Investigating the effects of factor variables

Jeff Pitblado
Associate Director, Statistical Software
StataCorp LP

German Stata User's Group 2011

Outline

This workshop's materials have been posted to the following website for your convenience.

http://www.stata.com/users/jpitblado/2011germany/

1 What are effects?

Effects

The effect of a factor variable is the change in a measurement between two or more levels of the factor.

Example:

- Difference in average cholesterol measurement between two age groups in a population.

2 Computing effects

```
. webuse cholesterol
(Artificial cholesterol data)
. describe chol agegrp
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
chol            float  %9.0g                  cholesterol level (mg/dL)
agegrp          float  %9.0g       ages
. label list ages
ages:
           1 10-19
           2 20-29
           3 30-39
           4 40-59
           5 60-79
```

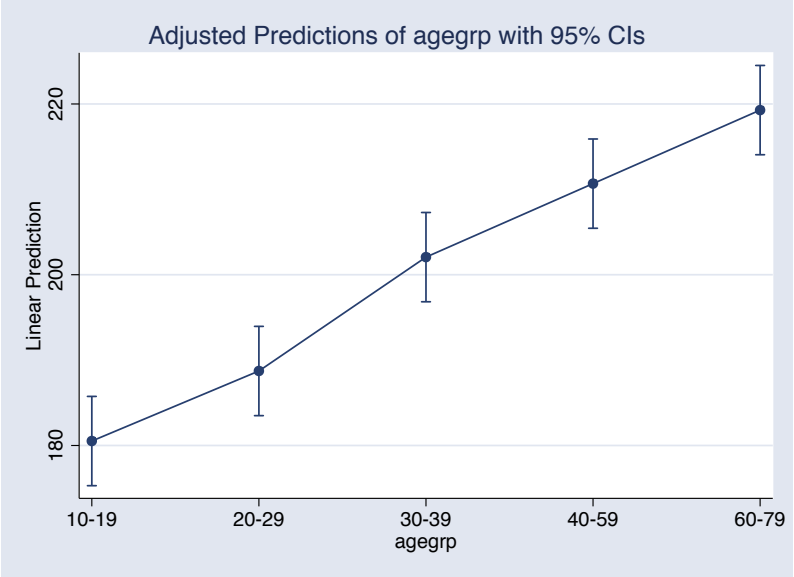
```
. regress chol i.agegrp
      Source |       SS       df       MS              Number of obs =      75
-------------+------------------------------           F(  4,    70) =   35.02
       Model |  14943.3997     4  3735.84993            Prob > F      =  0.0000
    Residual |  7468.21971    70  106.688853            R-squared     =  0.6668
-------------+------------------------------           Adj R-squared =  0.6477
       Total |  22411.6194    74  302.859722            Root MSE      =  10.329

------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agegrp |
          2  |   8.203575   3.771628     2.18   0.033     .6812991    15.72585
          3  |   21.54105   3.771628     5.71   0.000     14.01878    29.06333
          4  |   30.15067   3.771628     7.99   0.000      22.6284    37.67295
          5  |   38.76221   3.771628    10.28   0.000     31.23993    46.28448
             |
       _cons |   180.5198   2.666944    67.69   0.000     175.2007    185.8388
------------------------------------------------------------------------------
```

```
. margins agegrp
Adjusted predictions                              Number of obs   =         75
Model VCE    : OLS
Expression   : Linear prediction, predict()
------------------------------------------------------------------------------
             |            Delta-method
```

	Margin	Std. Err.	z	P>|z|	[95% Conf.	Interval]
agegrp						
1	180.5198	2.666944	67.69	0.000	175.2926	185.7469
2	188.7233	2.666944	70.76	0.000	183.4962	193.9504
3	202.0608	2.666944	75.76	0.000	196.8337	207.2879
4	210.6704	2.666944	78.99	0.000	205.4433	215.8975
5	219.282	2.666944	82.22	0.000	214.0548	224.5091

```
. marginsplot
```



Coefficient table

`regress` reports some simple tests on the effects of **`agegrp`** on **`chol`**.

