fvset — Declare factor-variable settings

Description

fvset base, fvset design, and fvset clear manage factor-variable settings, which identify the base level and specify how to accumulate statistics over levels. fvset base declares the base level for each specified variable; the default for factor variables without a declared base level is the lowest value. fvset design specifies how the margins command is to accumulate over the levels of a factor variable. fvset clear removes factor-variable settings for each variable in varlist. fvset clear _all removes all factor-variable settings from all variables.

fvset report reports the current factor-variable settings for each variable in varlist. fvset without arguments is a synonym for fvset report.

Quick start

Set the base category of categorical variable a1 to 3
   fvset base 3 a1

Set the base category of a2, a3, and a4 to each variable’s largest observed value
   fvset base last a2 a3 a4

Set the base category of a5 to the most frequent category
   fvset base frequent a5

Set a6 to have no base category
   fvset base none a6

Restore the default base category (first) for a5
   fvset base default a5

Specify that margins should treat a2 as though it is balanced
   fvset design asbalanced a2

Clear factor-variable settings for a2 to a4
   fvset clear a2-a4

List factor-variable settings for all factor variables
   fvset report
Syntax

Declare base settings

fvset base base_spec varlist

Declare design settings

fvset design design_spec varlist

Clear the current settings

fvset clear varlist

Report the current settings

fvset report [ varlist ] [ , base(base_spec) design(design_spec) ]

<table>
<thead>
<tr>
<th>base_spec</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>default base</td>
</tr>
<tr>
<td>first</td>
<td>lowest level value; the default</td>
</tr>
<tr>
<td>last</td>
<td>highest level value</td>
</tr>
<tr>
<td>frequent</td>
<td>most frequent level value</td>
</tr>
<tr>
<td>none</td>
<td>no base</td>
</tr>
<tr>
<td>#</td>
<td>nonnegative integer value</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>design_spec</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>default design</td>
</tr>
<tr>
<td>asbalanced</td>
<td>accumulate using $1/k$, $k =$ number of levels</td>
</tr>
<tr>
<td>asobserved</td>
<td>accumulate using observed relative frequencies; the default</td>
</tr>
</tbody>
</table>

Options

base(base_spec) restricts fvset report to report only the factor-variable settings for variables with the specified base_spec.

design(design_spec) restricts fvset report to report only the factor-variable settings for variables with the specified design_spec.
Remarks and examples

Example 1

Using auto2.dta, we include factor variable i.rep78 in a regression:

```
. use http://www.stata-press.com/data/r14/auto2
  (1978 Automobile Data)
. regress mpg i.rep78, baselevels
```

```
Source | SS      df     MS
--------+------------------
Model   | 549.415777  4  137.353944
Residual| 1790.78712  64  27.9810488
        | 2340.2029  68  34.4147485
        | Number of obs = 69  F(4, 64) = 4.91
        | Prob > F = 0.0016
        | R-squared = 0.2348
        | Adj R-squared = 0.1869
        | Root MSE = 5.2897

mpg    | Coef.  Std. Err.  t    P>|t|    [95% Conf. Interval]
--------+--------------------------------------------
rep78   |                      
Poor    | 0 (base)             
Fair    | -1.875  4.181884  -0.45 0.655  -10.22927  6.479274 
Average| -1.566667  3.863059  -0.41 0.686  -9.284014  6.150681 
Good    | .6666667  3.942718   0.17 0.866  -7.209818  8.543152 
Excellent| 6.363636  4.066234   1.56 0.123  -1.759599 14.486867 
        | _cons 21  3.740391  5.61 0.000  13.52771  28.47229 
```

We specified the `baselevels` option so that the base level would be included in the output. By default, the first level is the base level. We can change the base level to 2:

```
. fvset base 2 rep78
. regress mpg i.rep78, baselevels
```

```
Source | SS      df     MS
--------+------------------
Model   | 549.415777  4  137.353944
Residual| 1790.78712  64  27.9810488
        | 2340.2029  68  34.4147485
        | Number of obs = 69  F(4, 64) = 4.91
        | Prob > F = 0.0016
        | R-squared = 0.2348
        | Adj R-squared = 0.1869
        | Root MSE = 5.2897

mpg    | Coef.  Std. Err.  t    P>|t|    [95% Conf. Interval]
--------+--------------------------------------------
rep78   |                      
Poor    | 1.875  4.181884  0.45 0.655  -6.479274 10.22927 
Fair    | 0 (base)             
Average| .3083333  2.104836  0.15 0.884  -3.896559 4.513226 
Good    | 2.541667  2.247695  1.13 0.262  -1.948621 7.031954 
Excellent| 8.238636  2.457918  3.35 0.001  3.32838 13.14889 
        | _cons 19.125  1.870195 10.23 0.000 15.38886 22.86114 
```
Let's set `rep78` to have no base level and fit a cell-means regression:

```stata
.fvset base none rep78
.regress mpg i.rep78, noconstant
```

```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 69</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>31824.2129</td>
<td>5</td>
<td>6364.84258</td>
<td>F(5, 64) = 227.47</td>
</tr>
<tr>
<td>Residual</td>
<td>1790.78712</td>
<td>64</td>
<td>27.9810488</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>33615</td>
<td>69</td>
<td>487.173913</td>
<td>R-squared = 0.9467</td>
</tr>
</tbody>
</table>
```

```

| Coef.     | Std. Err.  | t    | P>|t| | [95% Conf. Interval] |
|-----------|------------|------|------|----------------------|
| Poor      | 21         | 3.740391 | 5.61 | 0.000  | 13.52771 28.47229 |
| Fair      | 19.125     | 1.870195 | 10.23 | 0.000  | 15.38886 22.86114 |
| Average   | 19.43333   | .9657648 | 20.12 | 0.000  | 17.504 21.36267 |
| Good      | 21.66667   | 1.246797 | 17.38 | 0.000  | 19.1759 24.15743 |
| Excellent | 27.36364   | 1.594908 | 17.16 | 0.000  | 24.17744 30.54983 |
```

