

# Fitting Models and Interpreting Results

Chuck Huber  
StataCorp  
[chuber@stata.com](mailto:chuber@stata.com)

Stata Webinar  
September 16, 2020

# Download Website

You can download the slides, datasets, and do-files here:

**<https://tinyurl.com/statareg>**

# Outline

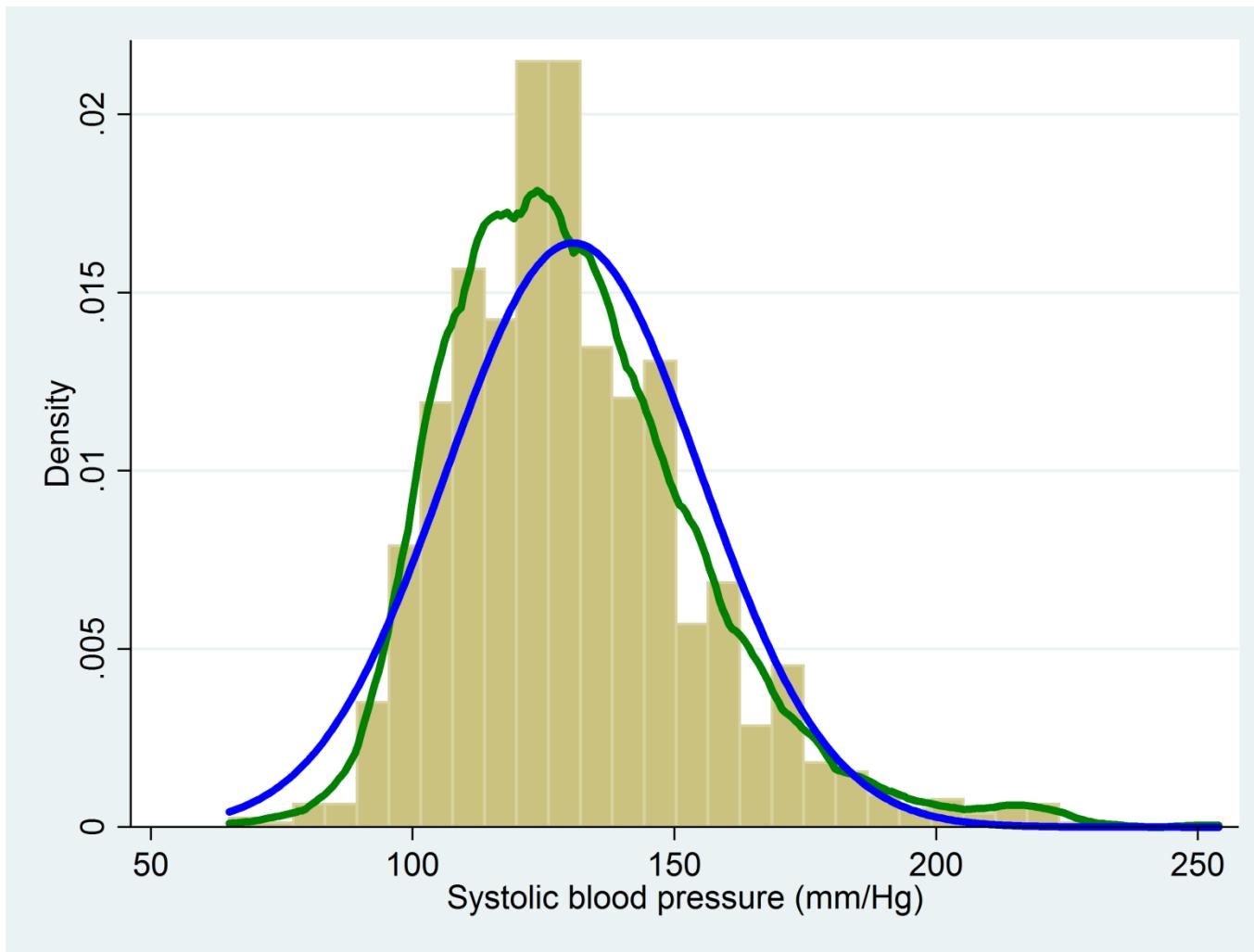
- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - Binary covariates
  - Categorical covariates
- Linear regression with multiple covariates
- Linear regression with interactions
- Linear regression with log-transformations

```
. use nhanes_clean.dta  
(Fictitious data based on the National Health and Nutrition Examination Survey)  
  
. describe id-diabetes dob bmi HighBP
```

| variable      | name | storage | display | value        | label                            |
|---------------|------|---------|---------|--------------|----------------------------------|
|               |      |         | format  | label        | variable label                   |
| id            |      | str6    | %9s     |              | Identification Number            |
| age           |      | byte    | %10.0g  |              | Age (years)                      |
| sex           |      | byte    | %9.0g   | SexLabel     | Sex                              |
| race          |      | byte    | %8.0g   | race_n       | Race                             |
| height        |      | float   | %9.0g   |              | height (cm)                      |
| weight        |      | float   | %9.0g   |              | weight (kg)                      |
| sbp           |      | int     | %9.0g   |              | Systolic blood pressure (mm/Hg)  |
| dbp           |      | int     | %9.0g   |              | Diastolic blood pressure (mm/Hg) |
| cholesterol   |      | int     | %9.0g   |              | serum cholesterol (mg/dL)        |
| triglycerides |      | int     | %9.0g   |              | serum triglycerides (mg/dL)      |
| hdl           |      | int     | %9.0g   |              | high density lipids (mg/dL)      |
| healthstatus  |      | byte    | %9.0g   | healthstatus | Overall health status            |
| heartattack   |      | byte    | %9.0g   | YesNo        | Previous heart attack            |
| diabetes      |      | byte    | %9.0g   | YesNo        | Diagnosed with diabetes          |
| dob           |      | float   | %td..   |              | Date of Birth                    |
| bmi           |      | float   | %5.1f   |              | Body Mass Index (kg/m^2)         |
| HighBP        |      | byte    | %12.0g  | HighBP       | Hypertension                     |

```
. summarize age-diabetes dob bmi HighBP
```

| Variable     | Obs   | Mean     | Std. Dev. | Min      | Max      |
|--------------|-------|----------|-----------|----------|----------|
| age          | 1,266 | 48.44076 | 16.98858  | 20       | 74       |
| sex          | 1,267 | .4664562 | .4990705  | 0        | 1        |
| race         | 1,267 | 2.704815 | .6973209  | 1        | 3        |
| height       | 1,266 | 167.1308 | 9.61487   | 144.199  | 193.398  |
| weight       | 1,266 | 72.17252 | 16.28185  | 39.12    | 175.88   |
|              |       |          |           |          |          |
| sbp          | 1,267 | 130.6843 | 24.34302  | 65       | 254      |
| dbp          | 1,267 | 80.54854 | 13.99187  | 35       | 150      |
| cholesterol  | 1,267 | 216.4917 | 46.86525  | 89       | 426      |
| triglyceri~s | 633   | 140.7962 | 89.34762  | 40       | 1191     |
| hdl          | 1,080 | 50.16019 | 14.35358  | 21       | 117      |
|              |       |          |           |          |          |
| healthstatus | 1,265 | 2.639526 | 1.229137  | 1        | 5        |
| heartattack  | 1,267 | .0528808 | .2238839  | 0        | 1        |
| diabetes     | 1,267 | .0591949 | .2360824  | 0        | 1        |
| dob          | 1,266 | 2561.307 | 6290.69   | -35161   | 13145    |
| bmi          | 1,265 | 25.78497 | 5.250611  | 15.36715 | 53.26417 |
|              |       |          |           |          |          |
| HighBP       | 1,267 | .6653512 | .4720539  | 0        | 1        |

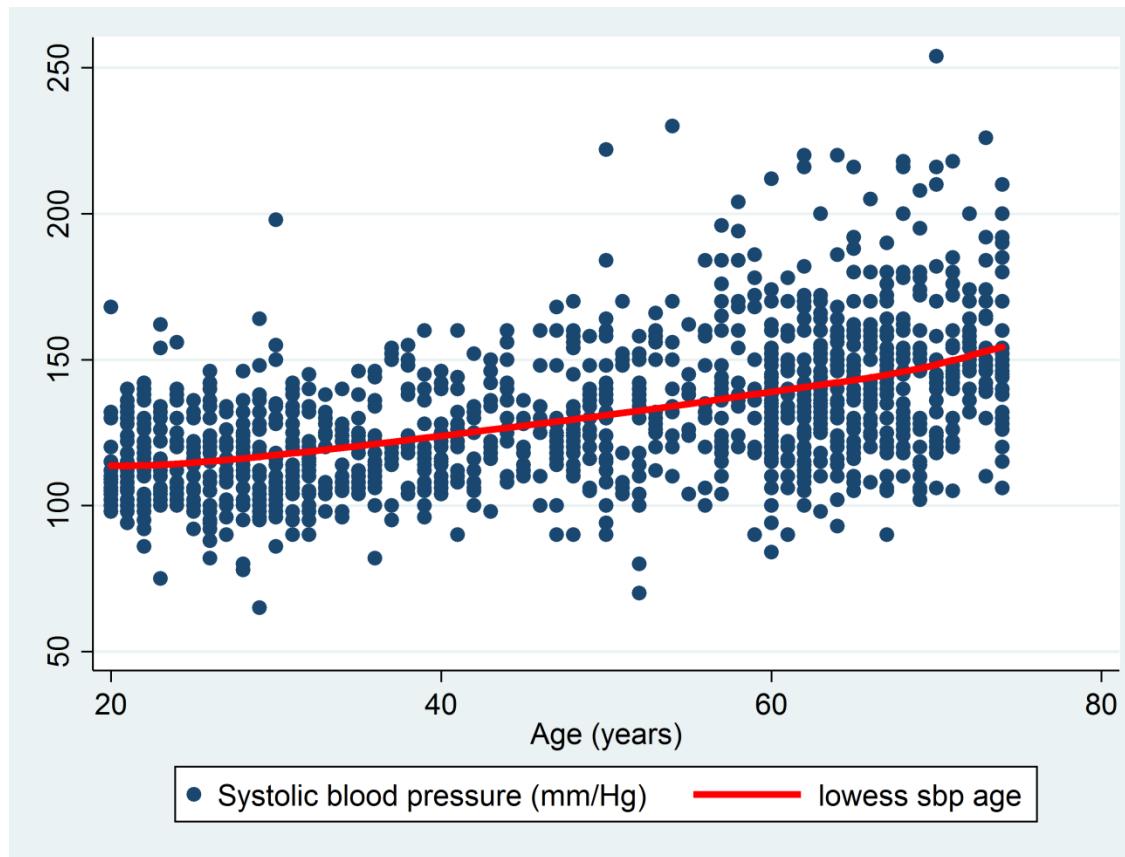


```
histogram sbp, normal normopts(lcolor(blue) lwidth(thick)) ///
kdensity kdenopts(lcolor(green) lwidth(thick))
```