How can we change the base level?

- Refit the model using the **`b.`** operator.
- Use **`test`** or **`lincom`** to perform the comparison.

```
. lincom 1.agegrp - 5.agegrp
 ( 1)  1b.agegrp - 5.agegrp = 0
```

chol	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-38.76221	3.771628	-10.28	0.000	-46.28448	-31.23993

```
. lincom 2.agegrp - 5.agegrp
 ( 1)  2.agegrp - 5.agegrp = 0
```

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

(1)	-30.55863	3.771628	-8.10	0.000	-38.08091	-23.03636

- New in Stata 12
- ANOVA-style tests of linear hypotheses involving factor variables and their interactions from the most recently fit model.
  - main effects
  - simple effects
  - interaction effects
  - nested effects
- Decompose tests into individual components/effects/constrasts.
  - built-in contrast operators
  - user defined contrasts

Syntax

contrast *op*.*varname* [, *options*]

op.	Description
r.	diff from a reference (base) level; the default
a.	diff from the next level (adjacent)
ar.	diff from the previous level (reverse adjacent)

```
. contrast rb5.agegrp, effects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
agegrp			
(1 vs 5)	1	105.62	0.0000
(2 vs 5)	1	65.65	0.0000
(3 vs 5)	1	20.85	0.0000
(4 vs 5)	1	5.21	0.0255
Joint	4	35.02	0.0000
Residual	70		

	Contrast	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegrp						
(1 vs 5)	-38.76221	3.771628	-10.28	0.000	-46.28448	-31.23993
(2 vs 5)	-30.55863	3.771628	-8.10	0.000	-38.08091	-23.03636
(3 vs 5)	-17.22115	3.771628	-4.57	0.000	-24.74343	-9.698877
(4 vs 5)	-8.611533	3.771628	-2.28	0.025	-16.13381	-1.089257

```
. contrast a.agegrp, effects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
agegrp			
(1 vs 2)	1	4.73	0.0330
(2 vs 3)	1	12.51	0.0007
(3 vs 4)	1	5.21	0.0255
(4 vs 5)	1	5.21	0.0255
Joint	4	35.02	0.0000
Residual	70		

	Contrast	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegrp						
(1 vs 2)	-8.203575	3.771628	-2.18	0.033	-15.72585	-.6812991
(2 vs 3)	-13.33748	3.771628	-3.54	0.001	-20.85976	-5.815204
(3 vs 4)	-8.60962	3.771628	-2.28	0.025	-16.1319	-1.087345
(4 vs 5)	-8.611533	3.771628	-2.28	0.025	-16.13381	-1.089257

As-balanced effects

Compute effects using linear combinations that weight each margin equally.

op.	Description
g.	diff from the balanced grand mean
h.	diff from the balanced mean of subsequent levels (Helmert)
j.	diff from the balanced mean of previous levels (reverse Helmert)
p.	orthogonal polynomial in the level values
q.	orthogonal polynomial in the level squence

```
. contrast h.agegrp, effects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
agegrp			
(1 vs >1)	1	68.42	0.0000
(2 vs >2)	1	50.79	0.0000
(3 vs >3)	1	15.63	0.0002
(4 vs 5)	1	5.21	0.0255
Joint	4	35.02	0.0000
Residual	70		

	Contrast	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegrp						
(1 vs >1)	-24.66438	2.981734	-8.27	0.000	-30.61126	-18.7175
(2 vs >2)	-21.94774	3.079522	-7.13	0.000	-28.08965	-15.80583
(3 vs >3)	-12.91539	3.266326	-3.95	0.000	-19.42987	-6.400905
(4 vs 5)	-8.611533	3.771628	-2.28	0.025	-16.13381	-1.089257

As-observed effects

Compute effects using linear combinations that weight each margin according to the sample frequencies of the levels.

op.	Description
gw.	diff from the weighted grand mean
hw.	diff from the weighted mean of subsequent levels (Helmert)
jw.	diff from the weighted mean of previous levels (reverse Helmert)
pw.	weighted orthogonal polynomial in the level values
qw.	weighted orthogonal polynomial in the level squence

3 Higher order effects

Interaction effect

When the effect of one factor depends on the level of other factors.

Example:

- Difference in average blood pressure measurement between two dosage levels for men and women.
- Factors: dosage and gender

