

## Assorted comments on the adjust, prvalue and prt看 commands

**NOTE:** If you type `help adjust` in Stata 11, you get the following message:

```
adjust has been superseded by margins. margins can do everything
that adjust did and more. margins syntax differs from adjust; see
margins. adjust continues to work but does not support factor
variables and will often fail if you do not run your estimation command
under version control, with the version set to less than 11. This help file
remains to assist those who encounter an adjust command in old do-
files and programs.
```

I haven't quite mastered the new `margins` command yet but I'll toss in a few examples.

1. The commands can often do the same thing, with slightly different syntax. Which you use may be just a matter of personal preference.

*a) prvalue versus adjust, Long & Freese (2006) p. 161*

```
. use http://www.nd.edu/~rwilliam/xsoc73994/long2006/binlfp2.dta, clear
(Data from 1976 PSID-T Mroz)
. quietly logit lfp k5 k618 age wc hc lwg inc
. prvalue, x(age=35 k5=2 wc=0 hc=0 inc=15) rest(mean)
```

logit: Predictions for lfp

Confidence intervals by delta method

```

                                95% Conf. Interval
Pr(y=inLF|x):      0.1318    [ 0.0556,    0.2081]
Pr(y=NotInLF|x):  0.8682    [ 0.7919,    0.9444]

      k5      k618      age      wc      hc      lwg      inc
x=    2  1.3532537      35      0      0  1.0971148      15
```

```
. adjust age=35 k5=2 wc=0 hc=0 inc=15 k618 lwg, pr
```

```
-----
Dependent variable: lfp      Command: logit
Covariates set to mean: k618 = 1.3532536, lwg = 1.0971148
Covariates set to value: age = 35, k5 = 2, wc = 0, hc = 0, inc = 15
-----
```

```
-----
All |          pr
-----+-----
    |          .131837
-----
Key: pr = Probability
```

```
. margins, at(age=35 k5=2 wc=0 hc=0 inc=15 k618 lwg)
```

```
Adjusted predictions      Number of obs   =      753  
Model VCE      : OIM
```

```
Expression   : Pr(lfp), predict()  
at           : k5           =           2  
             k618          =    1.353254 (mean)  
             age           =           35  
             wc            =           0  
             hc            =           0  
             lwg           =    1.097115 (mean)  
             inc           =           15
```

```
-----  
          |              Delta-method  
          |      Margin   Std. Err.      z    P>|z|    [95% Conf. Interval]  
-----+-----  
_cons | .1318369   .0389149    3.39  0.001   .0555651   .2081087  
-----
```

*b) prttab versus adjust, 1-way classification, Williams Logistic Regression III Handout*

```
. use http://www.nd.edu/~rwilliam/xsoc73994/statafiles/glm-logit.dta, clear  
. quietly logit grade gpa tuce psi  
. adjust gpa=3 tuce=20, by(psi) pr
```

```
-----  
Dependent variable: grade      Command: logit  
Covariates set to value: gpa = 3, tuce = 20  
-----
```

```
-----  
      psi |      pr  
-----+-----  
      0 | .066617  
      1 | .435077  
-----
```

Key: pr = Probability

```
. prttab psi, x(gpa=3 tuce=20)
```

logit: Predicted probabilities of positive outcome for grade

```
-----  
      psi | Prediction  
-----+-----  
      0 |    0.0666  
      1 |    0.4351  
-----
```

```
      gpa  tuce  psi  
x=      3    20 .4375
```

```
. margins, at( gpa=3 tuce=20) over(psi)
```

```
Predictive margins                                Number of obs   =           32
Model VCE      : OIM

Expression    : Pr(grade), predict()
over          : psi
at            : 0.psi
               gpa          =           3
               tuce         =           20
               1.psi
               gpa          =           3
               tuce         =           20
```

```
-----+-----
```

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
psi						
0	.066617	.0611322	1.09	0.276	-.0531999	.1864339
1	.4350765	.1812458	2.40	0.016	.0798413	.7903118

```
-----+-----
```

