Predicting a clear view with margins

Bill Rising

StataCorp LP

2013 Nordic and Baltic Stata Users Group Meeting
Stockholm
25 September 2013

Getting Started

- This will be an interactive demonstration
- We want to explore the `margins` command
- On the way we will learn about making nice graphs to help with explanations

Getting Ready

- We'll be doing a lot of work with categorical variables
- By default, Stata does not show the base reference class when using factor variables
- Let's fix this

```
. set showbaselevels on, permanently
```

Starting Simple

- We will start with the low birthweight dataset from Hosmer and Lemeshow's book on logistic regression

```
. webuse lbw
```

- We can take a peek at the dataset

```
. codebook, compact
```

Linear Regression Coefficients

- Let's start simple with an easy linear regression for birthweights

```
. regress bwt c.age##c.age lwt i.smoke i.race
```

- We know what the coefficients mean
  - The coefficient for `smoke` says: all other things being equal (i.e. weight and race), we think smokers' babies are about 400 grams lighter than non-smokers' babies on average
- We can see that age looks to have a quadratic effect
  - It would be nice to see this more clearly

Asking a Different Question

- Now suppose instead of being asked about the effect of age, we are asked

  > "What would the model take for the average weight of a baby for a mother aged 30?"

- There are two paths we can take here:
  - We could plug in 30 for the age for all the women in our sample, leave all other covariates the same, predict the birthweights, and average the results
    - Order: predict with partial info, then average
  - We could average all the other covariates, set age to 30, and predict
    - Order: average, then predict with partial info
- The first path is called 'predictive margins' or 'average predictive margins'—it is the one we will take

A Predictive Margin

- Stata implements predictive margins using the `margins` command
- Here is what we could use as our best guess of the mean weight of babies of women aged 30:

```
. margins, at(age=30)
```

- Computationally, the point estimate could be computed by
  - Changing age to 30 everywhere
  - Using `predict` to get predicted values
  - Finding the mean of those values
- `margins` does more—it computes standard errors

Margins Across Multiple Values

- One of the nice things about `margins` is that it can be used get average predicted values over a range
  - This, in some way, is a way to see a variable's effect over a range of values
- So, for example, if we wanted to show someone how birthweights change by mother's age, we could make the following table

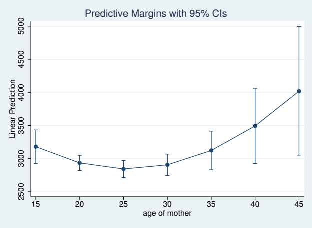
```
. margins, at(age=(15(5)45))
```

- This lets us see the weights drop and then rise
  - But it forces us to use factor variable-like notation for the ages

Picturing Predictive Margins

- Stata has a command which can draw graphs arising from the `margins` command: `marginsplot`
- Here is a very simple example

```
. marginsplot
```



  - This gives a good view of the parabolic shape
- We will now investigate margins

What About Comparing Groups?

- Suppose we would like to compare weights across the three race categories
- This is done by including `race` in the *varlist* for the `margins` command:

```
. margins race
```

  - We can do this because we specified `i.race` in the model
  - Aside: this is equivalent to specifying the race categories explicitly

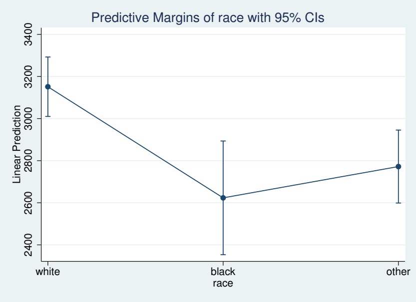
```
. margins, at(race=(1 2 3))
```

- These values can be interpreted as best-guesses for partial information for your sample

Graphing the Group Means

- We can graph these using `marginsplot`

```
. marginsplot
```

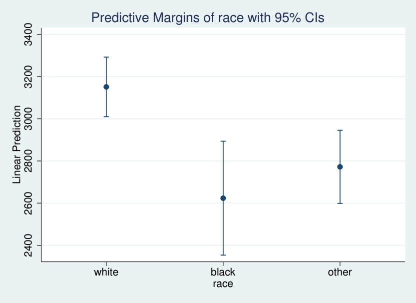


- We might want to make the graph look a little better

Making a Nicer Graph

- We can get this graph to look nicer by
  - Squelching the connecting lines with `recast(scatter)`
  - Expanding the horizontal axis with the `xscale` option
- Here is a prettier graph

```
. marginsplot, recast(scatter) xscale(range(0.5 3.5))
```



More Complex Margins

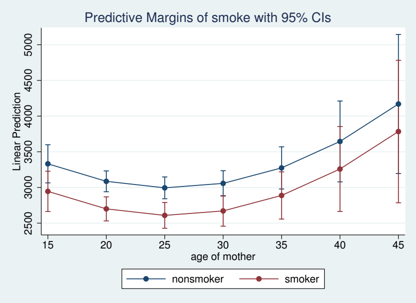
- There is no reason for us to limit our predictive margins to be computed over just one variable
- We could just as well look to see how age and smoking status work together
- Here is the the margins command

```
. margins smoke, at(age==(15(5)45))
```

Still a Simple Graph

- The `marginsplot` still makes a simple graph

```
. marginsplot
```



- Because there are no interactions, the parabolas are parallel

Driving the Point Home

- If we wanted to specify smoking and age, we could use the interaction notation in the `margins` command
  - This is true even though there were no interactions—all that is done is that all possible combinations of smoking status and race are included
- The command is not bad...

