**Reference Categories in Stata**

**Created by Nikki Tennessen for PSQF 6270**

// Factor Variable Example in Stata – using i. and baselevels

// Predicting Math Self-Assessment using Self-Identified Race

summarize mselfas // provides a numerical summary for math self-assessment

 Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

 mselfas | 14,047 .004475 .935765 -1.87 1.18

tabulate race // provides a frequency table for race/ethnicity categories reports value labels

Race/Ethnic |

 ity of |

 Student | Freq. Percent Cum.

------------+-----------------------------------

 white | 10,395 74.00 74.00

 Asian | 904 6.44 80.44

 Latinae | 1,526 10.86 91.30

 Black | 1,222 8.70 100.00

------------+-----------------------------------

 Total | 14,047 100.00

tabulate race, nolabel // provides a frequency table showing numerical values representing race/ethnicity categories; “white” is listed as the first category

Race/Ethnic |

 ity of |

 Student | Freq. Percent Cum.

------------+-----------------------------------

 1 | 10,395 74.00 74.00

 2 | 904 6.44 80.44

 3 | 1,526 10.86 91.30

 4 | 1,222 8.70 100.00

------------+-----------------------------------

 Total | 14,047 100.00

regress mselfas asian latinae black // the "white" dummy variable is omitted to serve as the reference group

 Source | SS df MS Number of obs = 14,047

-------------+---------------------------------- F(3, 14043) = 28.41

 Model | 74.1948087 3 24.7316029 Prob > F = 0.0000

 Residual | 12225.2709 14,043 .870559772 R-squared = 0.0060

-------------+---------------------------------- Adj R-squared = 0.0058

 Total | 12299.4657 14,046 .875656107 Root MSE = .93304

------------------------------------------------------------------------------

 mselfas | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

 asian | .2444849 .0323536 7.56 0.000 .1810675 .3079023

 latinx | -.1148812 .025578 -4.49 0.000 -.1650175 -.064745

 black | .0163823 .0282162 0.58 0.562 -.0389252 .0716898

 \_cons | -.0002039 .0091514 -0.02 0.982 -.0181419 .017734

------------------------------------------------------------------------------

regress mselfas i.race // i. treats the first category as the reference group; in this case, "white" serves as the reference category since it's the first value of the race variable

 Source | SS df MS Number of obs = 14,047

-------------+---------------------------------- F(3, 14043) = 28.41

 Model | 74.1948087 3 24.7316029 Prob > F = 0.0000

 Residual | 12225.2709 14,043 .870559772 R-squared = 0.0060

-------------+---------------------------------- Adj R-squared = 0.0058

 Total | 12299.4657 14,046 .875656107 Root MSE = .93304

------------------------------------------------------------------------------

 mselfas | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

 race |

 Asian | .2444849 .0323536 7.56 0.000 .1810675 .3079023

 Latinae | -.1148812 .025578 -4.49 0.000 -.1650175 -.064745

 Black | .0163823 .0282162 0.58 0.562 -.0389252 .0716898

 |

 \_cons | -.0002039 .0091514 -0.02 0.982 -.0181419 .017734

------------------------------------------------------------------------------

// as long as you have the correct reference group between dummy variables and the i. variable, the regression equations will be the same

// you can change the reference category using "ib" in place of "i"

regress mselfas ib2.race // changes reference group to 2; in this case, "Asian"

 Source | SS df MS Number of obs = 14,047

-------------+---------------------------------- F(3, 14043) = 28.41

 Model | 74.1948087 3 24.7316029 Prob > F = 0.0000

 Residual | 12225.2709 14,043 .870559772 R-squared = 0.0060

-------------+---------------------------------- Adj R-squared = 0.0058

 Total | 12299.4657 14,046 .875656107 Root MSE = .93304

------------------------------------------------------------------------------

 mselfas | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

 race |

 white | -.2444849 .0323536 -7.56 0.000 -.3079023 -.1810675

 Latinae | -.3593662 .0391598 -9.18 0.000 -.4361247 -.2826077

 Black | -.2281026 .0409318 -5.57 0.000 -.3083344 -.1478708

 |

 \_cons | .244281 .0310324 7.87 0.000 .1834534 .3051086

------------------------------------------------------------------------------

regress mselfas ib(last).race // changes reference group to the last category; in this case, "Black"

 Source | SS df MS Number of obs = 14,047

-------------+---------------------------------- F(3, 14043) = 28.41

 Model | 74.1948087 3 24.7316029 Prob > F = 0.0000

 Residual | 12225.2709 14,043 .870559772 R-squared = 0.0060

-------------+---------------------------------- Adj R-squared = 0.0058

 Total | 12299.4657 14,046 .875656107 Root MSE = .93304

------------------------------------------------------------------------------

 mselfas | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

 race |

 white | -.0163823 .0282162 -0.58 0.562 -.0716898 .0389252

 Asian | .2281026 .0409318 5.57 0.000 .1478708 .3083344

 Latinae | -.1312636 .0358175 -3.66 0.000 -.2014706 -.0610566

 |

 \_cons | .0161784 .0266909 0.61 0.544 -.0361394 .0684962

------------------------------------------------------------------------------

regress mselfas i.race, baselevels // you can include the option “baselevels” to explicitly list which category serves as the baselevel

 Source | SS df MS Number of obs = 14,047

-------------+---------------------------------- F(3, 14043) = 28.41

 Model | 74.1948087 3 24.7316029 Prob > F = 0.0000

 Residual | 12225.2709 14,043 .870559772 R-squared = 0.0060

-------------+---------------------------------- Adj R-squared = 0.0058

 Total | 12299.4657 14,046 .875656107 Root MSE = .93304

------------------------------------------------------------------------------

 mselfas | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 race |

 white | 0 (base)

 Asian | .2444849 .0323536 7.56 0.000 .1810675 .3079023

 Latinae | -.1148812 .025578 -4.49 0.000 -.1650175 -.064745

 Black | .0163823 .0282162 0.58 0.562 -.0389252 .0716898

 |

 \_cons | -.0002039 .0091514 -0.02 0.982 -.0181419 .017734

------------------------------------------------------------------------------