Interpreting and Visualizing Regression models with Stata Margins and Marginsplot

Boriana Pratt

May 2017

CC BY NC

Interpreting regression models

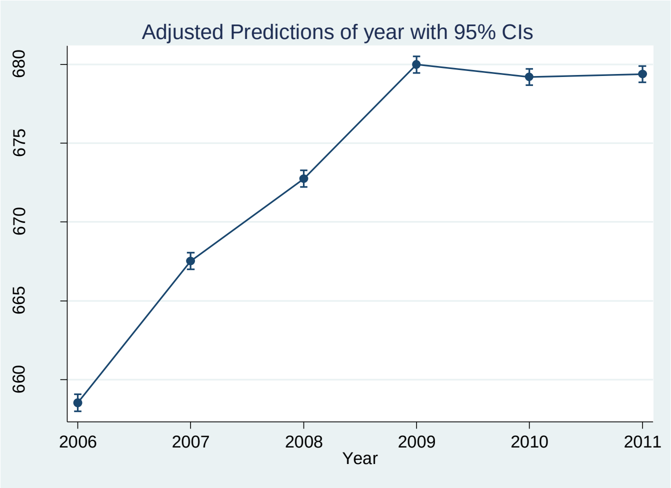
- Often regression results are presented in a table format, which makes it hard for interpreting effects of interactions, of categorical variables or effects in a non-linear models.
- For nonlinear models, such as logistic regression, the raw coefficients are often not of much interest. What we want to see for interpretation are effects on outcomes such as probabilities (instead of log odds).
- Stata has a number of handy commands such as **margins**, **marginsplot**, **contrast** for making sense of regression results and for visualizing such results.

Topics:

margins

year	Margin	Delta-method Std. Err.	t	P>|t|	[95% Conf.	Interval]
2006	658.5267	.2751133	2393.66	0.000	657.9875	659.0659
2007	667.5207	.2705957	2466.86	0.000	666.9903	668.0511
2008	672.7515	.26794	2510.83	0.000	672.2263	673.2767
2009	679.9947	.2651201	2564.86	0.000	679.4751	680.5143

marginsplot



Marginal Effects at the Mean

Average Marginal Effects

Marginal Effects at Representative values

margins, contrast

Margins : asbalanced

year	Contrast	Std. Err.	t	P>|t|
(2007 vs 2006)	8.994042	.3858877	23.31	0.000
(2008 vs 2006)	14.22482	.3840301	37.04	0.000
(2009 vs 2006)	21.46802	.382068	56.19	0.000
(2010 vs 2006)	20.67296	.380687	54.30	0.000
(2011 vs 2006)	20.86324	.379856	54.92	0.000

margins, pwcompare

Margins

What are "margins"?

Margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates. (from "margins" help)

- "conditional margin" – response at fixed values for all covariates
- "predictive margin" – response when at least one covariate is left to vary

With the "margins" command you can compute predicted levels for different covariate values or differences in levels (often called marginal effects), or even differences in differences.

Continuous vs. discrete marginal effects:

- For a continuous covariate, margins computes the first derivative of the response with respect to the covariate.
- For a discrete covariate, margins computes the effect of a discrete change of the covariate (discrete change effects).

Use margins command to get marginal means, predictive margins and marginal effects.

Datasets

NYC math assessment data for 2006-2011 by school and gender (from NYC Open Data: https://nycplatform.socrata.com/)

	dbn	grade	year	category	numbertested	meanscales~e	level1
1	01M015	3	2006	Female	23	675	0
2	01M015	3	2006	Male	16	657	2
3	01M015	3	2007	Female	11	679	2
4	01M015	3	2007	Male	20	668	0
5	01M015	3	2008	Female	17	661	0

nhanes2 (from Stata – webuse)

	sex	race	age	height	weight	bpsystol	bpdiast	tcresult
1	Male	White	54	174.598	62.48	106	80	226
2	Female	White	41	152.297	48.76	108	66	179
3	Female	Other	21	164.098	67.25	98	66	137
4	Female	White	63	162.598	94.46	180	80	189

Adjusted means

```
.use NYC_MATH_2006_2011_byschool, clear


. regress meanscore i.gender

      Source |       SS           df       MS      Number of obs   =    42,321
-------------+----------------------------------   F(1, 42319)     =    115.85
       Model |  65074.4271         1  65074.4271   Prob > F        =    0.0000
    Residual |  23771882.7    42,319  561.730728   R-squared       =    0.0027
-------------+----------------------------------   Adj R-squared   =    0.0027
       Total |  23836957.1    42,320   563.25513   Root MSE        =    23.701

------------------------------------------------------------------------------
   meanscore |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |
       Male  |    -2.4803   .2304428   -10.76   0.000    -2.931973   -2.028628
       _cons |   674.4093   .1641427  4108.68   0.000     674.0876     674.731
------------------------------------------------------------------------------
```

How to get mean scores by gender?

Adjusted means

How to get means by gender?

```
. di _b[_cons]
674.40932

. di _b[_cons] +_b[2.gender]
671.92902
```

Or, by using `margins':

```
. margins gender

Adjusted predictions                            Number of obs     =     42,321
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |
     Female  |   674.4093   .1641427  4108.68   0.000     674.0876     674.731
       Male  |    671.929   .1617439  4154.28   0.000      671.612     672.246
------------------------------------------------------------------------------
```

Predicted means

```
. regress meanscore i.gender year

      Source |       SS           df       MS      Number of obs   =    42,321
-------------+----------------------------------   F(2, 42318)     =   2114.36
       Model |  2165561.58         2  1082780.79   Prob > F        =    0.0000
    Residual |  21671395.5    42,318  512.108217   R-squared       =    0.0908
-------------+----------------------------------   Adj R-squared   =    0.0908
       Total |  23836957.1    42,320   563.25513   Root MSE        =     22.63

------------------------------------------------------------------------------
   meanscore |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |
       Male  |   -2.47607    .220029   -11.25   0.000    -2.907331   -2.044809
        year |   4.137436   .0646029    64.04   0.000     4.010813    4.264059
       _cons |  -7635.866   129.7587   -58.85   0.000    -7890.195   -7381.536
------------------------------------------------------------------------------
```