*c) prttab versus adjust, 2-way cross-classification, Long & Freese (2003) p. 136*

```
. use http://www.nd.edu/~rwilliam/xsoc73994/long2006/binlfp2.dta, clear
. quietly logit lfp k5 k618 age wc hc lwg inc
. prttab k5 wc, rest(mean)
```

logit: Predicted probabilities of positive outcome for lfp

```
-----+-----
```

# kids <	Wife College: 1=yes 0=no	NoCol	College
0	0.6069	0.7758	
1	0.2633	0.4449	
2	0.0764	0.1565	
3	0.0188	0.0412	

```
-----+-----
```

	k5	k618	age	wc	hc	lwg	inc
x=	.2377158	1.3532537	42.537849	.2815405	.39176627	1.0971148	20.128965

```
. adjust k618 age hc lwg inc, by (k5 wc) pr
```

```
-----+-----
```

Dependent variable: lfp      Command: logit  
Covariates set to mean: k618 = 1.3532536, age = 42.537849, hc = .39176628, lwg = 1.0971148, inc = 20.128965

```
-----+-----
```

```

-----
      |   Wife College:
# kids < |   1=yes 0=no
6       |   NoCol  College
-----+-----
      0 | .606877  .775821
      1 | .263334  .444866
      2 | .076447  .15652
      3 |           .041199
-----
Key: Probability

```

## 2. Special options on a command may make it appealing to you.

`prvalue` has assorted options that may be useful to you, e.g. `min`, `mean`, `max`. Also potentially useful are the `upper` & `lower` options, which choose the observed values that maximize/minimize the likelihood of success.

```
. prvalue, rest(lower)
```

```
logit: Predictions for lfp
```

```
Confidence intervals by delta method
```

```

                95% Conf. Interval
Pr(y=inLF|x):   0.0000  [-0.0001,  0.0001]
Pr(y=NotInLF|x): 1.0000  [ 0.9999,  1.0001]

      x=      k5      k618      age      wc      hc      lwg      inc
           3         8         60         0         0 -2.0541239      96

```

```
. prvalue, rest(upper)
```

```
logit: Predictions for lfp
```

```
Confidence intervals by delta method
```

```

                95% Conf. Interval
Pr(y=inLF|x):   0.9847  [ 0.9706,  0.9988]
Pr(y=NotInLF|x): 0.0153  [ 0.0012,  0.0294]

      x=      k5      k618      age      wc      hc      lwg      inc
           0         0         30         1         1  3.2188759  -.02900009

```

Don't just use the options mindlessly though. Make sure the combinations of values makes sense – for example, in the above, do you really think there are many 60 year old women out there with 3 kids under the age of 6, along with another 8 kids ages 6-18? Not to mention the fact that neither the husband nor the wife went to college but the husband still makes 96K?

## 3. Be careful about how values are being adjusted – the commands differ in their approaches and it isn't always obvious what is going on.

```
. use http://www.nd.edu/~rwilliam/xsoc73994/long2006/binlfp2.dta, clear
. quietly logit lfp k5 k618 age wc hc lwg inc
```

```
. prtab k5 wc
```

```
logit: Predicted probabilities of positive outcome for lfp
```

```
-----
# kids < | Wife College:
          | 1=yes 0=no
6        | NoCol College
-----+-----
          | 0 | 0.6069 0.7758
          | 1 | 0.2633 0.4449
          | 2 | 0.0764 0.1565
          | 3 | 0.0188 0.0412
-----

          k5      k618      age      wc      hc      lwg      inc
x= .2377158 1.3532537 42.537849 .2815405 .39176627 1.0971148 20.128965
```

```
. adjust, by (k5 wc) pr
```

```
-----
Dependent variable: lfp      Command: logit
Variables left as is: k618, age, hc, lwg, inc
-----
```

```
-----
# kids < | Wife College:
          | 1=yes 0=no
6        | NoCol College
-----+-----
          | 0 | .574653 .774515
          | 1 | .334376 .541162
          | 2 | .141362 .292142
          | 3 |          .032917
-----
```

```
Key: Probability
```

Why the differences? For variables not named on the command line, `prtab` fills in the *overall mean* by default. (Note how it tells you what those values are.) But, for `adjust`, the default for variables not explicitly named is to fill in the *subgroup means*, where the subgroups are defined by the `by` option (e.g. in the above there are 8 groups: women with no college and 0 kids, women with no college and 1 kid, women who went to college and have 0 kids, etc. There don't happen to be any women who didn't go to college and have 3 kids under age 6, which is why that entry is blank. Also, note that `adjust` does not report the subgroup means for each variable – with 8 groups and 5 variables, it would have to report 40 different values.) If the `by` option isn't used, then there is a single group and the overall mean for each variable is used. Both approaches have their uses but be clear as to what is going on.