# Outline

- The dataset
- Linear regression with one covariate
  - **Continuous covariates**
  - Binary covariates
  - Categorical covariates
- Linear regression with multiple covariates
- Linear regression with interactions
- Linear regression with log-transformations

# Scatterplot for SBP by Age



```
twoway (scatter sbp age) ///
(lowess sbp age, lcolor(red) lwidth(thick))
```

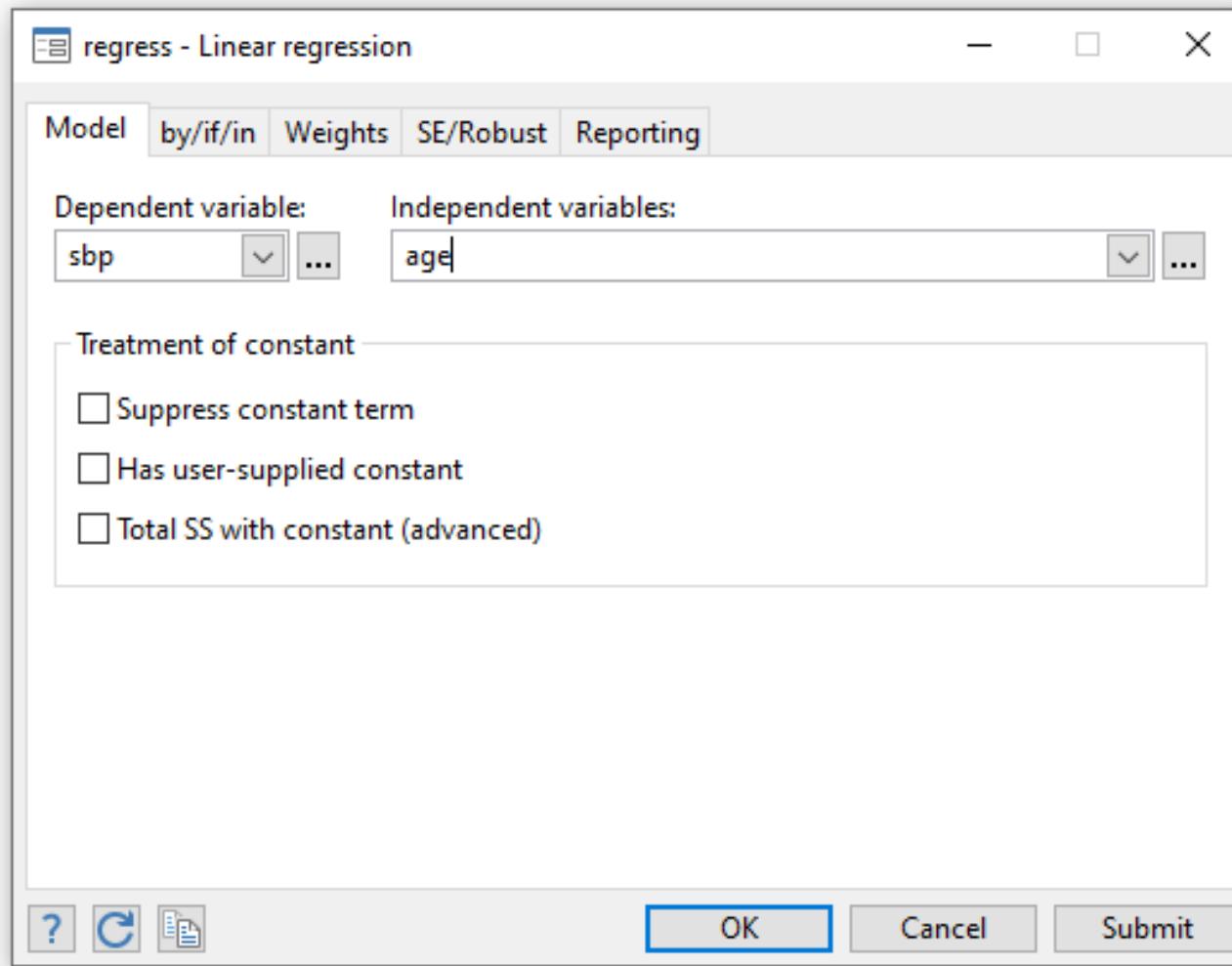
# Linear Regression

- Linear regression allows us to estimate the relationship between an outcome variable,  $y$ , and one or more predictor variables,  $x_i$ .

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$sbp_i = \beta_0 + \beta_1 age_i + \varepsilon_i$$

## Statistics &gt; Linear models and related &gt; Linear regression



# Linear Regression for Age

```
. regress sbp age
```

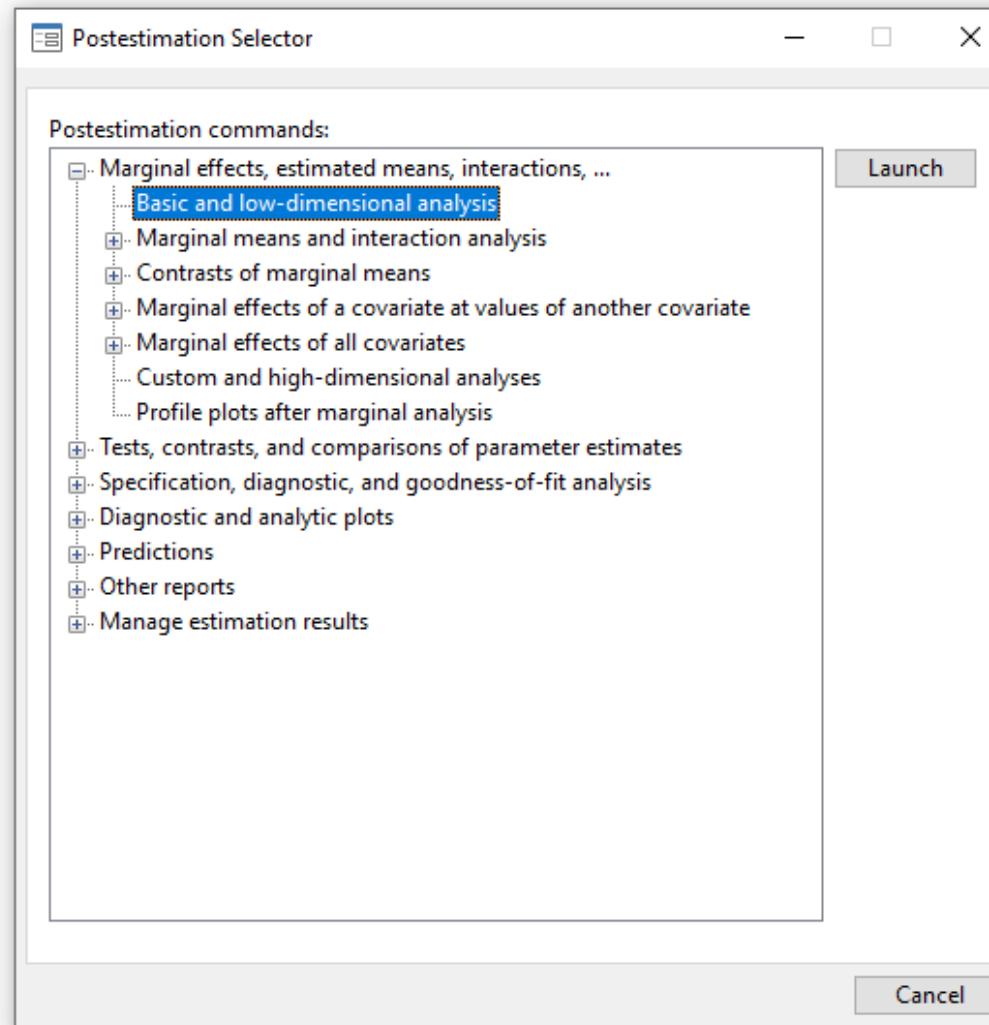
| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 195377.79  | 1     | 195377.79  | F(1, 1265)    | = | 445.46 |
| Residual | 554831.927 | 1,265 | 438.602314 | Prob > F      | = | 0.0000 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.2604 |
|          |            |       |            | Adj R-squared | = | 0.2598 |
|          |            |       |            | Root MSE      | = | 20.943 |

| sbp   | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|-------|----------|-----------|-------|-------|----------------------|----------|
| age   | .7315144 | .0346594  | 21.11 | 0.000 | .6635182             | .7995106 |
| _cons | 95.24655 | 1.779153  | 53.53 | 0.000 | 91.75613             | 98.73697 |

On average, SBP is 0.73 mm/Hg higher for each additional year of age.

# Statistics > Postestimation



# Margins

The image shows two Stata dialog boxes. On the left is the 'Postestimation Selector' dialog, which lists various postestimation commands under 'Postestimation commands'. A blue arrow points from the 'Launch' button in this dialog to the 'margins' dialog on the right. The 'margins' dialog is titled 'margins - Marginal means, predictive margins, and marginal effects'. It has two tabs: 'Main' (selected) and 'SE'. The 'Main' tab contains settings for computing margins: 'Covariate to compute margins, contrasts, or effects across', 'Type of covariate: Continuous', 'Covariate: age', 'Analysis type: Marginal means of outcome for levels of covariate', 'Contrast type: Differences from reference level', and 'Values to compute estimates at: 20(10)70'. Below these are sections for 'Interaction analysis with another covariate' (disabled), 'Draw profile plots of results' (disabled), 'Swap x-axis and plot dimensions' (disabled), 'Do not draw confidence intervals for point estimates' (disabled), and 'Plot options' (button). At the bottom, it asks 'How should other variables in the model be handled?' with three radio buttons: 'Population averaged (average estimates over estimation sample)' (selected), 'Compute estimates at average of estimation sample', and 'Treat factor variables as balanced -- equal weights to each level'. Buttons at the bottom include '?', 'OK', 'Cancel', and 'Submit'.

