Tidy marginal tables for regressions

M. Escobar (modesto@usal.es)

University of Salamanca

2023 Spanish Stata Users Group meeting

Madrid, 23th October

Presentation

Aims

The purposes of this presentation are:

- Present the results of several **regressions** at the same time thanks to the new etable command.
- Introduce the meaning and importance of **margins** and **marginal effects** in regression.
- Criticize the widespread use of the **category base contrasts** on the factor variables.
- Show marginals and marginal effects of several regressions in **parallel**.
- Extend these presentations to logistic and **multi-logistic models**.
- Offer an **ado program** to make these presentations easily arranged.

Definition: Marginal effects in regression refer to the change in the dependent or response variable for a small change in one of the independent variables, holding other variables constant.

In other words, they show how the dependent variable is affected by the explanatory variables.

Marginal effects

In simple regression

Suppose you have a simple linear regression:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Where:

- Y is the response variable.
- X_1 is the predictor.
- β_0 and β_1 are the regression coefficients.
- ϵ is the error term.

The marginal effect of X_1 on Y is simply β_1, meaning for each unit increase in X_1, Y will change by β_1 units, assuming all other variables are held constant.

Marginal effects

Estimation

To estimate marginal effects, the following formula is used:

$$\hat{p} = 1/w_{\cdot} \sum_{j=1}^{N} \delta_j(S_p) w_j h(z_j, \hat{\theta})$$

where $\delta_j(S_p)$ identifies elements within the supopulation S_p for the prediction of interest, $w_{\cdot}$ is $\sum_{j=1}^{N} \delta_j(S_p) w_j$, and $h(z_j, \hat{\theta})$ can be

- $\frac{\partial f(z_j, \hat{\theta})}{\partial x}$ in case of continuous predictors.
- $f(z_j, \hat{\theta}|A = k) - f(z_j, \hat{\theta}|A = base)$ in case of a discrete variable A for each k of its K values.

where $f(z_j, \hat{\theta})$ is the regression function expressing a general linear model.

Marginal effects

Importance

Marginal effects are crucial for understanding the relationship between variables in a regression model. They allow for a clearer, direct interpretation of regression coefficients, especially in non-linear models where the coefficients do not directly represent the changes in the dependent variable associated with a unit change in the independent variable.

Data

4 variables to analyze

Data from the CIS-3411 (2023). (A sample of 29,201 respondents).

Table: Variables and its categories

Variables	Categories
Probability of vote	[0-10] (4 variables: PSOE, PP, VOX, Sumar)
Vote intention	(PSOE, PP, VOX, Sumar)
Gender	(Male, Female)
Age	(continuous)
Ideology	(Left, Center, Right)
Religion	(Practitioner, Believer, Others)

Data distribution

Frequencies, percentages, means and standard deviations

Table: Descriptive statistics

	N/Mean	(Percent)/SD
Voting intention		
PSOE	7,200	(31%)
PP	7,592	(32%)
VOX	2,652	(11%)
Sumar	3,522	(15%)
Others	2,463	(11%)
vp(PSOE)	4.2	3.9
vp(PP)	3.7	3.8
vp(VOX)	2.0	3.4
vp(Sumar)	2.5	2.9
Gender		
Male	15,023	(51%)
Female	14,178	(49%)
Age	51.5	16.4
Ideology		
Left	12,188	(43%)
Center	8,750	(31%)
Right	7,328	(26%)
Religion		
Practicing	5,398	(19%)
Non-practicing	10,980	(38%)
Others	12,508	(43%)

Regression

Commands regress and margins

- Stata has good commands for obtaining regressions, margins and marginal effects.
- On the one hand, we have the classic **regress**, on the other hand, **margins**.
- Version 18 has also incorporated etable which works well in three conditions:
  - There is only one equation
  - There are several equations but always with the same dependent variable.
  - There are diferent dependent variables but with the same independent variables (mvregress)
  - Only for the case of regressions, but not of marginals.