How to get predicted mean scores for 2006 and 2008 by gender?

```
. di _b[_cons] +_b[year]*2006
663.83028

. di _b[_cons] +_b[2.gender] +_b[year]*2006
661.35421

. di _b[_cons] +_b[year]*2008
672.10515

. di _b[_cons] +_b[2.gender] +_b[year]*2008
669.62908
```

Predicted meanscore for female in 2006

Predicted meanscore for male in 2006

Predicted meanscore for female in 2008

Predicted meanscore for male in 2008

Predicted means

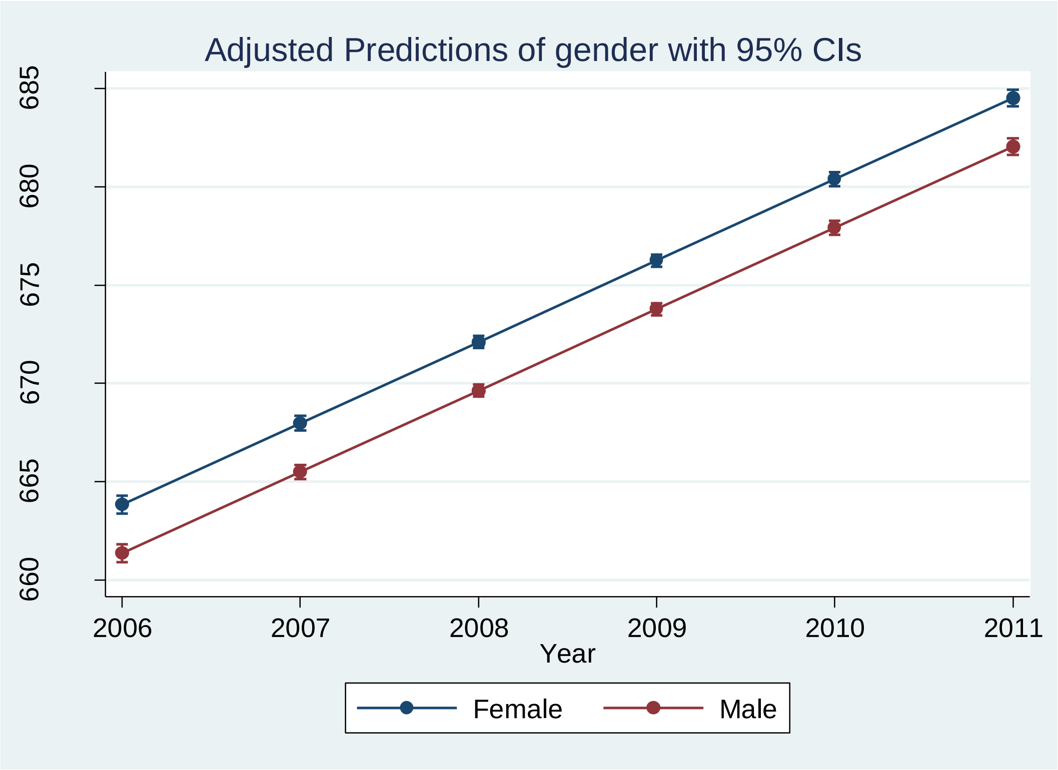
How to get predicted mean scores for 2006 and 2008 by gender?

```
. margins gender, at(year=(2006 2008)) vsquish

Adjusted predictions                            Number of obs     =     42,321
Model VCE    : OLS

Expression   : Linear prediction, predict()
1._at        : year            =        2006
2._at        : year            =        2008

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _at#gender |
   1#Female  |   663.8303   .2277025  2915.34   0.000      663.384    664.2766
     1#Male  |   661.3542   .2260839  2925.26   0.000     660.9111    661.7973
   2#Female  |   672.1051   .1608015  4179.72   0.000       671.79    672.4203
     2#Male  |   669.6291   .1585551  4223.32   0.000     669.3183    669.9398
------------------------------------------------------------------------------
```

!Note: gender has to be a factor variable in the model.

Marginsplot

```
. margins gender, at(year=(2006(1)2011)) vsquish

. marginsplot
```



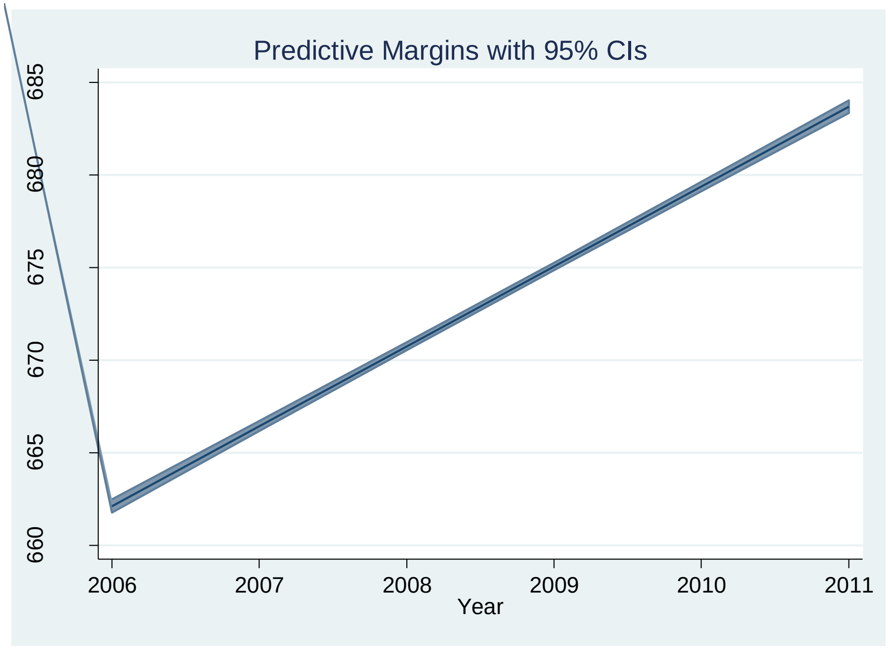
!Note: "marginsplot" has to be right after "margins".