Postestimation Selector

Postestimation commands:

- Marginal effects, estimated means, interactions, ...
  - Basic and low-dimensional analysis
  - Marginal means and interaction analysis
  - Contrasts of marginal means
  - Marginal effects of a covariate at values of another covariate
  - Marginal effects of all covariates
  - Custom and high-dimensional analyses
  - Profile plots after marginal analysis
- Tests, contrasts, and comparisons of parameter estimates
- Specification, diagnostic, and goodness-of-fit analysis
- Diagnostic and analytic plots
- Predictions
- Other reports
- Manage estimation results

Launch

margins - Marginal means, predictive margins, and marginal effects

Main SE

Covariate to compute margins, contrasts, or effects across

Type of covariate: Continuous

Covariate: age

Analysis type: Marginal means of outcome for levels of covariate

Contrast type: Differences from reference level

Values to compute estimates at: 20(10)70

Interaction analysis with another covariate

Type of covariate: Categorical

Covariate:

Values to compute estimates at:

Draw profile plots of results

Swap x-axis and plot dimensions

Do not draw confidence intervals for point estimates

Plot options

How should other variables in the model be handled?

(●) Population averaged (average estimates over estimation sample)

(○) Compute estimates at average of estimation sample

(□) Treat factor variables as balanced -- equal weights to each level

Custom analysis

?

OK

Cancel

Submit

# Margins for Age

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

```
Adjusted predictions                               Number of obs     =      1,267
Model VCE    : OLS

Expression   : Linear prediction, predict()
1._at        : age          =       20
2._at        : age          =       30
3._at        : age          =       40
4._at        : age          =       50
5._at        : age          =       60
6._at        : age          =       70
```

|     | Delta-method |           |        |       |                      |          |
|-----|--------------|-----------|--------|-------|----------------------|----------|
|     | Margin       | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
| _at |              |           |        |       |                      |          |
| 1   | 109.8768     | 1.148086  | 95.70  | 0.000 | 107.6245             | 112.1292 |
| 2   | 117.192      | .8688152  | 134.89 | 0.000 | 115.4875             | 118.8965 |
| 3   | 124.5071     | .6571402  | 189.47 | 0.000 | 123.2179             | 125.7963 |
| 4   | 131.8223     | .5908308  | 223.11 | 0.000 | 130.6632             | 132.9814 |
| 5   | 139.1374     | .7117466  | 195.49 | 0.000 | 137.7411             | 140.5337 |
| 6   | 146.4526     | .9509679  | 154.00 | 0.000 | 144.5869             | 148.3182 |

Our regression model predicts that the average SBP is 109.9 mm/Hg at age 20, 117.192 at age 30, 124.5 at age 40, 131.8 at age 50, 139.1 at age 60 and 146.5 at age 70.

# Marginsplot

The diagram illustrates the process of generating a marginsplot. It begins with the "Postestimation Selector" dialog box, which contains a list of postestimation commands. A blue arrow points from the "Launch" button in the Postestimation Selector to the "marginsplot - Graph results from margins" dialog box.

**Postestimation Selector Dialog:**

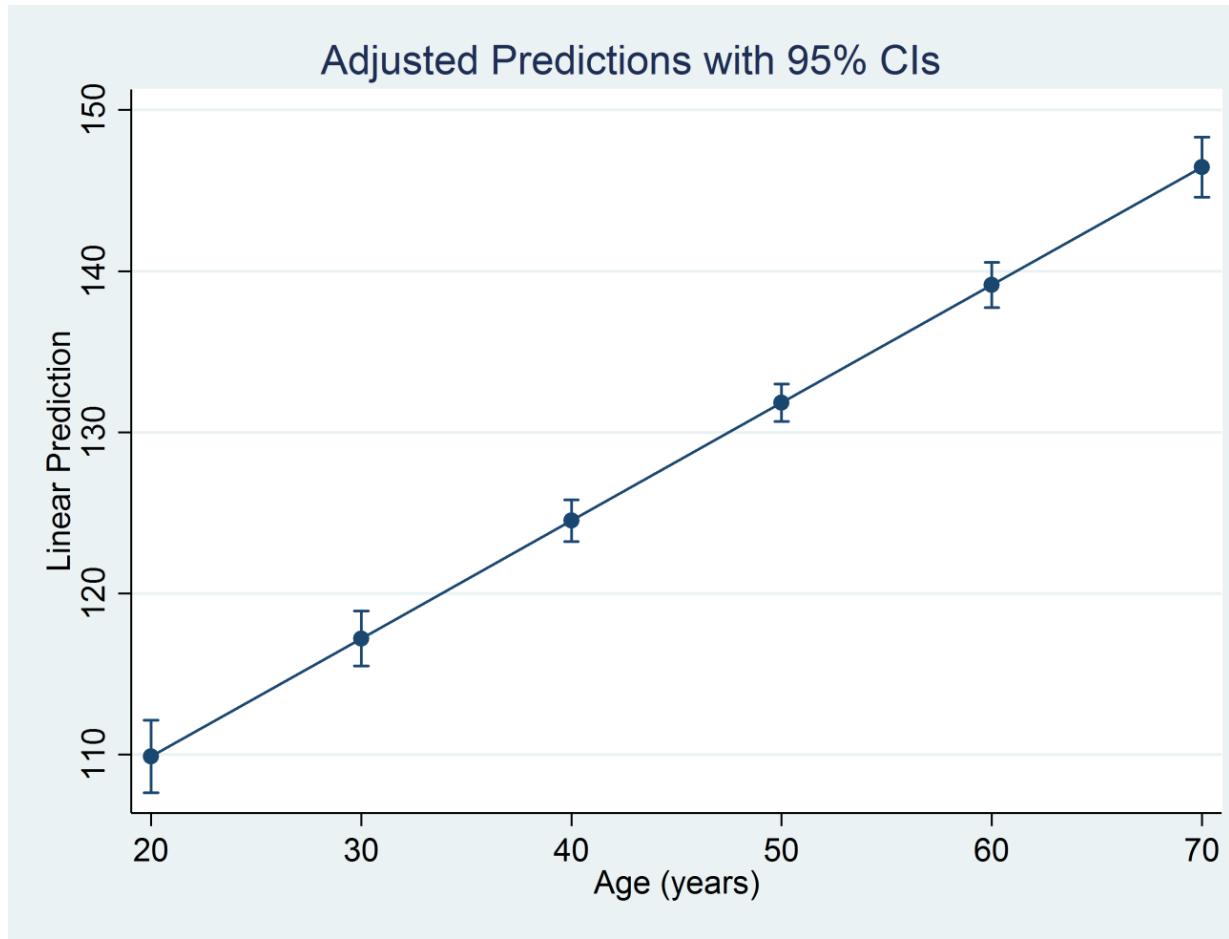
- Postestimation commands:
  - Marginal effects, estimated means, interactions, ...
    - Basic and low-dimensional analysis
    - Marginal means and interaction analysis
    - Contrasts of marginal means
    - Marginal effects of a covariate at values of another covariate
    - Marginal effects of all covariates
    - Custom and high-dimensional analyses
  - Profile plots after marginal analysis
  - Tests, contrasts, and comparisons of parameter estimates
  - Specification, diagnostic, and goodness-of-fit analysis
  - Diagnostic and analytic plots
  - Predictions
  - Other reports
  - Manage estimation results

**marginsplot - Graph results from margins Dialog:**

Main tab selected. The dialog includes the following sections:

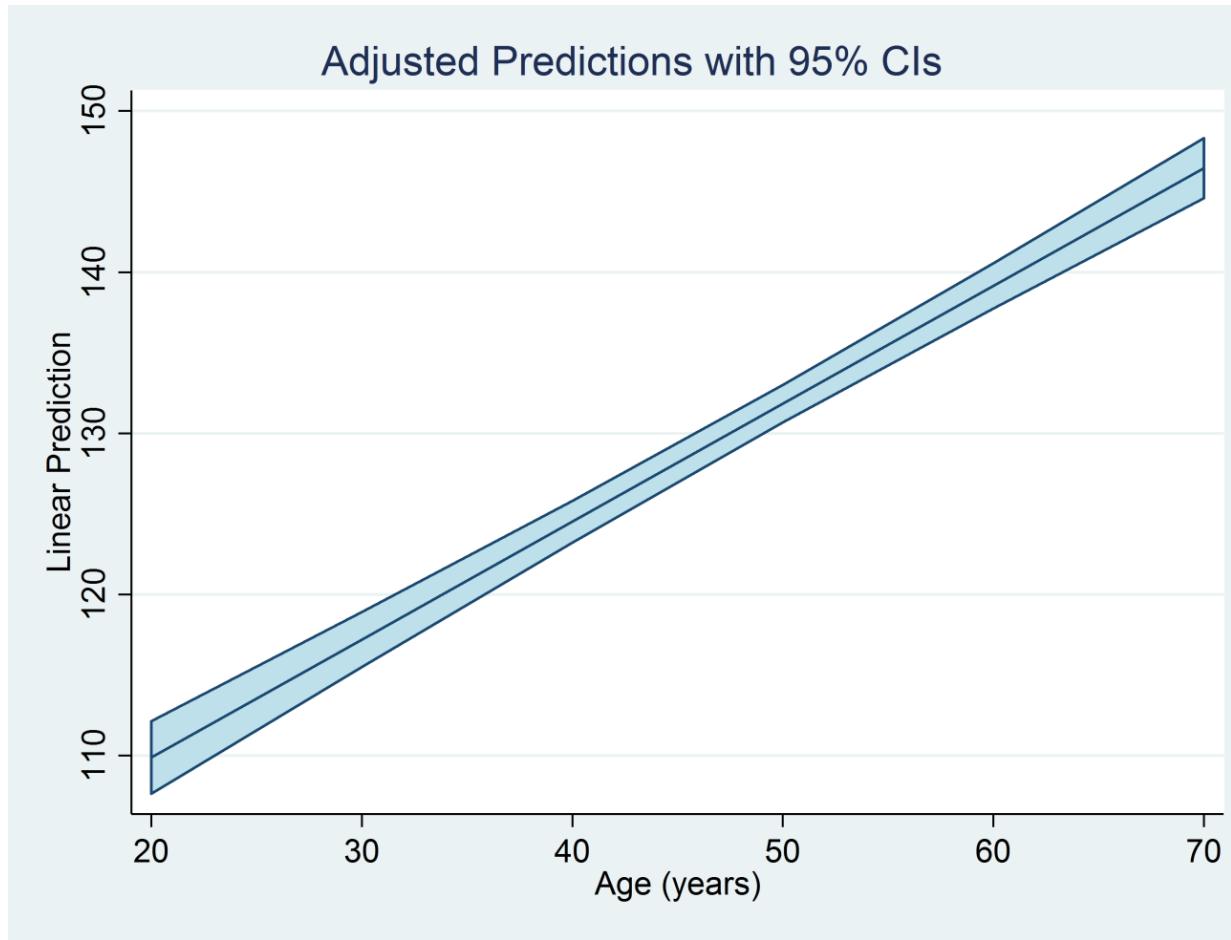
  - Dimensions**: Variable(s) that define the x axis:  Options
  - Create plots for groups defined by variables:**  Options
  - Create subgraphs for groups defined by variables:**  Options
  - Create graphs for groups defined by variables:**  Options
  - Checkboxes (bottom left):**
    - Swap the x and y axes (horizontal graph)
    - Do not plot confidence intervals
    - Use labels attached to marginal-effects variables
  - Checkboxes (bottom right):**
    - Name of graph, or stub if multiple graphs
    - Replace graph if it already exists in memory
  - Buttons (bottom):** ? C

