Interpretting Regression Coefficients

Ryan Safner

12/4/2018

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

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

- $\beta_1 = \frac{\Delta Y_{1i}}{\Delta X_{1i}}$: 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

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

- $\beta_1 = \frac{\Delta ln(Y_{2i})}{\Delta X_{1i}}$: 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

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

- $\beta_1 = \frac{\Delta Y_{3i}}{\Delta X_{1i}}$: 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. 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 (`ln.totalexp`, Y_2), and whether or not the person reports "`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.income`), and a dummy variable (`obese`$=$1 if a person is obese, = 0 otherwise). The sample description and summary statistics are presented below:

Variable	Obs	Min	Q1	Median	Q3	Max	Mean	Std. Dev.
age	3167	65.00	69.00	73.00	79.00	85.00	74.06	6.28
goodhealth	3167	0.00	0.00	1.00	1.00	1.00	0.59	0.49
ln.income	3167	9.22	9.22	9.22	9.91	9.91	9.56	0.35
ln.totalexp	3167	0.00	7.38	8.26	9.08	12.37	7.99	1.98
obese	3167	0.00	0.00	0.00	1.00	1.00	0.26	0.44
totalexp	3167	1.00	1596.00	3860.00	8793.50	235392.00	8308.89	13999.03

Model 1

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

```
##
## Call:
## lm(formula = totalexp ~ age + ln.income + obese, data = handout)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11246  -6388  -4159    427 228061
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6857.36    6951.23  -0.986   0.3240
## age           194.08      41.31   4.698 2.73e-06 ***
## ln.income      44.30     741.52   0.060   0.9524
## obese        1393.60     567.38   2.456   0.0141 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13950 on 3163 degrees of freedom
## Multiple R-squared:  0.00851,    Adjusted R-squared:  0.00757
## F-statistic: 9.049 on 3 and 3163 DF,  p-value: 5.784e-06
```

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

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 $0.01 \times 44.2 = \$0.442$
- `obses`: obese seniors spend \$1,393 more per year on medical care than non-obese seniors

Model 2

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

```
##
## Call:
## lm(formula = ln.totalexp ~ age + ln.income + obese, data = handout)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8086 -0.5943  0.2703  1.0835  4.5746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.166616   0.977608   6.308 3.22e-10 ***
## age          0.043713   0.005809   7.525 6.86e-14 ***
## ln.income   -0.160061   0.104286  -1.535    0.125
## obese        0.445888   0.079796   5.588 2.49e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.961 on 3163 degrees of freedom
## Multiple R-squared:  0.02396,    Adjusted R-squared:  0.02303
## F-statistic: 25.88 on 3 and 3163 DF,  p-value: < 2.2e-16
```

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

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

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

```
##
## Call:
## lm(formula = goodhealth ~ age + ln.income + obese, data = handout)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7681 -0.5322  0.2850  0.4380  0.5097
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.421380   0.242339  -1.739  0.08217 .
## age          0.002792   0.001440   1.939  0.05262 .
## ln.income    0.079197   0.025851   3.064  0.00221 **
## obese        0.167010   0.019780   8.443  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4862 on 3163 degrees of freedom
## Multiple R-squared:  0.0264, Adjusted R-squared:  0.02548
## F-statistic: 28.59 on 3 and 3163 DF,  p-value: < 2.2e-16
```

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

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