Marginsplot

```
. margins, at(year=(2006(1)2011)) vsquish
```

To plot with confidence interval lines:

```
. marginsplot, recast(line) recastci(rarea) ciopts(color(*.7))
```



Predicted means

Q: What does the following command produce?

```
. margins, at(year=(2006 2008)) vsquish
```

Q: How can you get the same numbers from the regression equation?

Predicted means

Q: What does the following command produce?

```
. margins, at(year=(2006 2008)) vsquish
```

```
. margins , at(year=(2006 2008)) vsquish

Predictive margins                              Number of obs     =     42,321
Model VCE    : OLS

Expression   : Linear prediction, predict()
1._at        : year            =        2006
2._at        : year            =        2008

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |    662.574   .1984318  3339.05   0.000     662.1851    662.9629
          2  |   670.8489   .1157263  5796.86   0.000     670.6221    671.0757
------------------------------------------------------------------------------
```

Q: How can you get the same numbers from the regression equation?

```
. di _b[_cons] +_b[2.gender]*0.5074 +_b[year]*2006
662.57392

. di _b[_cons] +_b[2.gender]*0.5074 +_b[year]*2008
670.84879
```

Predicted means

Margins option – treat all factor variables as balanced:

```
. margins , at(year=(2006 2008)) vsquish asbalanced

Adjusted predictions                            Number of obs     =     42,321
Model VCE    : OLS

Expression   : Linear prediction, predict()
1._at        : gender                       (asbalanced)
               year            =        2006
2._at        : gender                       (asbalanced)
               year            =        2008

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   662.5922   .1984388  3339.03   0.000     662.2033    662.9812
          2  |   670.8671   .1157378  5796.44   0.000     670.6403     671.094
------------------------------------------------------------------------------
```

In the data there are 50.7% boys and 49.3% girls, with the “asbalanced” option margins predicts the means if the data had 50% boys and 50% girls.

Average marginal effects

```
. margins , dydx(*)

Average marginal effects                          Number of obs     =     42,321
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 2.gender year

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |
       Male  |   -2.47607    .220029   -11.25   0.000    -2.907331   -2.044809
        year |   4.137436   .0646029    64.04   0.000     4.010813    4.264059
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Marginal effect (ME) measures the effect on the conditional mean of y of a change in one of the regressors .
In the linear regression model, the marginal effect equals the relevant slope coefficient. For non-linear models this is not the case and hence there are different methods for calculating marginal effects.

Binary outcome

Let's look at a logistic model:

```
. webuse nhanes2

. logit highbp i.sex i.agegrp bmi weight
```

```
Logistic regression                             Number of obs     =     10,351
                                                LR chi2(8)        =    2385.33
                                                Prob > chi2       =     0.0000
Log likelihood = -5858.1005                     Pseudo R2         =     0.1692

------------------------------------------------------------------------------
      highbp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     Female  |  -.3810961   .0630679    -6.04   0.000     -.504707   -.2574853
             |
      agegrp |
      30-39  |   .4919107   .0823427     5.97   0.000     .3305221    .6532994
      40-49  |   .9785078   .0839912    11.65   0.000     .8138882    1.143127
      50-59  |   1.594481   .0829633    19.22   0.000     1.431876    1.757086
      60-69  |   1.817423   .0714901    25.42   0.000     1.677305    1.957541
        70+  |   2.265552   .0930956    24.34   0.000     2.083088    2.448016
             |
         bmi |   .1137638   .0112739    10.09   0.000     .0916673    .1358603
      weight |    .008617   .0037796     2.28   0.023     .0012091    .0160248
       _cons |  -4.843599   .1488107   -32.55   0.000    -5.135263   -4.551935
------------------------------------------------------------------------------
```

Adjusted predictions at the means

```
. margins sex, atmeans

Adjusted predictions                            Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()
at           : 1.sex           =    .4748333 (mean)
               2.sex           =    .5251667 (mean)
               1.agegrp        =    .2241329 (mean)
               2.agegrp        =    .1566998 (mean)
               3.agegrp        =    .1228867 (mean)
               4.agegrp        =    .1247222 (mean)
               5.agegrp        =    .2763018 (mean)
               6.agegrp        =    .0952565 (mean)
               bmi             =     25.5376 (mean)
               weight          =    71.89752 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
       Male  |   .4490167   .0097773    45.92   0.000     .4298536    .4681799
     Female  |   .3576128   .0088545    40.39   0.000     .3402584    .3749672
------------------------------------------------------------------------------
```

The probability of an “average” man to have high blood pressure

The probability of an “average” woman to have high blood pressure

“average” here means with weight of 71.9 kg, bmi of 25.5, 9.5% in age group6, 27.6% in age group5, etc.

Average predictive margins

```
. margins sex

Predictive margins                              Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
       Male  |   .4601265   .0075996    60.55   0.000     .4452316    .4750214
     Female  |   .3864881    .007311    52.86   0.000     .3721589    .4008173
------------------------------------------------------------------------------
```

On average the probability for a man to have high blood pressure is 46%.

On average the probability for a woman to have high blood pressure is 38.6%.

Above is the same as if option 'asobserved' (the default) is used:

```
. margins sex, asobserved
```

Marginal effect at the mean (MEM)

```
. margins, dydx(sex) atmeans

Conditional marginal effects                    Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()
dy/dx w.r.t. : 2.sex
at           : 1.sex            =    .4748333 (mean)
               2.sex            =    .5251667 (mean)
               1.agegrp         =    .2241329 (mean)
               2.agegrp         =    .1566998 (mean)
               3.agegrp         =    .1228867 (mean)
               4.agegrp         =    .1247222 (mean)
               5.agegrp         =    .2763018 (mean)
               6.agegrp         =    .0952565 (mean)
               bmi              =     25.5376 (mean)
               weight           =    71.89752 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     Female  |  -.0914039   .0150619    -6.07   0.000    -.1209246   -.0618832
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

The probability of an "average" woman to have high blood pressure is 9% less than that for an "average" man, where "average" means a person with bmi=25.5376, weight=71.897 and part of all age groups.