Regression

Voting PSOE on gender, age, ideology and religion

Table: Regression of voting PSOE

```
. regress pPSOE i.gender age i.ideology i.religion
```

Source	SS	df	MS		
Model	136494.696	6	22749.116	Number of obs	= 27,806
Residual	286559.494	27,799	10.3082663	F(6, 27799)	= 2206.88
				Prob > F	= 0.0000
				R-squared	= 0.3226
				Adj R-squared	= 0.3225
Total	423054.19	27,805	15.2150401	Root MSE	= 3.2106

pPSOE	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
gender						
Female	.703684	.0388779	18.10	0.000	.6274813	.7798866
age	.0272726	.0012246	22.27	0.000	.0248723	.0296728
ideology						
Center	-3.403105	.0474241	-71.76	0.000	-3.496059	-3.310152
Right	-5.015372	.0522556	-95.98	0.000	-5.117796	-4.912949
religion						
Non-practicing	.5224136	.0562724	9.28	0.000	.4121169	.6327102
Others	.3425385	.060363	5.67	0.000	.224224	.460853
_cons	4.505728	.0958141	47.03	0.000	4.317927	4.693528

Margins

Margins of voting PSOE on gender, ideology and religion

Table: Margins of voting PSOE vs. mean

```
. margins gender religion ideology
Predictive margins                                Number of obs = 27,806
Model VCE: OLS
Expression: Linear prediction, predict()
```

	Margin	Delta-method std. err.	t	P>|t|	[95% conf.	interval]
gender						
Male	3.901896	.0268702	145.21	0.000	3.849229	3.954563
Female	4.60558	.0278595	165.31	0.000	4.550974	4.660186
religion						
Practicing	3.892569	.0480662	80.98	0.000	3.798357	3.986781
Non-practicing	4.414983	.0316367	139.55	0.000	4.352973	4.476992
Others	4.235108	.0311933	135.77	0.000	4.173967	4.296248
ideology						
Left	6.588571	.0307639	214.17	0.000	6.528272	6.648869
Center	3.185465	.0350743	90.82	0.000	3.116718	3.254213
Right	1.573198	.0395775	39.75	0.000	1.495624	1.650772

Alternative margins

Margins of voting PSOE on gender, ideology and religion vs. mean

Table: Margins of voting PSOE vs. mean

```
. quietly: summarize pPSOE
. margins gender ideology religion, exp(predict(xb) - `r(mean)´)
Predictive margins                                 Number of obs = 27,806
Model VCE: OLS
Expression: predict(xb) - 4.204353251074449
```

	Margin	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
gender						
Male	-.3024574	.0268702	-11.26	0.000	-.3551221	-.2497927
Female	.4012266	.0278595	14.40	0.000	.346623	.4558301
ideology						
Left	2.384217	.0307639	77.50	0.000	2.323921	2.444513
Center	-1.018888	.0350743	-29.05	0.000	-1.087632	-.9501435
Right	-2.631155	.0395775	-66.48	0.000	-2.708726	-2.553585
religion						
Practicing	-.3117839	.0480662	-6.49	0.000	-.4059919	-.2175759
Non-practicing	.2106297	.0316367	6.66	0.000	.1486229	.2726365
Others	.0307546	.0311933	0.99	0.324	-.0303831	.0918924

Regular marginal effects

Marginal effects of gender, age, ideology and religion

Table: Marginal effects on voting PSOE

```
. margins, dydx(*)
Average marginal effects                        Number of obs = 27,806
Model VCE: OLS
Expression: Linear prediction, predict()
dy/dx wrt:  2.gender age 2.ideology 3.ideology 2.religion 3.religion
```

	Delta-method dy/dx	std. err.	t	P>|t|	[95% conf.	interval]
gender						
Female	.703684	.0388779	18.10	0.000	.6274813	.7798866
age	.0272726	.0012246	22.27	0.000	.0248723	.0296728
ideology						
Center	-3.403105	.0474241	-71.76	0.000	-3.496059	-3.310152
Right	-5.015372	.0522556	-95.98	0.000	-5.117796	-4.912949
religion						
Non-practicing	.5224136	.0562724	9.28	0.000	.4121169	.6327102
Others	.3425385	.060363	5.67	0.000	.224224	.460853

```
Note: dy/dx for factor levels is the discrete change from the base level.
```