4. Likewise, be careful about sample selection and missing data – the default behavior of commands is sometimes different. I find it much easier to screw up my sample selection in Stata than I do in SPSS. In SPSS, you might use a filter command to select cases, and the selection continues to hold for all analyses until you change it; in Stata you often have to keep remembering to use an `if` qualifier with every command, or else drop the cases you don't want to be analyzing. You especially have to worry about this with several Stata post-estimation commands – do they apply only to the sample used by the last estimation command, or do they apply to all cases? (Or perhaps to a different sample entirely!) The `predict` command, for example, generally applies to all cases whether they were used by the last command or not. As the help for `adjust` says,

If you have restricted your estimation command to a portion of the data using `if` or `in`, then you will generally want to use the same conditions with `adjust`. This is easily done by including `if e(sample)` with the `adjust` command. However, there may be legitimate reasons for using different data to perform the estimation and to obtain adjusted predictions.

So, for example, if you did a logistic regression using blacks only, if you did not make the same restriction using `adjust` then the adjustment would be done using both blacks and whites. This would primarily matter, I believe, for those variables whose values you did not explicitly specify on the `adjust` command (e.g. the overall mean could get used instead of the black mean). By way of contrast, with the `prvalue` command, the use of the `all` option “specifies that any calculations of means, medians, etc., should use the entire sample instead of the sample used to estimate the model.” In other words, `prvalue` limits itself to the sample you used in the estimation command, unless you tell it otherwise.

```
. use http://www.nd.edu/~rwilliam/xsoc73994/statafiles/glm-reg.dta, clear
* Logistic regression, blacks only
. quietly logit incbinary educ if black, nolog
. * Probability of success, using the mean of education for the entire sample, 13.16.
. adjust educ, pr
```

```
-----
Dependent variable: incbinary      Command: logit
Covariate set to mean: educ = 13.16
-----
```

```
-----
All |          pr
-----+-----
    |    .088808
-----
```

Key: pr = Probability

```
. prvalue, all
```

logit: Predictions for incbinary

Confidence intervals by delta method

		95% Conf. Interval	
Pr(y=1 x):	0.0888	[ 0.0022,	0.1754]
Pr(y=0 x):	0.9112	[ 0.8246,	0.9978]

```
educ
x= 13.16
```

```
. * Probability of success, using the mean for the black sample only.
. adjust educ if e(sample), pr
```

```
-----
Dependent variable: incbinary      Command: logit
Covariate set to mean: educ = 10.2
-----
```

```
-----
All |          pr
-----+-----
    |    .00844
-----
```

Key: pr = Probability

```
. prvalue
```

logit: Predictions for incbinary

Confidence intervals by delta method

		95% Conf. Interval	
Pr(y=1 x):	0.0084	[-0.0112,	0.0281]
Pr(y=0 x):	0.9916	[ 0.9719,	1.0112]

educ  
x= 10.2

The logistic regression was run on blacks only. However, the first `adjust` command (and the `prvalue` command with the `all` option) did not limit the sample to blacks only; whites were also included. As a result, `adjust` used the overall sample mean for education, 13.16, resulting in a predicted probability of success of almost 9%. The second `adjust` command (and the `prvalue` command with its default option) limited the sample to blacks (the same as what was used in the logistic regression). The average black has only 10.2 years of education, and a person with 10.2 years of education has a predicted probability of success of 0.84%. If you wanted to only analyze blacks throughout your analysis, failing to restrict all of your analyses to blacks could greatly distort your conclusions.