Average marginal effect (AME)

```
. margins, dydx(sex)

Average marginal effects                        Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()
dy/dx w.r.t. : 2.sex

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     Female  |  -.0736384   .0121553    -6.06   0.000    -.0974624   -.0498144
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

On average, the probability of a woman to have high blood pressure is 7% less than that for a man.

There are other ways to get the same AME (because sex has only two categories):

```
. margins sex, contrast(nowald pveffects)
. margins r.sex
```

How is AME computed?

- Go to the first case. Treat that person as though s/he were male, regardless of their sex. Leave all other independent variables as they are. Compute the probability that this person (if male) would have high blood pressure.
- Do the same thing, this time treating the person as though they were a female.
- The difference in the two probabilities just computed is the marginal effect for that case.
- Repeat the process for every case in the sample.
- Compute the average of all the marginal effects you have computed. This gives you the AME for female.

How is AME computed?

Compute log-odds for everyone if “male”:

```
.gen mlodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight
.replace mlodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.agegrp] if agegrp==2
.replace mlodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[3.agegrp] if agegrp==3
.replace mlodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[4.agegrp] if agegrp==4
.replace mlodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[5.agegrp] if agegrp==5
.replace mlodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[6.agegrp] if agegrp==6
```

Compute log-odds for everyone if “female”:

```
.gen flodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.sex]
.replace flodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.sex] +_b[2.agegrp] if agegrp==2
.replace flodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.sex] +_b[3.agegrp] if agegrp==3
.replace flodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.sex] +_b[4.agegrp] if agegrp==4
.replace flodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.sex] +_b[5.agegrp] if agegrp==5
.replace flodds_ = _b[_cons] +_b[bmi]*bmi +_b[weight]*weight +_b[2.sex] +_b[6.agegrp] if agegrp==6
```

Compute predicted probabilities for everyone:

```
gen fpp = exp(flodds_)/ (1+ exp(flodds_))
gen mpp = exp(mlodds_)/ (1+ exp(mlodds_))
```

Compute marginal effect for everyone and get the average of these:

```
.gen me = fpp - mpp
.summ me
```

Marginal effect at a Representative value (MER)

```
. margins, dydx(sex) at(agegrp=(1 4 6)) vsquish

Average marginal effects                        Number of obs     =     10,351
Model VCE     : OIM

Expression    : Pr(highbp), predict()
dy/dx w.r.t. : 2.sex
1._at         : agegrp          =           1
2._at         : agegrp          =           4
3._at         : agegrp          =           6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
2.sex        |
         _at |
          1  |  -.0556457   .0094367    -5.90   0.000    -.0741413   -.0371501
          2  |  -.0868295   .0143506    -6.05   0.000    -.1149562   -.0587027
          3  |  -.0789143   .0132599    -5.95   0.000    -.1049033   -.0529253
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

The probability of a woman, who is 70+ years of age, to have high blood pressure is almost 8% less than that for a man of similar age. Where the probability of a young woman (20-29years old) to have high blood pressure is 5.5% less than that of a man of similar age.

Another way to get the same estimate:

```
. margins r.sex, at(agegrp=(1 4 6)) vsquish
```

Predicted margins and marginal effects

Q: What do the following commands give you (for the model for high blood pressure)?

```
. margins, dydx(sex) asbalanced
```

```
. margins
```

Q. Would you expect the previous number to be the same as the following:

```
. margins, atmeans
```

```
. margins, at(agegrp=(1 6) ) vsquish
```

Predicted Margins

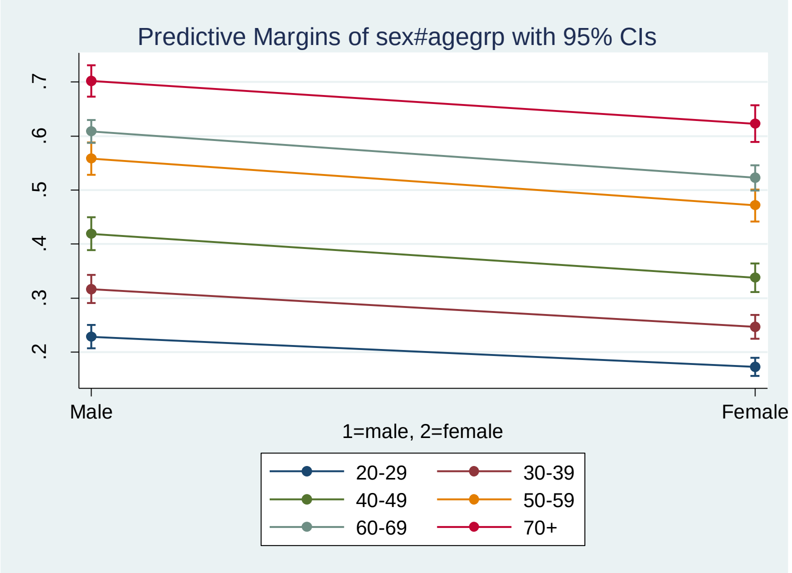
To get predicted probability separate for men and women in each age category:

```
. logit highbp i.sex i.agegrp bmi weight


. margins sex#agegrp

Predictive margins                              Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  sex#agegrp |
 Male#20-29  |   .2285314   .0109208    20.93   0.000     .2071271    .2499358
 Male#30-39  |      .3167   .0133626    23.70   0.000     .2905098    .3428902
 Male#40-49  |   .4189904   .0154765    27.07   0.000      .388657    .4493238
 Male#50-59  |   .5585883   .0153028    36.50   0.000     .5285954    .5885812
 Male#60-69  |   .6084001   .0108451    56.10   0.000     .5871441    .6296561
   Male#70+  |   .7019608    .014816    47.38   0.000     .6729221    .7309995
Female#20-29 |   .1728857   .0086179    20.06   0.000     .1559949    .1897766
Female#30-39 |   .2468291   .0110757    22.29   0.000     .2251212    .2685371
Female#40-49 |   .3377913   .0136866    24.68   0.000      .310966    .3646166
Female#50-59 |   .4717589   .0150666    31.31   0.000      .442229    .5012888
Female#60-69 |   .5226117   .0120877    43.23   0.000     .4989202    .5463033
 Female#70+  |   .6230465   .0173389    35.93   0.000     .5890629    .6570301
------------------------------------------------------------------------------
```

On average the probability of a woman, 70+ years of age, to have high blood pressure is 62.3%.