Multiple regression

Several response variables with the same predictors

Table: Regression of voting on gender, age, religion and ideology

```
. quietly: mvreg pPSOE-pSumar = i.gender age i.religion i.ideology
. quietly: etable, showeq column(index) showstars
. collect layout (colname#result[_r_b _r_se] result[N]) (coleq#stars)
```

	vp(PSOE)	vp(PP)	vp(VOX)	vp(Sumar)
Gender				
Female	0.701 **	0.065	-0.615 **	0.268 **
	(0.039)	(0.035)	(0.032)	(0.027)
Age	0.027 **	-0.006 **	-0.022 **	0.004 **
	(0.001)	(0.001)	(0.001)	(0.001)
Religion				
Non-practicing	0.533 **	-0.693 **	-0.357 **	0.198 **
	(0.056)	(0.050)	(0.047)	(0.040)
Others	0.353 **	-1.844 **	-0.941 **	1.184 **
	(0.061)	(0.054)	(0.050)	(0.042)
Ideology				
Center	-3.398 **	3.347 **	1.705 **	-3.078 **
	(0.047)	(0.042)	(0.039)	(0.033)
Right	-5.011 **	4.888 **	4.545 **	-3.405 **
	(0.052)	(0.047)	(0.043)	(0.037)
Intercept	4.489 **	2.709 **	2.284 **	3.411 **
	(0.096)	(0.086)	(0.080)	(0.067)

Ado dime

An ado with different

dime (differenciated marginal effects) has 3 modalities.

- **regress** uses multiple regress commands with equations separated with parentheses, and obtains:
  - means of response variables and R2 for each equation
  - Marginal effects vs. means with their significance
- **sureg** uses sureg command (Zellner's seemingly unrelated regression) with equations separated with parentheses, and obtains:
  - Frequency, margin, marginal effect and its significance for every category of the independent variables.
  - N, mean and R squared for each response variable.
- **mlogit** uses a mlogit command, and obtains.
  - Frequency of every category of the independent variables (number of cases for quantitative variables).
  - Margins, marginal effects and their significance on every category of the response for every category of independent variables.

Command

dime

`dime regress` (*depvar1 varlist1*) (*depvar2 varlist2*) ... (*depvarN varlistN*) [*if*] [*in*] [*weight*] [, *options*]

It accepts the same options as the regress command.

`dime sureg` (*depvar1 varlist1*) (*depvar2 varlist2*) ... (*depvarN varlistN*) [*if*] [*in*] [*weight*] [, *options*]

It has three kinds of options:

- vce(vcetype) specifies the type of standard error reported.
- export(filename.suffix) to export the table to a file.
- graph, pvalue, bonferroni, and linkbipolar are for network graphs.

`dime mlogit` *depvar varlist* [*if*] [*in*] [*weight*] [, *options*]

It as the same options as `dime sureg`.

Tidy marginal effects

Tidy marginal effects on gender, age, ideology and religion vs. mean

Table: Marginal effects of voting PSOE vs. mean

```
. dime regress (pPSOE i.gender age i.ideology i.religion)
```

	vp(PSOE)
mean	4.2
Gender	
Male	-0.3 ***
Female	0.4 ***
Ideology	
Left	2.3 ***
Center	-1.1 ***
Right	-2.7 ***
Religion	
Practicing	-0.3 ***
Non-practicing	0.2 ***
Others	-0.0
R-squared	0.32
Number of observations	(27,806)

*** p<.001, ** p<.01, * p<.05

Multiple tidy marginal effects

Tidy marginal effects of gender, age, ideology and religion vs. mean

Table: Margins and marginal effects on mean of voting

```
. dime sureg (pPSOE i.gender age i.ideology i.religion) ///
             (pPP i.gender age i.ideology i.religion) ///
             (pVOX i.gender age i.ideology i.religion) ///
             (pSumar i.gender age i.ideology i.religion)
```

Table of margins and marginal effects from the mean.