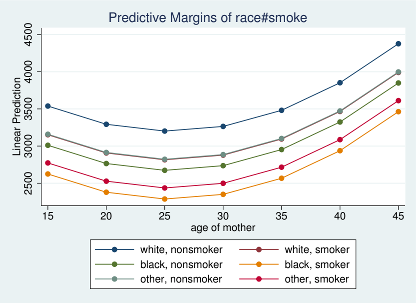
```
. margins race#smoke, at(age==(15(5)45))
```

- ... but the output is

We Can Still Visualize This

- We can still make a picture

```
. marginsplot, noci
```

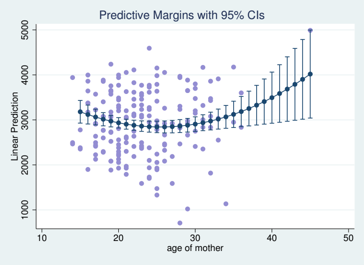


- The `noci` option squelches the confidence intervals to make a cleaner graph
  - Possibly a too-clean graph!

A Fancy Overlay

- In this particular dataset, there is an outlier: a combination of a woman who is much older with a baby whichi is much heavier
- We can make a graph which shows the effect of age together with a scatterplot

```
. quietly margins, at(age==(15(1)45))
. marginsplot, legend(off) ///
      addplot(scatter bwt age, mcolor(lavender) below)
```



A Richer Dataset

- Now will switch over to the nhanes2 dataset

```
. webuse nhanes2
```

- This is a nice dataset, because it has good things for both linear and logistic regression
- These are survey data

```
. svyset
```

  - We will need to use the svy: prefix for estimation

Looking at Diabetes

- Now we would like to look at the chances of having diabetes
- Here is a simple model

```
. svy: logistic diabetes age i.sex i.race bmi
```

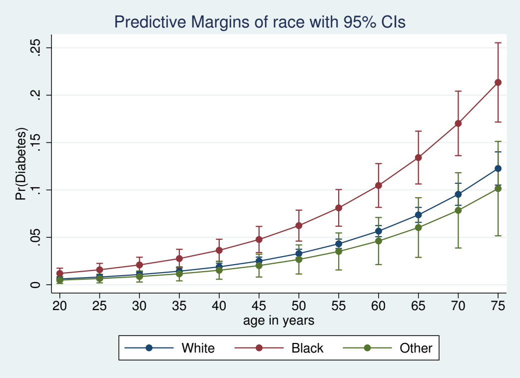
- We can see that age and bmi both increase the odds of diabetes by about 6% for each unit increase
- How does this play out for race and age?

```
. margins race, at(age==(20(5)75)) vce(uncond)
```

Here Is the Picture

- We can get a nice picture

```
. marginsplot, legend(rows(1))
```



This is Better than Odds Ratios

- This type of graph is something that makes explaining a logistic model much easier than via odds ratios
- It is as applicable to the general population as much as your belief that your sample is representative of the general population
  - Which is important for the odds ratios also
- Here, a picture is worth a thousand hard words

For Probit Fans

- If you prefer probit models, we can use the same type of logic

```
. svy: probit diabetes age i.sex i.race bmi
```

  - Now the coefficients are not very interpretable

- We can still get margins

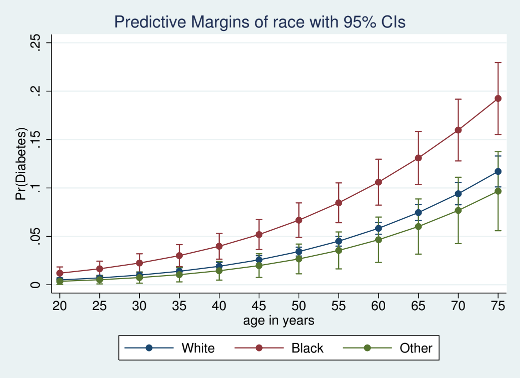
```
. margins race, at(age==(20(5)75)) vce(uncond)
```

- Creating the predictive margins still works the same

Picturing a Probit

- We can still get a very similar nice picture

```
. marginsplot, legend(rows(1))
```



Interactions

- Here is a model with interactions

```
. svy: logit diabetes c.age##c.age##race bmi i.sex
```

- If we look at the output, the higher-level interactions are needed in the model
- They are nearly impossible to picture or to talk about, however

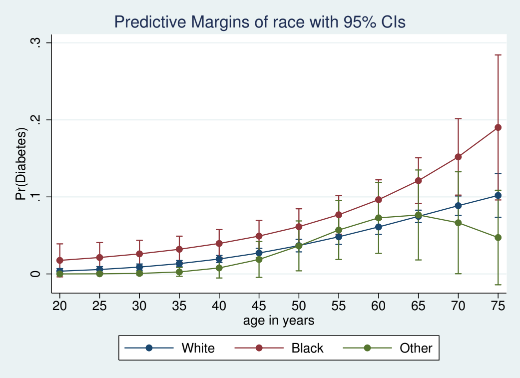
Visualizing Interactions with `marginsplot`

- Here are the margins for this complex model

```
. margins race, at(age==(20(5)75)) vce(uncond)
```

- And a nice, informative picture

```
. marginsplot, legend(rows(1))
```



Working Within Groups

- When we stated

```
. margins race, ...
```

  `margins` set every observation to each race category while computing the predictive margins

- If you would rather compute the predictive margins within each race category, use the `over` option

```
. margins, at(age==(20(5)75)) vce(uncond) over(race)
```

- The differences from before are small, and the picture is similar

```
. marginsplot, legend(rows(1))
```

Conclusion

- `marginsplot` is very good for showing how models work
- This can be used to good effect when explaining even simple non-linear models
- This can be used to good effect when explaining any type of interactions

A Fun Plot

- Just For Fun
  - Believe it or not, it is possible to make a contour plot of predictive margins

    ```
    . do margcon
    ```
  - Here is the picture