# Marginsplot for Age



marginsplot

# Marginsplot for Age

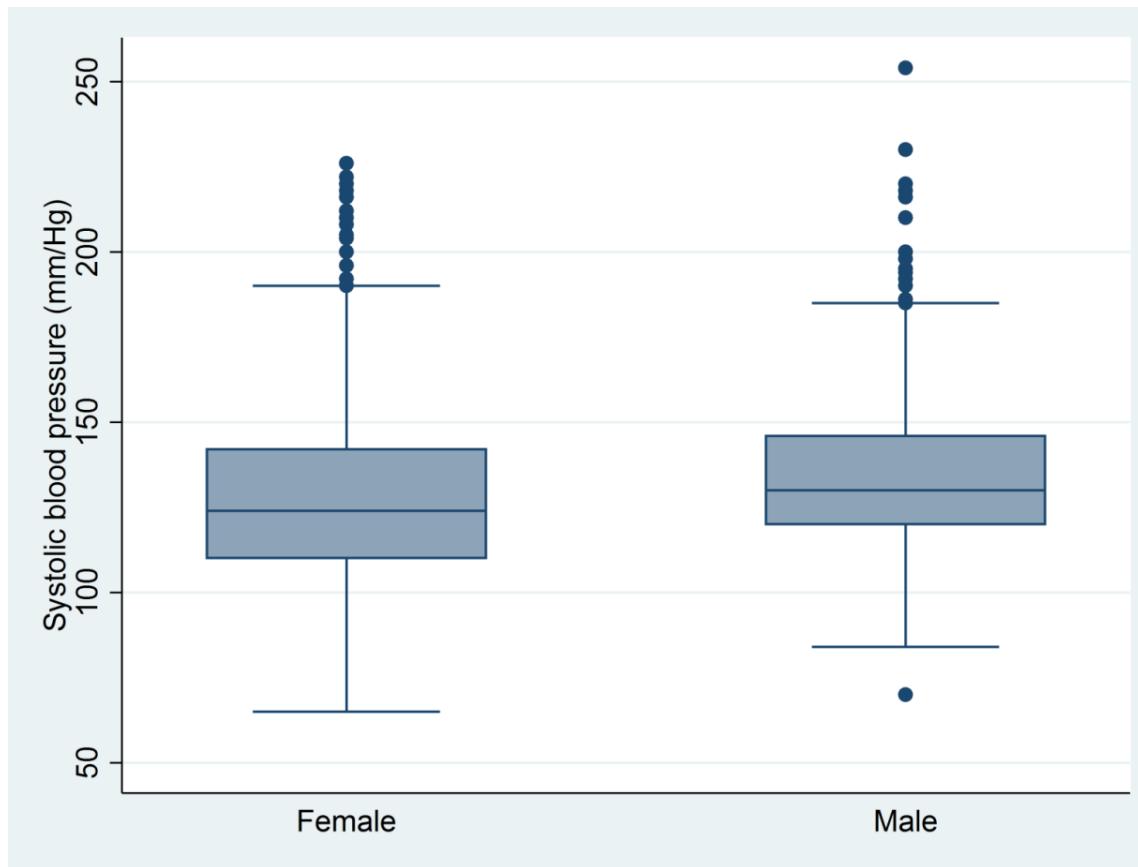


```
marginsplot, recast(line) recastci(rarea) ciopts(fcolor(ltblue))
```

# Outline

- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - **Binary covariates**
  - Categorical covariates
- Linear regression with multiple covariates
- Linear regression with interactions
- Linear regression with log-transformations

# Boxplots for SBP by Sex



```
graph box sbp, over(sex)
```

# Linear Regression for SBP and Sex

```
. regress sbp sex
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 8124.53576 | 1     | 8124.53576 | F(1, 1265)    | = | 13.85  |
| Residual | 742085.182 | 1,265 | 586.628602 | Prob > F      | = | 0.0002 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.0108 |
|          |            |       |            | Adj R-squared | = | 0.0100 |
|          |            |       |            | Root MSE      | = | 24.22  |

|       | Coef.    | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
|-------|----------|-----------|--------|-------|----------------------|----------|
| sex   | 5.075987 | 1.363964  | 3.72   | 0.000 | 2.400107             | 7.751867 |
| _cons | 128.3166 | .9315545  | 137.74 | 0.000 | 126.489              | 130.1441 |

Females are coded “0” and males are coded “1” so the average sbp for females is 128.3 mm/Hg and the sbp for males is 5.1 mm/Hg higher than females.

# Factor Variable Notation

- Stata assumes that covariates are continuous unless you tell it otherwise
- We can use the “**i.**” prefix to tell Stata that a covariate is a categorical variable
  - regress sbp i.sex
- We can use the “**c.**” prefix to tell Stata explicitly that a covariate is continuous
  - regress sbp c.age

# Factor Variable Notation

```
. list sex i.sex in 1/10
```

|     | sex    | 0b.<br>sex | 1.<br>sex |
|-----|--------|------------|-----------|
| 1.  | Male   | 0          | 1         |
| 2.  | Female | 0          | 0         |
| 3.  | Male   | 0          | 1         |
| 4.  | Female | 0          | 0         |
| 5.  | Male   | 0          | 1         |
| 6.  | Female | 0          | 0         |
| 7.  | Female | 0          | 0         |
| 8.  | Female | 0          | 0         |
| 9.  | Female | 0          | 0         |
| 10. | Male   | 0          | 1         |

The “i.” operator in front of the variable sex specifies that sex is a categorical variable. Stata will automatically create indicator variables for sex and label them in the output.

# Factor Variable Notation

```
. regress sbp i.sex
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 8124.53576 | 1     | 8124.53576 | F(1, 1265)    | = | 13.85  |
| Residual | 742085.182 | 1,265 | 586.628602 | Prob > F      | = | 0.0002 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.0108 |
|          |            |       |            | Adj R-squared | = | 0.0100 |
|          |            |       |            | Root MSE      | = | 24.22  |

| sbp   | Coef.    | Std. Err. | t      | P> t  | [95% Conf. Interval] |
|-------|----------|-----------|--------|-------|----------------------|
| sex   |          |           |        |       |                      |
| Male  | 5.075987 | 1.363964  | 3.72   | 0.000 | 2.400107             |
| _cons | 128.3166 | .9315545  | 137.74 | 0.000 | 126.489              |
|       |          |           |        |       | 7.751867             |
|       |          |           |        |       | 130.1441             |

The “i.” operator in front of the variable sex specifies that sex is a categorical variable. Stata will automatically create indicator variables for sex and label them in the output.

# The Margins Command

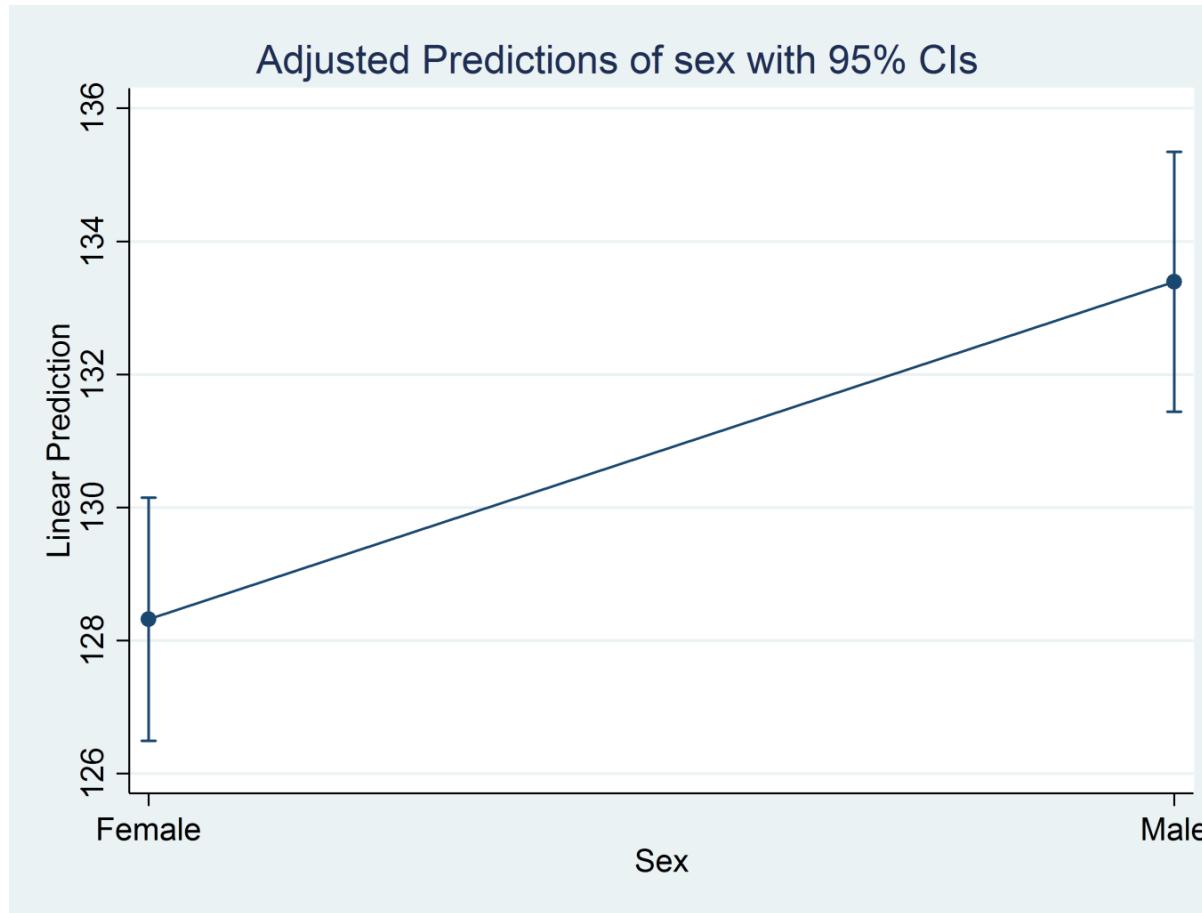
```
. margins sex
```

```
Adjusted predictions                               Number of obs     =      1,267
Model VCE    : OLS
Expression   : Linear prediction, predict()
```

|        | Delta-method |           |        |       |                      |          |
|--------|--------------|-----------|--------|-------|----------------------|----------|
|        | Margin       | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
| sex    |              |           |        |       |                      |          |
| Female | 128.3166     | .9315545  | 137.74 | 0.000 | 126.489              | 130.1441 |
| Male   | 133.3926     | .9962948  | 133.89 | 0.000 | 131.438              | 135.3471 |