Marginsplot

```
. marginsplot

  Variables that uniquely identify margins: sex agegrp
```

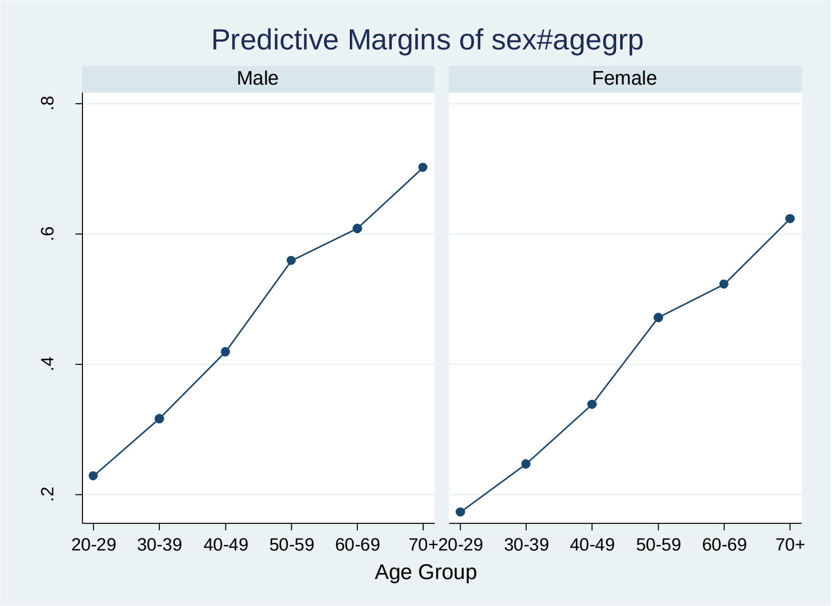


Predicted probabilities of high blood pressure are higher for men than for women in all age groups.

The probability of getting high blood pressure goes up as age goes up for both men and women.

To get separate graphs by sex (on the same plot):

```
. marginsplot, by(sex) noci

  Variables that uniquely identify margins: sex agegrp
```



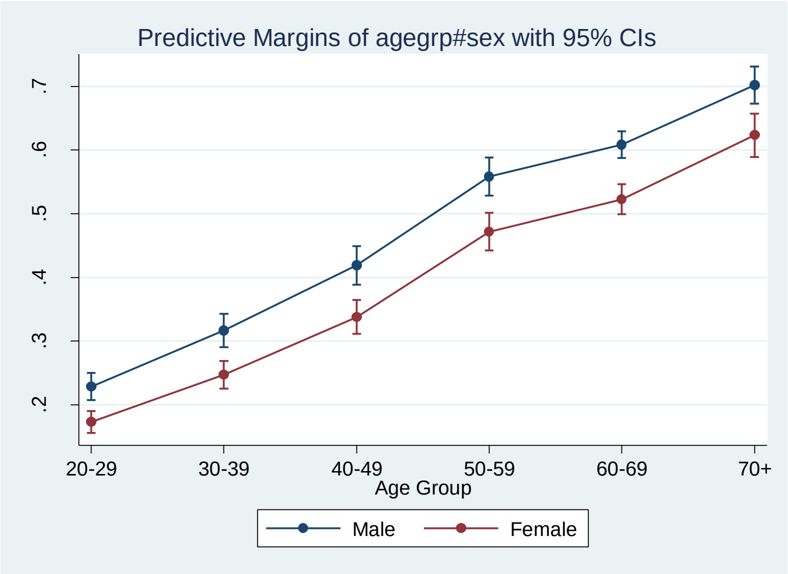
Marginsplot

To plot the two curves for sex on the same plot, you could either use the x option:

```
. marginsplot, x(agegrp)

  Variables that uniquely identify margins: sex agegrp
```

Or change the margins command:

```
. margins agegrp#sex
. marginsplot

  Variables that uniquely identify margins: sex agegrp
```



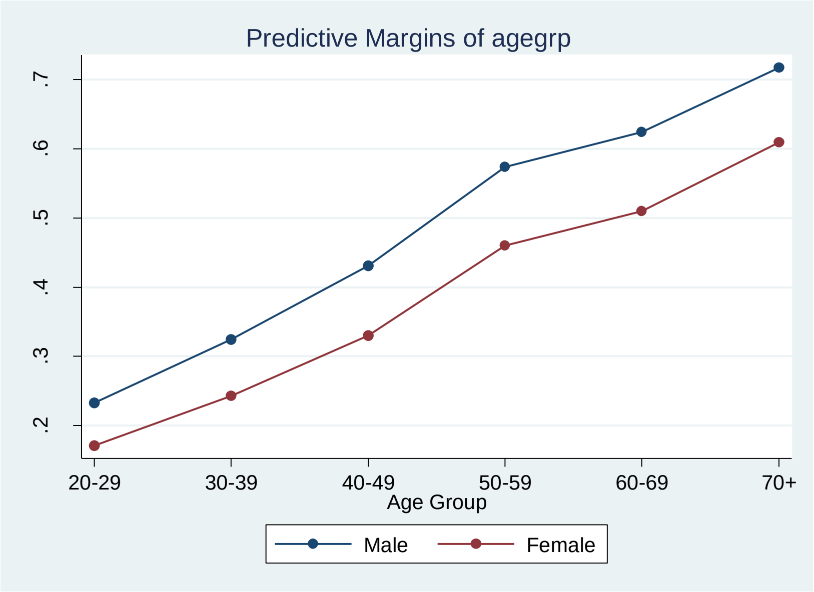
Margins

In the previous command, each observation was treated as part of each category when getting the predicted probabilities. If you would like predicted probabilities for each age category (for example) using the data on sex as is (not assuming once the person being female, then the same person being male) then you should use the over() option:

```
. margins agegrp, over(sex)
(output omitted)
```