**Example 2**

By default, `margins` assumes that factor variables are to be treated as observed and accumulates a margin by using the observed relative frequencies of the factor levels or the sum of the weights if weights have been specified.

```stata
.regress mpg i.foreign
```

```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>378.153515</td>
<td>1</td>
<td>378.153515</td>
<td>F(1, 72) = 13.18</td>
</tr>
<tr>
<td>Residual</td>
<td>2065.30594</td>
<td>72</td>
<td>28.6848048</td>
<td>Prob &gt; F = 0.0005</td>
</tr>
<tr>
<td>Total</td>
<td>2443.45946</td>
<td>73</td>
<td>33.4720474</td>
<td>R-squared = 0.1548</td>
</tr>
</tbody>
</table>
```

```

| Coef.     | Std. Err.  | t    | P>|t| | [95% Conf. Interval] |
|-----------|------------|------|------|----------------------|
| Foreign   | 4.945804   | 1.362162 | 3.63 | 0.001  | 2.230384 7.661225 |
| _cons     | 19.82692   | .7427186 | 26.70 | 0.000  | 18.34634 21.30751 |
```
Let's set foreign to always accumulate using equal relative frequencies:

```
.fvset design asbalanced foreign
.regress mpg i.foreign
```

```
. margins
Predictive margins Number of obs = 74  
Model VCE : OLS 
Expression : Linear prediction, predict()
```

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin</td>
<td>Std. Err.</td>
<td>t</td>
<td>P&gt;</td>
<td>t</td>
</tr>
<tr>
<td>_cons</td>
<td>21.2973</td>
<td>.6226014</td>
<td>34.21</td>
<td>0.000</td>
<td>20.05616 22.53843</td>
</tr>
</tbody>
</table>

Suppose that we issued the `fvset design` command earlier in our session and that we cannot remember which variables we set as `asbalanced`. We can retrieve this information by using the `fvset report` command:

```
. fvset report, design(asbalanced)
```

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>Base</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>foreign</td>
<td></td>
<td>asbalanced</td>
</tr>
</tbody>
</table>
```
Facebook viewset — Declare factor-variable settings

Technical note

margins is aware of a factor variable’s design setting only through the estimation results it is working with. The design setting is stored by the estimation command; thus changing the design setting between the estimation command and margins will have no effect. For example, the output from the following two calls to margins yields the same results:

```
. fvset clear foreign
. regress mpg i.foreign
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>378.153515</td>
<td>1</td>
<td>378.153515</td>
<td>F(1, 72) = 13.18</td>
</tr>
<tr>
<td>Residual</td>
<td>2065.30594</td>
<td>72</td>
<td>28.6848048</td>
<td>Prob &gt; F = 0.0005</td>
</tr>
<tr>
<td>Total</td>
<td>2443.45946</td>
<td>73</td>
<td>33.4720474</td>
<td>R-squared = 0.1548</td>
</tr>
</tbody>
</table>

| mpg      | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|------|------|---------------------|
| foreign  | 4.945804 | 1.362162  | 3.63 | 0.001 | 2.230384 7.661225   |
| _cons    | 19.82692 | .7427186  | 26.70| 0.000 | 18.34634 21.30751   |

```
. margins
```

Predictive margins Number of obs = 74
Model VCE : OLS
Expression : Linear prediction, predict()

| Delta-method | Margin | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------------|--------|-----------|------|------|---------------------|
| _cons        | 21.2973 | .6226014  | 34.21| 0.000 | 20.05616 22.53843   |

```
. fvset design asbalanced foreign
. margins
```

Predictive margins Number of obs = 74
Model VCE : OLS
Expression : Linear prediction, predict()

| Delta-method | Margin | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------------|--------|-----------|------|------|---------------------|
| _cons        | 21.2973 | .6226014  | 34.21| 0.000 | 20.05616 22.53843   |

Stored results

fvset stores the following in r():

Macros
- r(varlist) varlist
- r(baselist) base setting for each variable in varlist
- r(designlist) design setting for each variable in varlist