}) + \beta_3 X_{3i} + u_i \tag{1}$$

- $\beta_1 = \frac{\Delta Y_{1i}}{\Delta X_{1i}}$: a one unit change in X_1 causes a β_1 unit change in Y_{1i}
- $\beta_2 = \frac{\Delta Y_{1i}}{\Delta ln(X_{2i})}$: a 1% change in X_2 causes a $0.01 \times \beta_2$ unit change in Y_{1i}
- $\beta_3 = \frac{\Delta Y_{1i}}{\Delta X_{3i}}$: the change in X_3 from 0 to 1 causes a β_3 unit change in Y_{1i}

Model 2

$$ln(Y_{2i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 ln(X_{2i}) + \beta_3 X_{3i} + u_i \tag{2}$$

- $\beta_1 = \frac{\Delta ln(Y_{2i})}{\Delta X_{1i}}$: a one unit change in X_1 causes a $100 \times \beta_1$ percent change in Y_{2i}
- $\beta_2 = \frac{\Delta ln(Y_{2i})}{\Delta ln(X_{2i})}$: a 1% change in X_2 causes a β_2 percent change in Y_{2i}
- $\beta_3 = \frac{\Delta Y_{1i}}{\Delta X_{3i}}$: the change in X_3 from 0 to 1 causes a $100 \times \beta_3$ percent change in Y_{2i}

Model 3

$$Y_{3i} = \beta_0 + \beta_1 X_{1i} + \beta_2 ln(X_{2i}) + \beta_3 X_{3i} + u_i \tag{3}$$

- $\beta_1 = \frac{\Delta Y_{3i}}{\Delta X_{1i}}$: a one unit change in X_1 causes a $100 \times \beta_1$ percentage point change in the probability of Y_{3i} occurring (=1)
- $\beta_2 = \frac{\Delta Y_{3i}}{\Delta ln(X_{2i})}$: a 1% change in X_2 causes a β_2 percentage point change in the probability of Y_{3i} occurring (=1)
- $\beta_3 = \frac{\Delta Y_{3i}}{\Delta X_{3i}}$: the change in X_3 from 0 to 1 causes a $100 \times \beta_3$ percentage point change in the probability of Y_{3i} occurring (=1)

Example with Data

Below are the results from three regressions using the same data set. The results parallel the three general models outlined above. The dataset `meps2005.dta` can be found under Blackboard/Datasets, along with a `.do` file. It contains responses from a sample of senior citizens all on Medicare.

The regressions have three different outcome measures (analogous to Y_1, Y_2, and Y_3 above): total expenditures on medical care (totalexp, Y_1), the natural log of total expenditures on medical care (lntotalexp, Y_2), and whether or not the person has "goodhealth" (goodhealth, Y_3).

For each of these three dependent variables, we regress three potential independent variables, a continuous variable (age), the natural log of a continuous variable (ln of family income), and a dummy variable (obese=1 if a person is obese, =0 otherwise). The sample description and summary statistics are presented below:

```
. desc  totalexp ln_totalexp  goodhealth age ln_income obese

              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------
totalexp        long    %12.0g                total annual expenditures on health care
ln_totalexp     float   %9.0g                 log of total expenditures
goodhealth      float   %9.0g                 =1 if person reports excellent, very good, or good health
age             byte    %8.0g                 age in years
ln_income       float   %9.0g                 log of income
obese           float   %9.0g                 =1 if bmi>=30, =0 otherwise
```

```
. sum totalexp ln_totalexp  goodhealth age ln_income obese
```

Variable	Obs	Mean	Std. Dev.	Min	Max
totalexp	3167	8308.891	13999.03	1	235392
ln_totalexp	3167	7.992219	1.984316	0	12.36901
goodhealth	3167	.5866751	.4925079	0	1
age	3167	74.06157	6.278366	65	85
ln_income	3167	9.558831	.3464525	9.220389	9.913537
obese	3167	.2649195	.44136	0	1

Model 1

$$\widehat{Totalexp} = \hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 ln(income) + \hat{\beta}_3 obese$$

```
. reg totalexp age ln_income obese, r

Linear regression                               Number of obs =     3167
                                                F(  3,  3163) =    12.16
                                                Prob > F      =   0.0000
                                                R-squared     =   0.0085
                                                Root MSE      =    13946
```

totalexp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	194.0764	36.41287	5.33	0.000	122.6812	265.4716
ln_income	44.29933	746.3682	0.06	0.953	-1419.115	1507.714
obese	1393.604	517.8289	2.69	0.007	378.2893	2408.918
_cons	-6857.356	6860.138	-1.00	0.318	-20308.13	6593.415

$$\widehat{Totalexp} = -6857.36 + 194.08 age + 44.30 ln(income) + 1393.60 obese$$

Interpreting the coefficients:

- **age:** a one year increase in age will increase annual medical expenditures by \$194
- **ln_income**: a 1% increase in income will increase medical spending by $0.01 \times 44.2 =$ \$0.442
- **obese**: obese seniors spend \$1,393 more per year on medical care than non-obese seniors

Model 2

$$\widehat{ln(Totalexp)} = \hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 ln(income) + \hat{\beta}_3 obese$$

```
. reg ln_totalexp age ln_income obese, r

Linear regression                               Number of obs =     3167
                                                F(  3,  3163) =    26.68
                                                Prob > F      =   0.0000
                                                R-squared     =   0.0240
                                                Root MSE      =   1.9613
```

ln_totalexp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0437133	.0058519	7.47	0.000	.0322394	.0551872
ln_income	-.1600613	.1059448	-1.51	0.131	-.3677887	.0476662
obese	.4458879	.0751149	5.94	0.000	.298609	.5931667
_cons	6.166616	.9685833	6.37	0.000	4.267501	8.065731

$$\widehat{ln(Totalexp)} = 6.17 + 0.044 age - 0.16 ln(income) + 0.45 obese$$

Interpreting the coefficients:

- **age:** a one year increase in age will increase annual medical expenditures by 4.37%
- **ln_income**: a 1% increase in income will reduce medical spending by 0.16%
- **obese**: obese seniors spend 44.6% more per year on medical care than non-obese seniors

Model 3

$$\widehat{Goodhealth} = \hat{\beta}_0 + \hat{\beta}_1 age + \hat{\beta}_2 ln(income) + \hat{\beta}_3 obese$$

```
.  reg goodhealth age ln_income obese, r

Linear regression                               Number of obs =     3167
                                                F(  3,  3163) =    30.59
                                                Prob > F      =   0.0000
                                                R-squared     =   0.0264
                                                Root MSE      =   .48619

------------------------------------------------------------------------------
             |               Robust
  goodhealth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .002792   .0014438     1.93   0.053    -.0000389    .0056228
   ln_income |   .0791972   .0257435     3.08   0.002     .0287215    .1296728
       obese |   .1670099   .0189927     8.79   0.000     .1297707    .2042492
       _cons |  -.4213802    .243067    -1.73   0.083     -.897965    .0552047
------------------------------------------------------------------------------
```

$$\widehat{Goodhealth} = -0.421 + 0.003 age + 0.079 ln(income) + 0.167 obese$$

Interpreting the coefficients:

- **age:** a one year increase in age will increase the probability of reporting good health by 0.3 percentage points
- **ln_income**: a 1% increase in income will increase the probability of reporting good health by 0.079 percentage points
- **obese**: obese seniors have 16.7 higher percentage point probability of reporting good health than non-obese seniors