To plot without confidence intervals:

```
. marginsplot, noci

  Variables that uniquely identify margins: sex agegrp
```



MEM (marginal effects at the means)

```
. margins , dydx(*) atmeans

Conditional marginal effects                    Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()
dy/dx w.r.t. : 2.sex 2.agegrp 3.agegrp 4.agegrp 5.agegrp 6.agegrp bmi weight
at           : 1.sex           =    .4748333 (mean)
               2.sex           =    .5251667 (mean)
               1.agegrp        =    .2241329 (mean)
               2.agegrp        =    .1566998 (mean)
               3.agegrp        =    .1228867 (mean)
               4.agegrp        =    .1247222 (mean)
               5.agegrp        =    .2763018 (mean)
               6.agegrp        =    .0952565 (mean)
               bmi             =     25.5376 (mean)
               weight          =    71.89752 (mean)
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     Female  |  -.0914039   .0150619    -6.07   0.000    -.1209246   -.0618832
             |
      agegrp |
      30-39  |   .0840426   .0142145     5.91   0.000     .0561826    .1119026
      40-49  |    .188469   .0165182    11.41   0.000     .1560939    .2208441
      50-59  |   .3392582   .0170405    19.91   0.000     .3058594    .3726569
      60-69  |   .3944535    .013179    29.93   0.000     .3686231    .4202838
        70+  |   .4988221   .0179319    27.82   0.000     .4636762     .533968
             |
         bmi |    .027307   .0027072    10.09   0.000      .022001    .0326131
      weight |   .0020684   .0009072     2.28   0.023     .0002903    .0038464
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

AME (average marginal effects)

```
. margins , dydx(*)

Average marginal effects                        Number of obs     =     10,351
Model VCE    : OIM

Expression   : Pr(highbp), predict()
dy/dx w.r.t. : 2.sex 2.agegrp 3.agegrp 4.agegrp 5.agegrp 6.agegrp bmi weight

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     Female  |  -.0736384   .0121553    -6.06   0.000    -.0974624   -.0498144
             |
      agegrp |
      30-39  |    .081033   .0136612     5.93   0.000     .0542574    .1078085
      40-49  |   .1773893   .0154534    11.48   0.000     .1471012    .2076774
      50-59  |   .3134343   .0159017    19.71   0.000     .2822675    .3446011
      60-69  |   .3634504   .0125983    28.85   0.000     .3387581    .3881427
        70+  |   .4599028   .0171395    26.83   0.000       .42631    .4934956
             |
         bmi |   .0218737   .0021304    10.27   0.000     .0176982    .0260491
      weight |   .0016568   .0007261     2.28   0.022     .0002338    .0030799
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Contrast - overall

To test if the probability of having high blood pressure is the same in all age groups:

```
. margins agegrp, contrast

Contrasts of predictive margins
Model VCE    : OIM

Expression   : Pr(highbp), predict()

------------------------------------------------
             |         df        chi2     P>chi2
-------------+----------------------------------
      agegrp |          5     1305.03     0.0000
------------------------------------------------
```

Contrasts – comparing to base level

Apply contrast operator to get specific comparisons, for example - to compare each level to the base level of covariate 'agegrp' (and get the estimates with p-values):

```
. margins r.agegrp, contrast(nowald pveffects)

Contrasts of predictive margins
Model VCE     : OIM

Expression    : Pr(highbp), predict()

------------------------------------------------------------
                  |            Delta-method
                  |   Contrast   Std. Err.        z    P>|z|
------------------+-----------------------------------------
           agegrp |
(30-39 vs 20-29)  |    .081033    .0136612     5.93    0.000
(40-49 vs 20-29)  |   .1773893    .0154534    11.48    0.000
(50-59 vs 20-29)  |   .3134343    .0159017    19.71    0.000
(60-69 vs 20-29)  |   .3634504    .0125983    28.85    0.000
  (70+ vs 20-29)  |   .4599028    .0171395    26.83    0.000
------------------------------------------------------------
```

Contrasts – comparing to next level, or previous level

To get differences of predictive margins from the next level of 'agegrp':

```
. margins a.agegrp, contrast(nowald pveffects)

Contrasts of predictive margins
Model VCE    : OIM

Expression   : Pr(highbp), predict()

--------------------------------------------------------
                  |            Delta-method
                  |   Contrast   Std. Err.      z    P>|z|
------------------+-------------------------------------
           agegrp |
(20-29 vs 30-39)  |   -.081033   .0136612    -5.93   0.000
(30-39 vs 40-49)  |  -.0963563   .0166994    -5.77   0.000
(40-49 vs 50-59)  |   -.136045   .0184844    -7.36   0.000
(50-59 vs 60-69)  |  -.0500161   .0159939    -3.13   0.002
  (60-69 vs 70+)  |  -.0964524   .0169502    -5.69   0.000
--------------------------------------------------------
```

To get differences of predictive margins compared to the previous level:

```
. margins ar.agegrp, contrast(nowald pveffects)
```