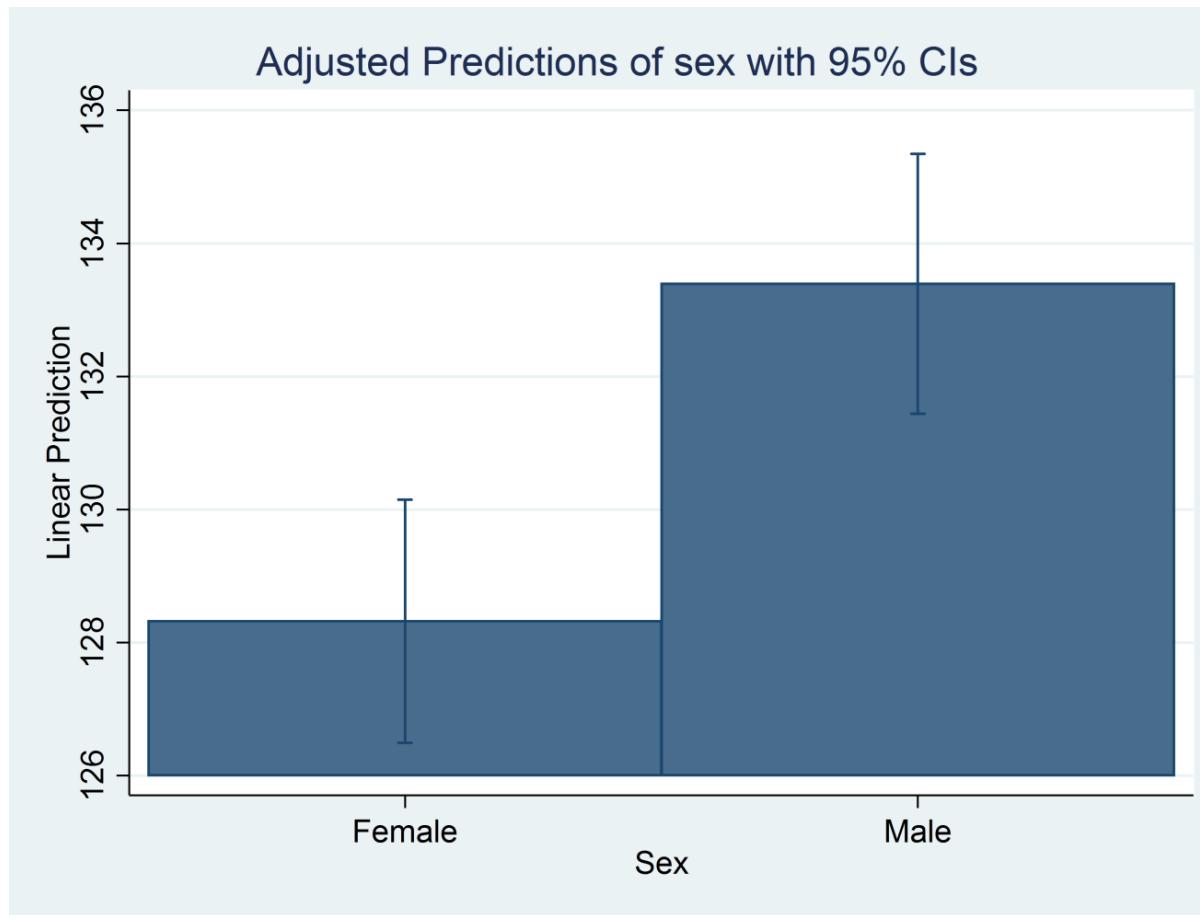
Our regression model predicts that the average sbp for females is 128.3 and the average sbp for males is 133.4.

# The Marginsplot Command



`marginsplot`

# The Marginsplot Command

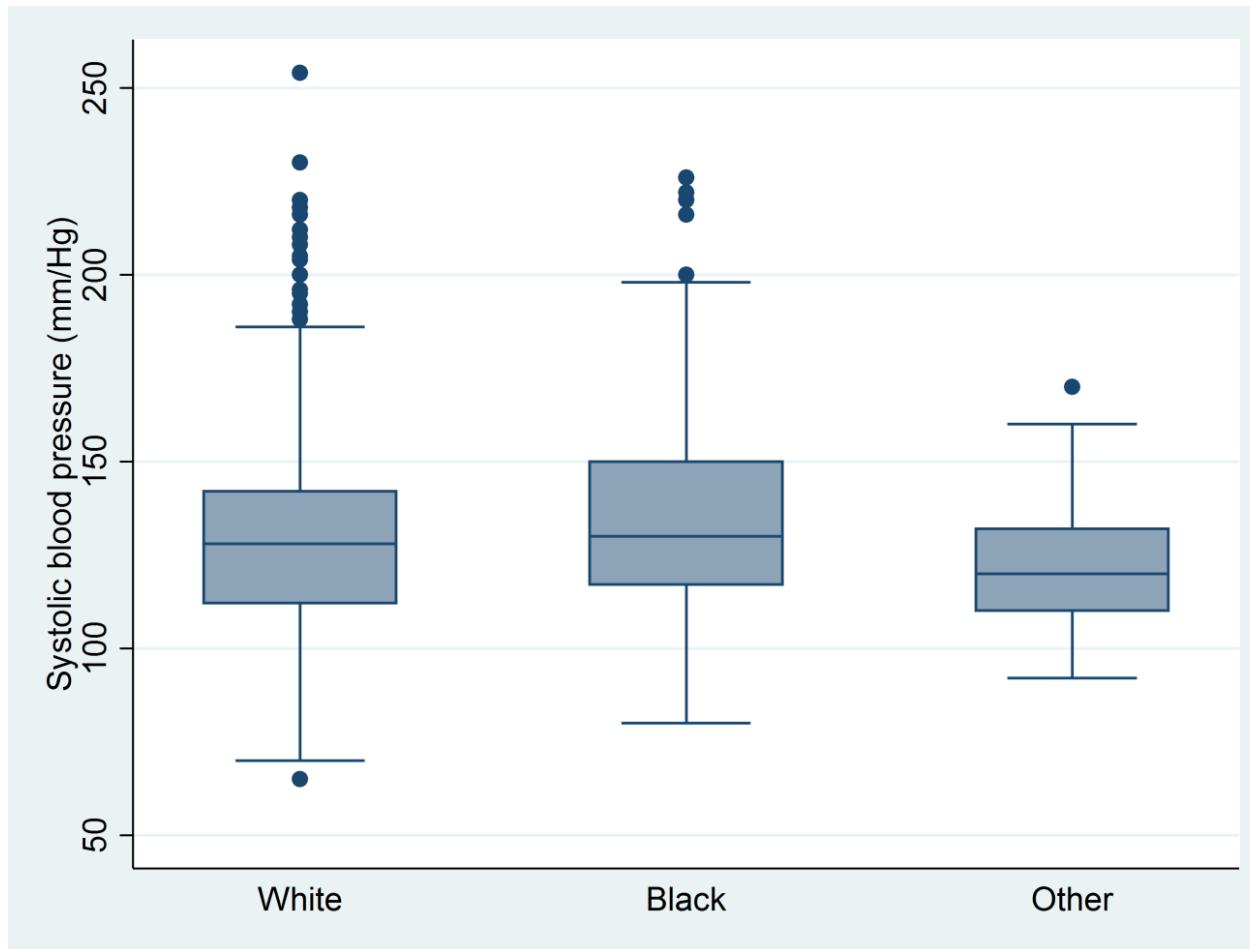


`marginsplot, recast(bar)`

# Outline

- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - Binary covariates
  - **Categorical covariates**
- Linear regression with multiple covariates
- Linear regression with interactions
- Linear regression with log-transformations

# Boxplot for Race



```
graph box sbp, over(race)
```

# Linear Regression for SBP and Race

```
. regress sbp i.race
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 4656.49666 | 2     | 2328.24833 | F(2, 1264)    | = | 3.95   |
| Residual | 745553.221 | 1,264 | 589.836409 | Prob > F      | = | 0.0195 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.0062 |
|          |            |       |            | Adj R-squared | = | 0.0046 |
|          |            |       |            | Root MSE      | = | 24.287 |

| sbp   | Coef.     | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
|-------|-----------|-----------|--------|-------|----------------------|----------|
| race  |           |           |        |       |                      |          |
| Black | 4.936873  | 1.975628  | 2.50   | 0.013 | 1.061001             | 8.812745 |
| Other | -6.011991 | 5.230919  | -1.15  | 0.251 | -16.27423            | 4.250248 |
| _cons | 130.1029  | .7428086  | 175.15 | 0.000 | 128.6456             | 131.5602 |

The average sbp for the referent category “White” is 130.1. The average sbp for the category “Black” is 4.9 mm/Hg higher than the category “White”. The average sbp for the category “Other” is 6.0 mm/Hg lower than the category “White”.

# Margins for Race