	Total	vp(PSOE)		vp(PP)		vp(VOX)		vp(Sumar)	
Total	(27,713)	4.2		3.7		2.0		2.5	
Male	(14,364)	3.9	-0.3 ***	3.6	-0.0	2.3	0.3 ***	2.4	-0.1 ***
Female	(13,349)	4.6	0.4 ***	3.7	0.0	1.7	-0.3 ***	2.7	0.1 ***
Age	(29,201)	4.3	0.0 ***	3.7	-0.0 ***	2.0	-0.0 ***	2.5	0.0 ***
Left	(12,029)	6.6	2.3 ***	1.4	-2.3 ***	0.3	-1.7 ***	4.3	1.8 ***
Center	(8,505)	3.2	-1.1 ***	4.7	1.1 ***	2.0	0.0	1.3	-1.3 ***
Right	(7,179)	1.6	-2.7 ***	6.3	2.6 ***	4.9	2.8 ***	0.9	-1.6 ***
Practicing	(5,027)	3.9	-0.4 ***	4.7	1.1 ***	2.6	0.5 ***	1.9	-0.6 ***
Non-practicing	(10,569)	4.4	0.2 ***	4.0	0.4 ***	2.2	0.2 ***	2.1	-0.4 ***
Others	(12,117)	4.2	-0.0	2.9	-0.8 ***	1.6	-0.4 ***	3.1	0.6 ***
R2		0.32		0.43		0.37		0.41	

*** p<.001, ** p<.01, * p<.05

Multiple tidy marginal effects in mlogit regression

Tidy marginal effects of gender, age, ideology and religion vs. mean

Table: Margins and marginal effects vs. main percentages of voting

```
. dime mlogit intencionr i.gender age i.ideology i.religion
Table of marginals and global mean differences
```

	Total	PSOE		PP		VOX		Sumar	
Total	(22,777) 30.7			32.4		11.0		15.3	
Male	(11,966) 27.7	-3.0	***	31.4	-1.0 **	14.0	3.0 ***	15.5	0.1
Female	(10,811) 34.1	3.4	***	33.5	1.1 **	7.7	-3.4 ***	15.2	-0.1
Age	(29,201) 31.1	0.4	***	32.5	0.1 ***	10.8	-0.3 ***	15.2	-0.1 ***
Left	(10,030) 56.4	25.7	***	3.2	-29.3 ***	0.9	-10.2 ***	25.8	10.5 ***
Center	(6,190) 21.2	-9.6	***	49.3	16.9 ***	11.2	0.2	4.4	-10.9 ***
Right	(6,557) 8.8	-21.9	***	56.4	24.0 ***	26.9	15.9 ***	3.6	-11.8 ***
Practicing	(4,298) 32.8	2.1	**	38.9	6.5 ***	12.1	1.0 *	6.8	-8.5 ***
Non-practicing	(8,768) 35.8	5.1	***	34.6	2.1 ***	11.4	0.4	8.5	-6.8 ***
Others	(9,711) 31.1	0.3		24.3	-8.1 ***	10.3	-0.8 *	21.1	5.8 ***

```
*** p<.001, ** p<.01, * p<.05
Pseudo R2: 0.279; Nagelkerke´s R2: 0.596; chi2: 19033.47; p: 0
```

Remarks

About coincidence analysis

- A manner of presenting margins and marginal effects has been proposed.
- It has used the new capabilities of the Stata `collect` command.
- This may be useful to represent friendly and tidily several (multiple) regressions and (multinomial) logistic models
- New models as poisson and negative binomial are going to be incorporated.
- Among other possible challenges would be the consideration of interaction terms.

Availability of dime

How to use it

- `dime` can be used from Stata version 18 onwards.
- Soon it will be available on Git-Hub.
- Later it will be added to the coin package and available on SSC.
- If you want to try it, you could ask for a copy of the program.
- Comments and suggestions will be welcome!!

Last slide

Thanks

¡Gracias por la atención prestada!

modesto@usal.es