BETA VERSION – Send Comments to @adamrossnelson

Regression Interpretation

Functional Forms Cheat Sheet v 1.0

Prepared by Adam Ross Nelson JD PhD @adamrossnelson

All models are wrong, some are useful.[1]

	Example Notation(s)	Interpretations
Linear	$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome, β_0 is the Y intercept, $\beta_1 \ldots \beta_k$ are coefficients (slopes) on $X_1 \ldots X_k$ which are input variables, and u is the unexplained. $y = mx + b + e$ Where y is the outcome variable, m is the slope, x is the input variable, b is the y intercept, and e is the unexplained, unobservable, and error.	A one unit change in X_1 is associated with a B_1 change in Y when holding all other X_n (X_2, $X_3 \ldots X_k$) constant. This specification models **change at a steady rate**. Conversationally, ß terms are sometimes also called slopes. Also, B_1 will equal the change in Y over the change in X_1 when holding all other X_n (X_2, $X_3 \ldots X_k$) constant. Noted as: $\beta_1 = \frac{\Delta Y}{\Delta X_1}$
Log Linear	$\log(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome (logged), β_0 is the Y intercept, $\beta_1 \ldots \beta_k$ are coefficients on $X_1 \ldots X_k$ which are input variables, and u is the unexplained.	$\beta_1 \times 100$ is the percent change in Y that is associated with a one unit change in X_1, when holding all other X_n (X_2, $X_3 \ldots X_k$) constant. Because we interpret the outcome as a percent change this specification is **nonlinear**.
Linear Log	$Y_i = \beta_0 + \beta_1 \log(X_{1i}) + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome, B_0 is the Y intercept, β_1 is the coefficient of interest, $\beta_2 \ldots \beta_k$ are additional coefficients on $X_2 \ldots X_k$ which are input variables, and u is the unexplained.	$\beta_1 \times .01$ is the change in Y that is associated with a 1% change in X_1 when holding all other X_n (X_2, $X_3 \ldots X_k$) constant. Because we interpret the outcome as a percent change this specification is **nonlinear**.
Log Log	$\log(Y_i) = \beta_0 + \beta_1 \log(X_{1i}) + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome (logged), B_0 is the Y intercept, β_1 is the coefficient of interest, $\beta_2 \ldots \beta_k$ are additional, and u is the unexplained.	β_1 is the percent change in Y associated with a 1% change in X_1 when holding all other X_n (X_2, $X_3 \ldots X_k$) constant. Because we interpret both the outcome and inputs as a percent changes this specification is **nonlinear**.
Exponential Quadratic	$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome, B_0 is the Y intercept, β_1 & β_2 are coefficients of interest, $\beta_n \ldots \beta_k$ are additional coefficients, and u is the unexplained. Also where X_2 is the square of X_1.	To interpret the results, compute marginal effects. This specification models **change at changing rates**. First, compute the predicted value of Y for a particular value of X_1. (e.g. mean or median). Second, compute the predicted value for Y at $X_1 + 1$. Third, the *difference of predicted values* is the change in Y associated with a unit change in X_1. Also, the marginal effect of X on Y for a given value of X will be: $\beta_1 + 2(\beta_2)X$
Dummy Variables	$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome, B_0 is the Y intercept, β_1 & β_2 are coefficients of interest, $\beta_n \ldots \beta_k$ are additional coefficients, and u is the unexplained. Also where D_1 is a dummy variable only equal to 0 or 1.	This specification is often called a **multiple intercept & single slope** model because the coefficient (β_1) plus the Y intercept (β_0) will be the intercept on Y when the corresponding dummy variable (D_1) = 1. Interpreting this dummy variable's coefficient essentially means comparing the means of Y between two groups. The coefficient (β_1) is the difference in the mean of Y between the two groups when holding all other X_n (X_2, $X_3 \ldots X_k$) constant.
Dummy × Continuous	$Y_i = \beta_0 + \beta_1 X_{1i} D_{1i} + \beta_2 X_{1i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome, B_0 is the Y intercept, β_1 & β_2 are coefficients of interest, $\beta_k \ldots \beta_n$ are additional coefficients, and u is the unexplained. Also where D_1 is a dummy variable only equal to 0 or 1.	This specification is often called a **single intercept & multiple slope** model because the coefficient (β_1) on the interaction term ($X_1 D_1$) plus the corresponding slope (β_2) will be the slope on Y when the corresponding dummy variable (D_1) = 1. Interpreting a dummy variable's coefficient means showing whether the association of X and Y will be the same among two different groups. The coefficient (β_1) is the difference in the slope of Y between the two groups when holding all other X_n (X_2, $X_3 \ldots X_k$) constant.
Dummy + Dummy × Continuous	$Y_i = \beta_0 + \beta_1 X_{1i} D_{1i} + \beta_2 X_{1i} + \beta_3 D_{1i} + \ldots + \beta_k X_{ki} + u$ Where Y is the outcome, B_0 is the Y intercept, β_1, β_2, & β_3 are the coefficients of interest, $\beta_n \ldots \beta_k$ are additional coefficients, and u is the unexplained. Also where D_1 is a dummy variable only equal to 0 or 1.	This specification is often called a **multiple intercept & multiple slope** model because the coefficient (β_1) on the interaction term ($X_1 D_1$) plus the corresponding slope (β_2) will be the slope on Y when the corresponding dummy variable (D_1) = 1. Likewise, the coefficient (β_3) plus the Y intercept (β_0) will be the intercept on Y when the corresponding dummy variable (D_1) = 1 Interpreting a dummy variable's coefficient means showing whether the association of X and Y will be the same among two different groups.