Contrasts - pwcompare

To get pairwise comparisons across all levels of 'agegrp':

```
. margins agegrp, pwcompare

Pairwise comparisons of predictive margins
Model VCE    : OIM

Expression   : Pr(highbp), predict()

----------------------------------------------------------------
                |            Delta-method         Unadjusted
                |   Contrast   Std. Err.     [95% Conf. Interval]
----------------+-----------------------------------------------
         agegrp |
30-39 vs 20-29  |    .081033   .0136612      .0542574    .1078085
40-49 vs 20-29  |   .1773893   .0154534      .1471012    .2076774
50-59 vs 20-29  |   .3134343   .0159017      .2822675    .3446011
60-69 vs 20-29  |   .3634504   .0125983      .3387581    .3881427
  70+ vs 20-29  |   .4599028   .0171395        .42631    .4934956
40-49 vs 30-39  |   .0963563   .0166994      .0636262    .1290865
50-59 vs 30-39  |   .2324014   .0170872      .1989111    .2658916
60-69 vs 30-39  |   .2824175   .0140897      .2548021    .3100328
  70+ vs 30-39  |   .3788699   .0182794      .3430428    .4146969
50-59 vs 40-49  |    .136045   .0184844      .0998163    .1722738
60-69 vs 40-49  |   .1860611   .0157572      .1551775    .2169448
  70+ vs 40-49  |   .2825135   .0195961      .2441058    .3209212
60-69 vs 50-59  |   .0500161   .0159939      .0186686    .0813637
  70+ vs 50-59  |   .1464685   .0197407      .1077775    .1851595
  70+ vs 60-69  |   .0964524   .0169502      .0632306    .1296742
----------------------------------------------------------------
```

Note: Stata has contrast and pwcompare also as free-standing commands.

Contrasts – with at() option

To get average marginal effects of sex for each age category:

```
. margins sex, contrast(nowald pveffects) vsquish at(agegrp=(1(1)6))

Contrasts of predictive margins
Model VCE    : OIM

Expression   : Pr(highbp), predict()
1._at        : agegrp          =           1
2._at        : agegrp          =           2
3._at        : agegrp          =           3
4._at        : agegrp          =           4
5._at        : agegrp          =           5
6._at        : agegrp          =           6

------------------------------------------------------------
                    |            Delta-method
                    |   Contrast   Std. Err.       z    P>|z|
--------------------+---------------------------------------
            sex@_at |
(Female vs base) 1  |  -.0556457   .0094367    -5.90   0.000
(Female vs base) 2  |  -.0698708   .0117327    -5.96   0.000
(Female vs base) 3  |  -.0811991   .0135129    -6.01   0.000
(Female vs base) 4  |  -.0868295   .0143506    -6.05   0.000
(Female vs base) 5  |  -.0857883   .0142134    -6.04   0.000
(Female vs base) 6  |  -.0789143   .0132599    -5.95   0.000
------------------------------------------------------------
```

Contrasts – difference in difference

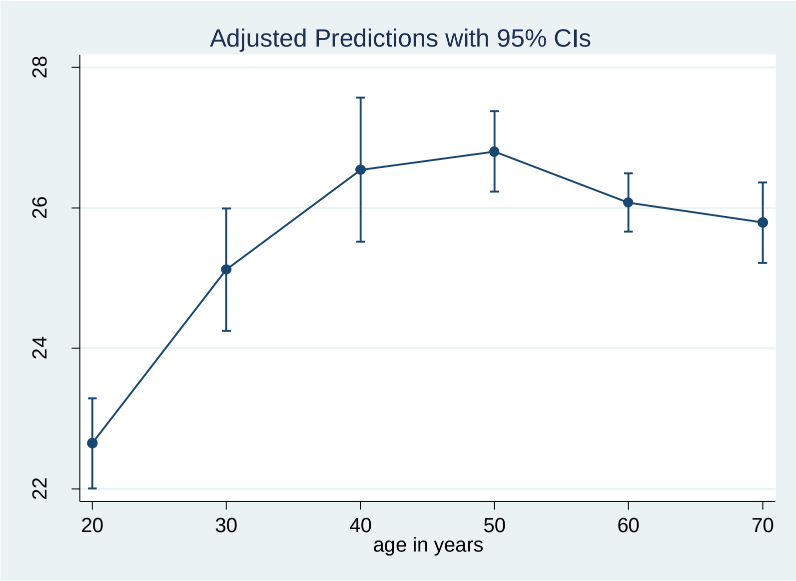
Even though our model does not have an interaction term, we can get differences of differences:

```
. margins sex#agegrp, contrast(nowald pveffects) vsquish

Contrasts of predictive margins
Model VCE    : OIM

Expression   : Pr(highbp), predict()

--------------------------------------------------------------------------
                                  |            Delta-method
                                  |   Contrast   Std. Err.      z    P>|z|
----------------------------------+---------------------------------------
                       sex#agegrp |
(Female vs base) (30-39 vs base)  |  -.0142251   .0033203    -4.28   0.000
(Female vs base) (40-49 vs base)  |  -.0255534   .0046694    -5.47   0.000
(Female vs base) (50-59 vs base)  |  -.0311838   .0053607    -5.82   0.000
(Female vs base) (60-69 vs base)  |  -.0301426   .0052348    -5.76   0.000
  (Female vs base) (70+ vs base)  |  -.0232686   .0045155    -5.15   0.000
--------------------------------------------------------------------------
```

The estimate -.0311 “says” that the marginal effects between female and male is smaller in group 1 (20-29 years old) than in group 4 (50-59 years old).