```
. margins race
```

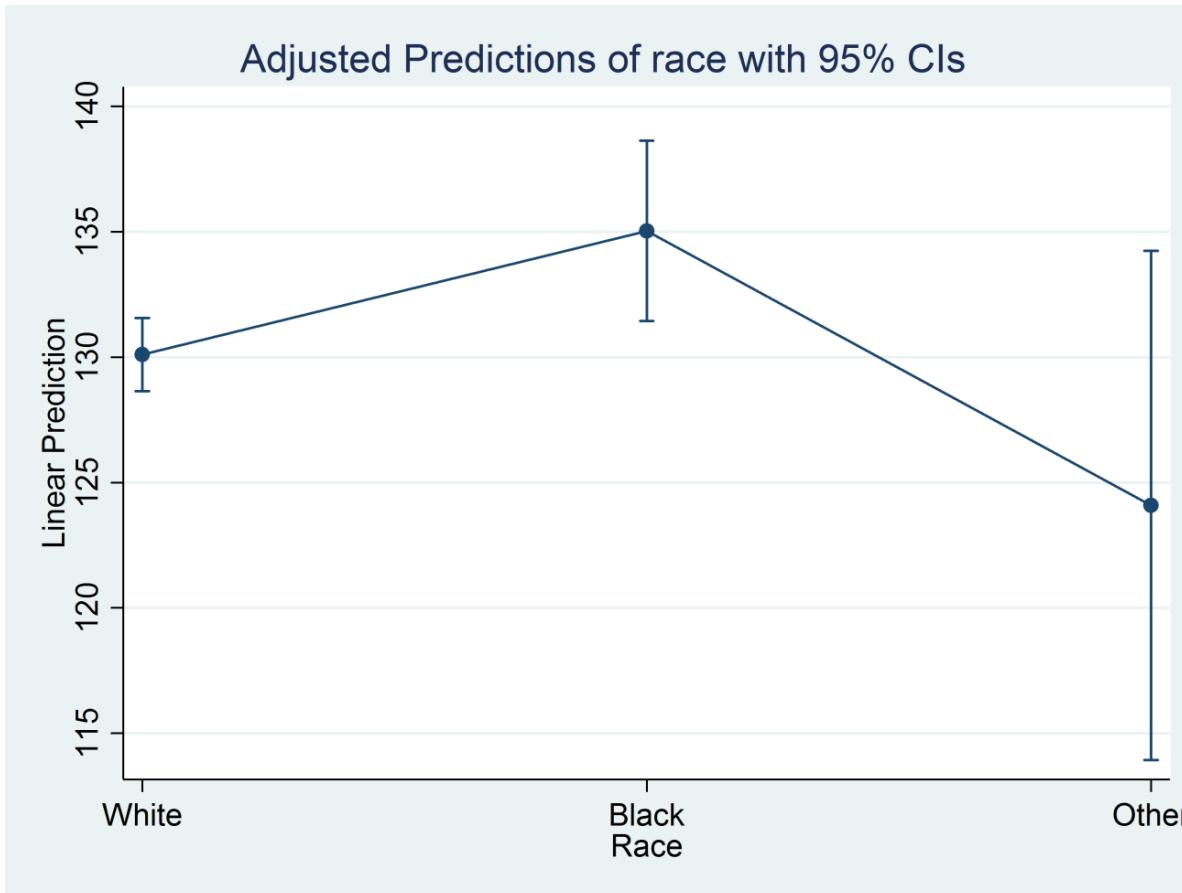
```
Adjusted predictions                               Number of obs     =      1,267
Model VCE    : OLS

Expression   : Linear prediction, predict()
```

|       | Delta-method |           |        |       |                      |          |
|-------|--------------|-----------|--------|-------|----------------------|----------|
|       | Margin       | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
| race  |              |           |        |       |                      |          |
| White | 130.1029     | .7428086  | 175.15 | 0.000 | 128.6456             | 131.5602 |
| Black | 135.0398     | 1.830667  | 73.77  | 0.000 | 131.4483             | 138.6313 |
| Other | 124.0909     | 5.177909  | 23.97  | 0.000 | 113.9327             | 134.2492 |

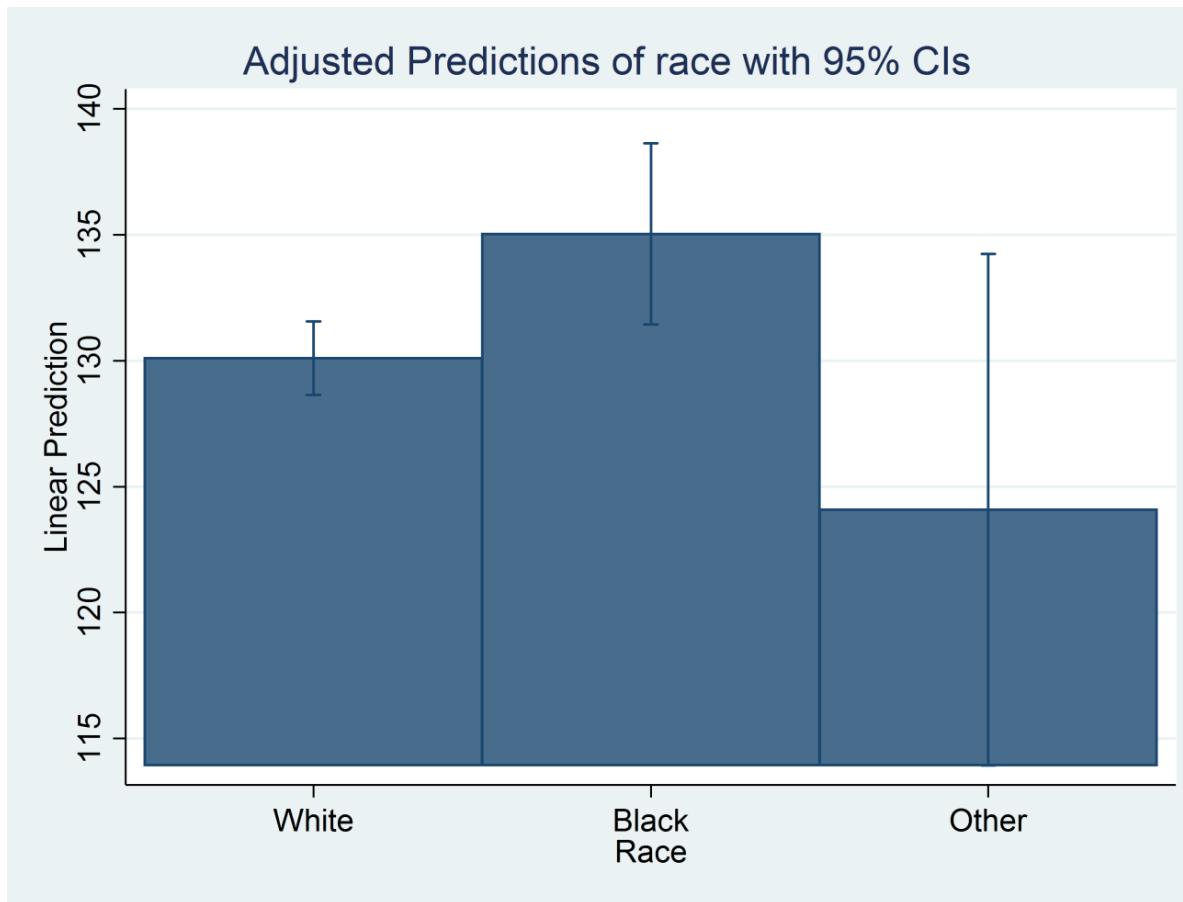
Our model predicts that the average sbp is 130.1 for the “White” category, 135.0 for the “Black” category, and 124.1 for the “Other” category.

# Marginsplot for Race



**marginsplot**

# Marginsplot for Race



`marginsplot, recast(bar)`

# Contrast Operators

| <i>operators (op.)</i> | Description   |
|------------------------|---|
| r.                     | differences from the reference (base) level; the default                          |
| a.                     | differences from the next level (adjacent contrasts)                              |
| ar.                    | differences from the previous level (reverse adjacent contrasts)                  |
| As-balanced operators  |   |
| g.                     | differences from the balanced grand mean  |
| h.                     | differences from the balanced mean of subsequent levels (Helmert contrasts)       |
| j.                     | differences from the balanced mean of previous levels (reverse Helmert contrasts) |
| p.                     | orthogonal polynomial in the level values   |
| q.                     | orthogonal polynomial in the level sequence                                       |
| As-observed operators  |   |
| gw.                    | differences from the observation-weighted grand mean                              |
| hw.                    | differences from the observation-weighted mean of subsequent levels               |
| jw.                    | differences from the observation-weighted mean of previous levels                 |
| pw.                    | observation-weighted orthogonal polynomial in the level values                    |
| qw.                    | observation-weighted orthogonal polynomial in the level sequence                  |

# Contrasts of Margins

