

Creating dummy variables with xi

with associated Stata code and output

xi expands terms containing categorical variables into indicator (also called dummy) variable sets by creating new variables, and executes the specified command with the expanded terms. By default, xi make the first value of the variable the referent group(how to specify otherwise is discussed below). The new dummy variables are left in your dataset. The names of the new dummy variables start with _I. xi may be used with any Stata command.

Using xi

In order to use the interaction expansion, preface the command you want to use it on with "xi:" and preface any variable needing expansion with "i."

- **xi: reg alc93 i.educ3**

```
. xi: reg alc93 i.educ3
i.educ3          _Ieduc3_1-3          (naturally coded; _Ieduc3_1 omitted)

-----+-----
Source |           SS       df       MS              Number of obs =   1364
-----+-----+-----+-----
Model  |  3.43716241         2   1.71858121          F( 2, 1361) =    0.72
Residual | 3239.97716   1361   2.38058572          Prob > F      =   0.4860
-----+-----+-----+-----
Total  | 3243.41433   1363   2.37961433          R-squared     =   0.0011
                                           Adj R-squared = -0.0004
                                           Root MSE    =   1.5429

-----+-----
alc93 |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----
_Ieduc3_2 |   .0763008   .1122574     0.68   0.497   -0.1439156   0.2965171
_Ieduc3_3 |  -0.0408742   .132318    -0.31   0.757   -0.3004435   0.2186951
 _cons |   7.925572   .0983726   80.57   0.000    7.732594    8.118551
```

Clarifying output

The default variables names are a bit hard to recognize. To fix this use the reformat command immediately following your regression. It provides clearer output.

- **reformat**

REGRESS formatted output

Outcome variable: alc93 , n=1364

```
-----+-----+-----+-----+-----
Covariate                Coef.   Std. Err.      P>|t|     95% Conf. Interval
-----+-----+-----+-----+-----
Education - 3 Cats
  <12yr*                   0
  12-15yr                  0.076     0.112     0.497   (-0.144 to 0.297)
  4yr College              -0.041     0.132     0.757   (-0.300 to 0.219)
Constant                  7.926     0.098    <0.001   (7.733 to 8.119)
-----+-----+-----+-----+-----
```

* Baseline category

Setting the referent group

By default it has made <12yr the referent group. If we would like to make 12-15 yr the referent group instead we would issue the command

- `char educ3 [omit] 2`

Rerunning the regression we see that this made 12-15yr the referent group.

```
. xi: reg alc93 i.educ3
i.educ3          _Ieduc3_1-3          (naturally coded; _Ieduc3_2 omitted)

-----+-----
Source |           SS          df           MS          Number of obs =      1364
-----+-----+-----+-----+-----
Model |   3.43716241           2   1.71858121          F( 2, 1361) =      0.72
Residual | 3239.97716   1361   2.38058572          Prob > F      =  0.4860
-----+-----+-----+-----+-----
Total | 3243.41433   1363   2.37961433          R-squared      =  0.0011
                                           Adj R-squared  = -0.0004
                                           Root MSE      =  1.5429

-----+-----
alc93 |           Coef.      Std. Err.      t      P>|t|      [95% Conf. Interval]
-----+-----+-----+-----+-----+-----
_Ieduc3_1 |  -.0763008      .1122574      -0.68   0.497   - .2965171      .1439156
_Ieduc3_3 |  -.117175      .1037084     -1.13   0.259   - .3206206      .0862706
 _cons |   8.001873      .0540791     147.97  0.000   7.895786      8.107961
-----+-----
```

Using xi with globals

If you are in the habit of creating globals you will need to create additional globals with the variables prefaced by "i." and use the second global in any commands with the "xi:" function. So for example if you have a global for demographics:

```
global demog "female age race educ3"
```

you'll need to make

```
global demogxi "female age i.race i.educ3"
```