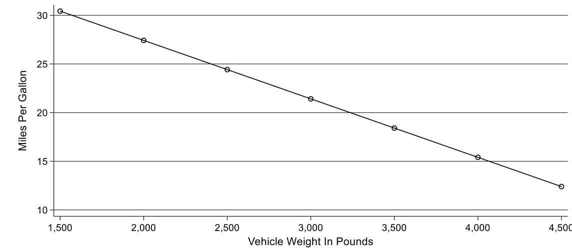
BETA VERSION – Send Comments to @adamrossnelson

Selected Visualizations

These visualizations model vehicle efficiency (miles per gallon) as a function of vehicle weight.[2]

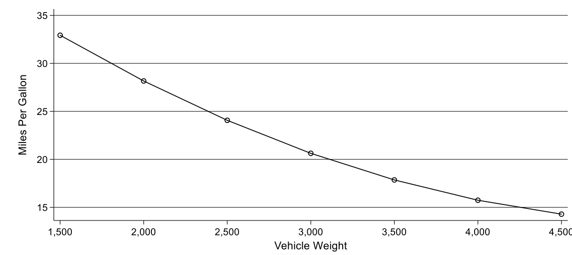
Linear Regression – Change At A Steady Rate

$$mpg = 39.44 + -.01 \times vehicle\ weight$$



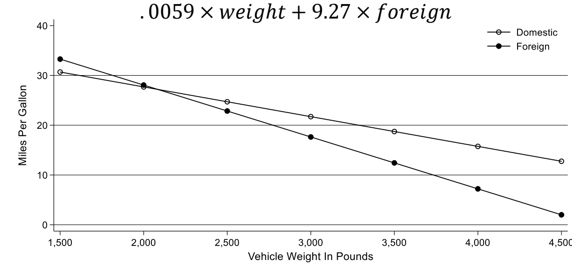
Exponential Quadratic – Change At A Changing Rate

$$mpg = 51.18 - .014 * weight + .00000013 \times weight^2$$



Dummy + Dummy × Continuous – Multiple Slope & Multiple Intercept

$$mpg = 39.65 - .0044 \times weight \times foreign - .0059 \times weight + 9.27 \times foreign$$



[1] Attributed to George Box. Box, G. E. P.(1976),". Science and Statistics," *Journal of the American Statistical Association*, 71, 791-799. See also: https://en.wikipedia.org/wiki/All_models_are_wrong

[2] Using data from Stata available at http://www.stata-press.com/data/r15/auto2.dta

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Adam Ross Nelson