```
. margins g.race, mcompare(scheffe) contrast(nowald)
```

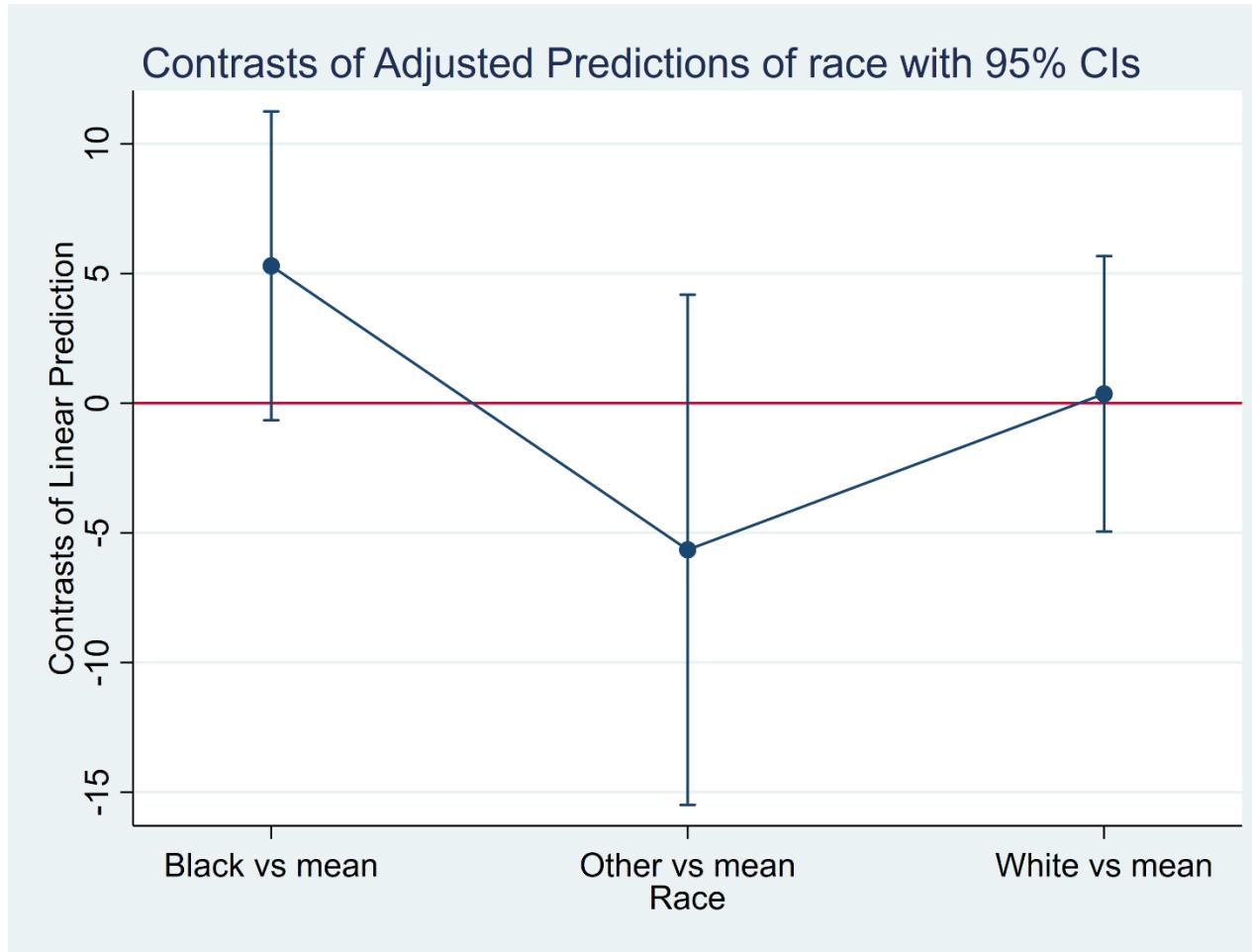
Contrasts of adjusted predictions Number of obs = 1,267  
Model VCE : OLS

Expression : Linear prediction, predict()

|      | Number of Comparisons |
|------|-----------------------|
| race | 3                     |

|                 | Delta-method |           | Scheffe              |          |
|-----------------|--------------|-----------|----------------------|----------|
|                 | Contrast     | Std. Err. | [95% Conf. Interval] |          |
| race            |              |           |                      |          |
| (Black vs mean) | 5.295245     | 2.128324  | -.6624773            | 11.25297 |
| (Other vs mean) | -5.653618    | 3.514195  | -15.49075            | 4.183509 |
| (White vs mean) | .3583727     | 1.896463  | -4.950311            | 5.667056 |

# Contrasts of Margins



```
marginsplot, yline(0) plotregion(margin(l=15 r=15))
```

# Pairwise Comparisons of Margins

```
. margins race, pwcompare mcompare(bonferroni)
```

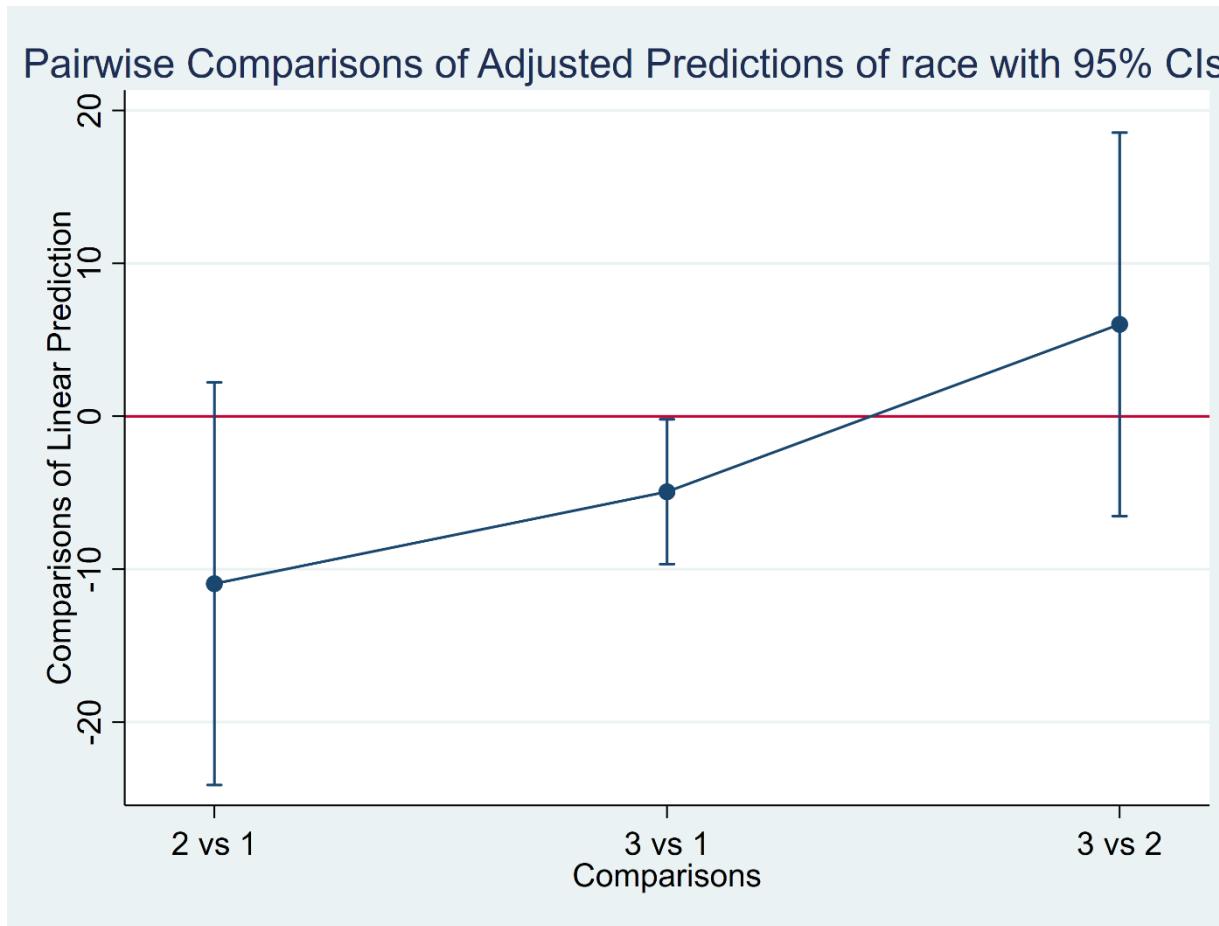
```
Pairwise comparisons of adjusted predictions      Number of obs      =      1,267
Model VCE      : OLS
```

```
Expression    : Linear prediction, predict()
```

|      | Number of Comparisons |
|------|-----------------------|
| race | 3                     |

|                | Contrast  | Delta-method<br>Std. Err. | Bonferroni<br>[95% Conf. Interval] |
|----------------|-----------|---------------------------|------------------------------------|
| race           |           |                           |                                    |
| Other vs Black | -10.94886 | 5.492002                  | -24.11413 2.216405                 |
| White vs Black | -4.936873 | 1.975628                  | -9.672792 -.2009535                |
| White vs Other | 6.011991  | 5.230919                  | -6.527416 18.5514                  |

# Pairwise Comparisons of Margins



```
 marginsplot, xdimension(_pw) unique yline(0) ///
    plotregion(margin(l=10 r=10))
```

# Outline

- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - Binary covariates
  - Categorical covariates
- **Linear regression with multiple covariates**
- Linear regression with interactions
- Linear regression with log-transformations

# Multiple Regression

```
. regress sbp i.sex i.race age
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 210817.929 | 4     | 52704.4823 | F(4, 1262)    | = | 123.31 |
| Residual | 539391.788 | 1,262 | 427.410292 | Prob > F      | = | 0.0000 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.2810 |
|          |            |       |            | Adj R-squared | = | 0.2787 |
|          |            |       |            | Root MSE      | = | 20.674 |

| sbp   | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|-------|----------|-----------|-------|-------|----------------------|----------|
| sex   |          |           |       |       |                      |          |
| Male  | 5.478602 | 1.16495   | 4.70  | 0.000 | 3.19315              | 7.764053 |
| race  |          |           |       |       |                      |          |
| Black | 6.560476 | 1.683872  | 3.90  | 0.000 | 3.25698              | 9.863972 |
| Other | .8661509 | 4.464017  | 0.19  | 0.846 | -7.891561            | 9.623863 |
| age   | .7381332 | .0343219  | 21.51 | 0.000 | .670799              | .8054674 |
| _cons | 91.44402 | 1.880863  | 48.62 | 0.000 | 87.75406             | 95.13398 |