Margins and marginsplot to assess linearity

```
. webuse nhanes2
. svy: regress bmi i.age
. margins, at(age=(20(10)70)) vsquish
. marginsplot
```

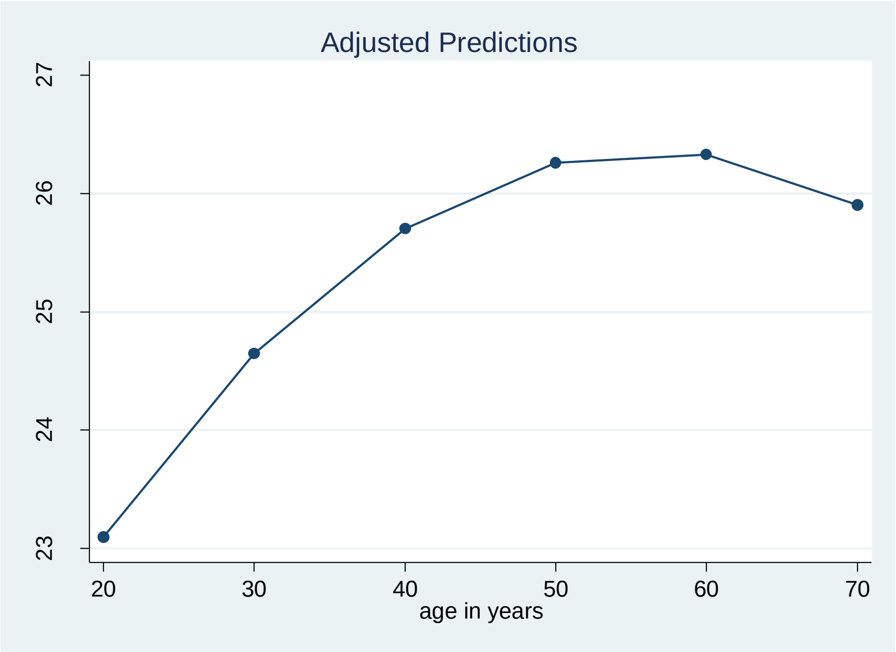


One way to use margins is to plot predicted margins at values of a continuous predictor and look whether linear model is appropriate. If the plot shows a curve (there is non-linearity) - introduce a quadratic term:

```
. svy: regress bmi age c.age#c.age
```

Margins and marginsplot to access linearity

```
. webuse nhanes2
. svy: regress bmi age c.age#c.age
(output ommited)

. margins, at(age=(20(10)70)) vsquish
. marginsplot, noci
```



The quadratic term in the model above is significant, and the plot shows that the quadratic equation seems to fit the data better.

Resources:

- **"Interpreting and visualizing regression models using Stata" by Michael Mitchell**
- **Stata Margins page: http://www.stata.com/manuals13/rmargins.pdf**
- **UCLA page: http://www.ats.ucla.edu/stat/stata/dae/predictive_margins.htm**
- **Richard Williams presentation (**and article http://www.statajournal.com/article.html?article=st0260**) : https://www3.nd.edu/~rwilliam/stats/Margins01.pdf**
- **On using margins and contour: http://www.stata.com/stata-news/news32-1/spotlight/?utm_source=statanews&utm_campaign=news32-1&utm_medium=email&utm_content=spotlight**
- **Online videos (on Stata channel): https://www.youtube.com/watch?v=XAG4CbIbH0k**
- **DSS presentation (**by Oscar Torres-Reyna) for ordered logit**: http://dss.princeton.edu/training/Margins.pdf**
- **http://www.stata.com/meeting/germany13/abstracts/materials/de13_jann.pdf**
- **For MEM vs AME see: http://www.michaelnormanmitchell.com/stow/marginal-effect-at-mean-vs-average-marginal-effect.html**

- Stata `help` command

Additional options:

.margins, asbalanced - treat all factor variables as balanced (groups of equal size)

To post the results from margins:

.margins, post −Estimates get stored in e() and could be used for further computations.

.margins, coeflegend −Useful if you need to access the coefficients posted (by the post option).

Questions?

Thank you.

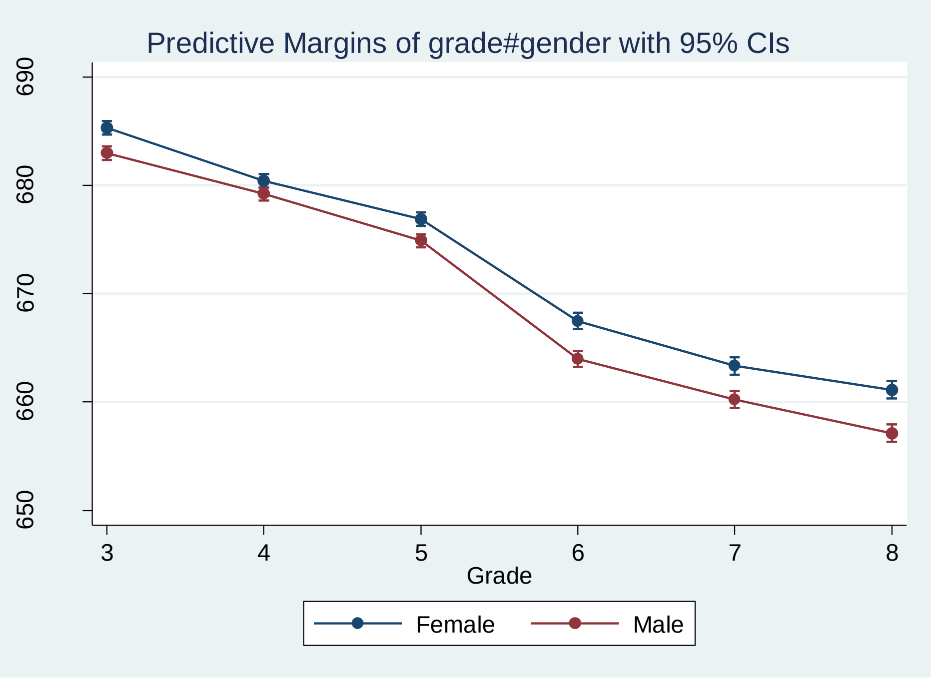
Marginsplot – model with interaction

```
. regress meanscore i.gender grade i.gender#grade year
(output omitted)
```

Not all of the interaction terms are significant.; some are positive, some are negative. What is going on? (Note that Stata treats grade as factor variable for the interaction, but as continuous for the main effect.)

To get predictive margins and plot:

```
. margins grade#gender
(output omitted)
. marginsplot
```



The meanscore is going down as grade goes up for both boys and girls. The scores for girls is higher than the score for boys in each grade. The differences seem to get bigger.

To get marginal effects (differences of predicted margins) of gender by grade:

```
. margins r.gender@grade, contrast(nowald pveffects)
(output omitted)
```

These are all significantly different from zero.