```
. webuse bpchange
(Artificial blood pressure data)
. describe
Contains data from http://localpress.stata.com/data/r12/bpchange.dta
  obs:            30                          Artificial blood pressure data
 vars:             3                          21 Feb 2011 16:59
 size:           360
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
bpchange        float  %9.0g                  change in diastolic blood
                                                pressure
dose            float  %9.0g                  dosage in milligrams per day
gender          float  %9.0g       gender
-------------------------------------------------------------------------------
Sorted by:


. anova bpchange dose##gender
                          Number of obs =      30     R-squared     =  0.9647
                          Root MSE      =  1.4677     Adj R-squared =  0.9573
                 Source | Partial SS    df       MS           F     Prob > F
------------------------+----------------------------------------------------
                  Model |  1411.9087     5  282.381741     131.09     0.0000
```

dose	963.481795	2	481.740897	223.64	0.0000
gender	355.118817	1	355.118817	164.85	0.0000
dose#gender	93.3080926	2	46.6540463	21.66	0.0000
Residual	51.699253	24	2.15413554		
Total	1463.60796	29	50.4692399		

```
. contrast dose#gender
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
dose#gender	2	21.66	0.0000
Residual	24		

```
. contrast dose##gender
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
dose	2	223.64	0.0000
gender	1	164.85	0.0000
dose#gender	2	21.66	0.0000
Residual	24		

```
. contrast a.dose@gender, effects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
dose@gender			
(250 vs 500) 1	1	47.24	0.0000
(250 vs 500) 2	1	122.90	0.0000
(500 vs 750) 1	1	11.06	0.0028
(500 vs 750) 2	1	70.68	0.0000
Joint	4	122.65	0.0000
Residual	24		

	Contrast	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dose@gender						
(250 vs 500) 1	6.380018	.9282533	6.87	0.000	4.464198	8.295839
(250 vs 500) 2	10.29073	.9282533	11.09	0.000	8.374914	12.20655
(500 vs 750) 1	3.087217	.9282533	3.33	0.003	1.171396	5.003038
(500 vs 750) 2	7.803784	.9282533	8.41	0.000	5.887963	9.719605

- Factor effects on slopes
  - **contrast** *fvar***#c.***xvar*
- Nonlinear models
  - **clogit**, **glm**, **logit**, **heckman**, **ivregress**, **nbreg**, **poisson**, ...
- Multiple equations
  - **manova**, **mlogit**, **mprobit**, **mvreg**
  - Special **_eqns** factor for effects between equations
- Adjusting for multiple comparisons
  - Bonferroni
  - Šidák
- [R] **contrast** — over 50 pages of informaton

4 Pairwise comparisons

Syntax

pwcompare *marginlist* [, *options*]

- Intercept and slope effects
- Nonlinear models
- Multiple equations
- Adjusting for multiple comparisons
  - Generally applicable
    - Bonferroni, Scheffe, Šidák
  - Linear models only
    - Tukey, Student-Newman-Keuls, Duncan, Dunnett
- [R] **pwcompare** — almost 30 pages of informaton

5 Summary

- **marginsplot** graphs results from **margins**
- **contrast** provides a short and simple syntax for testing all kinds of factor effects
- **pwcompare** performs pairwise comparisons of marginal linear predictions
- **margins** has new **contrast** and **pwcompare** features