# Margins for Age and Sex

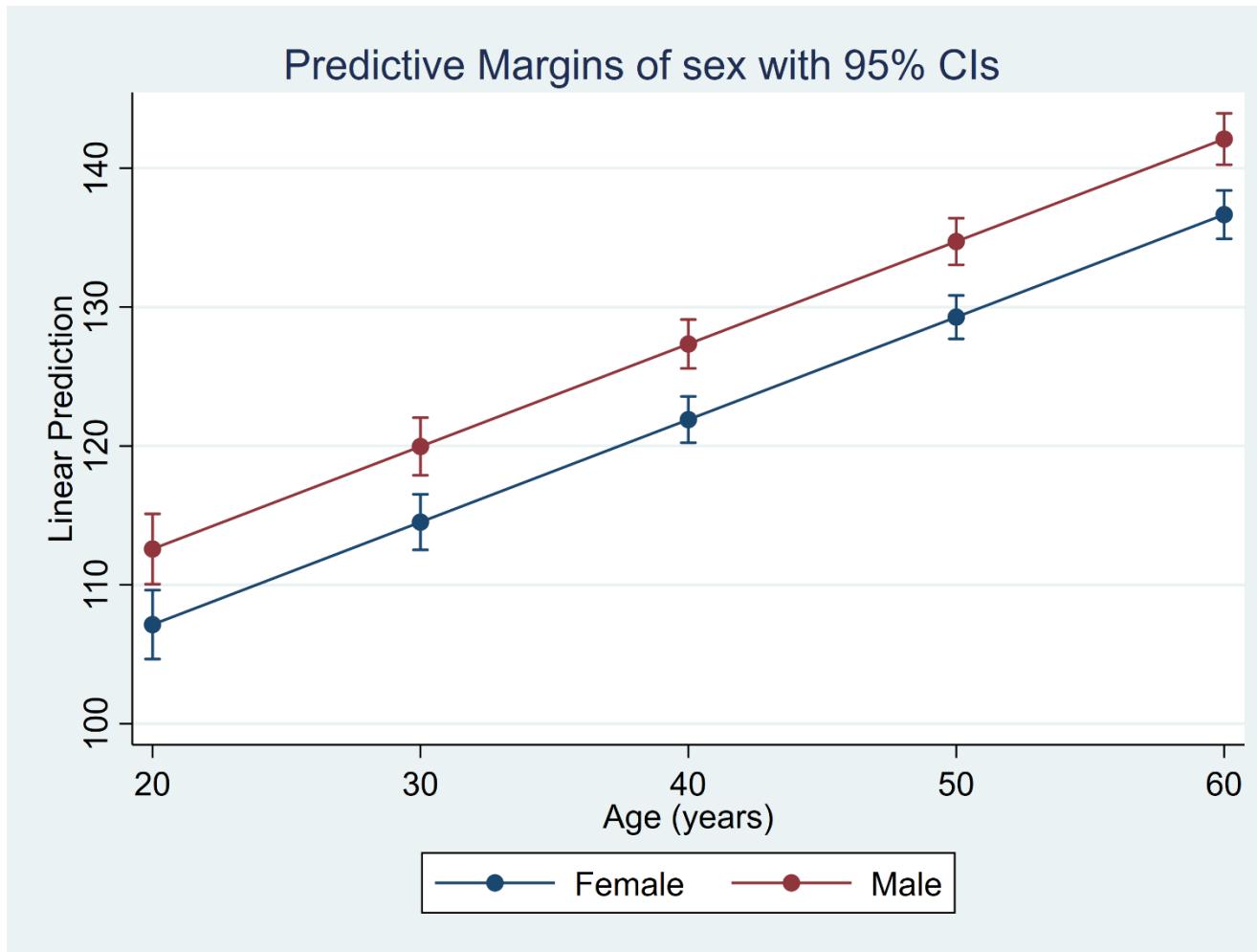
```
. margins sex, at(age=(20(10)60)) vsquish
```

| Predictive margins |                                | Number of obs | = | 1,266 |
|--------------------|--------------------------------|---------------|---|-------|
| Model VCE          | : OLS                          |               |   |       |
| Expression         | : Linear prediction, predict() |               |   |       |
| 1._at              | : age = 20                     |               |   |       |
| 2._at              | : age = 30                     |               |   |       |
| 3._at              | : age = 40                     |               |   |       |
| 4._at              | : age = 50                     |               |   |       |
| 5._at              | : age = 60                     |               |   |       |

|                | Delta-method |           |        |       |                      |          |
|----------------|--------------|-----------|--------|-------|----------------------|----------|
|                | Margin       | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
| <u>_at#sex</u> |              |           |        |       |                      |          |
| 1#Female       | 107.1411     | 1.264075  | 84.76  | 0.000 | 104.6612             | 109.621  |
| 1#Male         | 112.5828     | 1.289671  | 87.30  | 0.000 | 110.0527             | 115.113  |
| 2#Female       | 114.5198     | 1.020415  | 112.23 | 0.000 | 112.5179             | 116.5217 |
| 2#Male         | 119.9615     | 1.056442  | 113.55 | 0.000 | 117.889              | 122.0341 |
| 3#Female       | 121.8985     | .8486427  | 143.64 | 0.000 | 120.2336             | 123.5634 |
| 3#Male         | 127.3402     | .8969245  | 141.97 | 0.000 | 125.5806             | 129.0999 |
| 4#Female       | 129.2772     | .7966981  | 162.27 | 0.000 | 127.7142             | 130.8402 |
| 4#Male         | 134.7189     | .8535008  | 157.84 | 0.000 | 133.0445             | 136.3934 |
| 5#Female       | 136.6559     | .8859161  | 154.25 | 0.000 | 134.9179             | 138.3939 |
| 5#Male         | 142.0976     | .9423586  | 150.79 | 0.000 | 140.2489             | 143.9464 |

The predictions for each combination of age and sex are averaged over race.

# Marginsplot for Age and Sex



marginsplot

# Margins for Age and Race

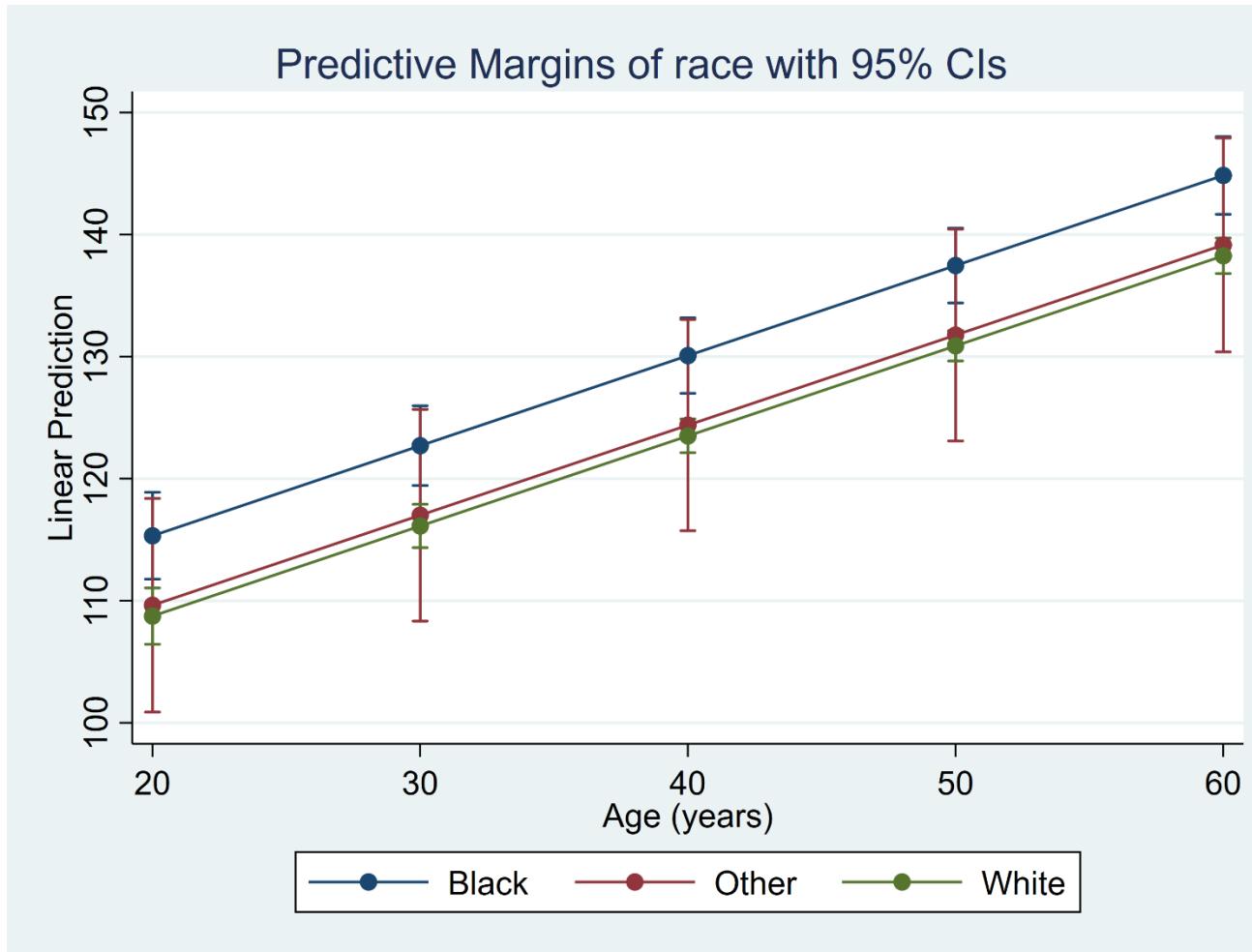
```
. margins race, at(age=(20(10)60)) vsquish
```

|                    |       |                              |   |       |
|--------------------|-------|------------------------------|---|-------|
| Predictive margins |       | Number of obs                | = | 1,266 |
| Model VCE          | : OLS |                              |   |       |
| Expression         | :     | Linear prediction, predict() |   |       |
| 1._at              | :     | age = 20                     |   |       |
| 2._at              | :     | age = 30                     |   |       |
| 3._at              | :     | age = 40                     |   |       |
| 4._at              | :     | age = 50                     |   |       |
| 5._at              | :     | age = 60                     |   |       |

|          | Delta-method |           |        |       |                      |          |
|----------|--------------|-----------|--------|-------|----------------------|----------|
|          | Margin       | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
| _at#race |              |           |        |       |                      |          |
| 1#Black  | 115.326      | 1.813503  | 63.59  | 0.000 | 111.7682             | 118.8838 |
| 1#Other  | 109.6308     | 4.458924  | 24.59  | 0.000 | 100.8831             | 118.3785 |
| 1#White  | 108.7472     | 1.175279  | 92.53  | 0.000 | 106.4415             | 111.0529 |
| 2#Black  | 122.7047     | 1.664516  | 73.72  | 0.000 | 119.4392             | 125.9702 |
| 2#Other  | 117.0095     | 4.419981  | 26.47  | 0.000 | 108.3382             | 125.6808 |
| 2#White  | 116.1259     | .9050882  | 128.30 | 0.000 | 114.3503             | 117.9015 |
| 3#Black  | 130.0834     | 1.577346  | 82.47  | 0.000 | 126.9889             | 133.1779 |
| 3#Other  | 124.3882     | 4.407498  | 28.22  | 0.000 | 115.7414             | 133.035  |
| 3#White  | 123.5046     | .7019079  | 175.96 | 0.000 | 122.1276             | 124.8816 |
| 4#Black  | 137.4621     | 1.562374  | 87.98  | 0.000 | 134.397              | 140.5273 |
| 4#Other  | 131.7669     | 4.421701  | 29.80  | 0.000 | 123.0922             | 140.4416 |
| 4#White  | 130.8833     | .6338382  | 206.49 | 0.000 | 129.6398             | 132.1268 |
| 5#Black  | 144.8408     | 1.621602  | 89.32  | 0.000 | 141.6595             | 148.0221 |
| 5#Other  | 139.1456     | 4.462334  | 31.18  | 0.000 | 130.3912             | 147.9    |
| 5#White  | 138.262      | .7391976  | 187.04 | 0.000 | 136.8118             | 139.7122 |

The predictions for each combination of age and race are averaged over sex.

# Marginsplot for Age and Race

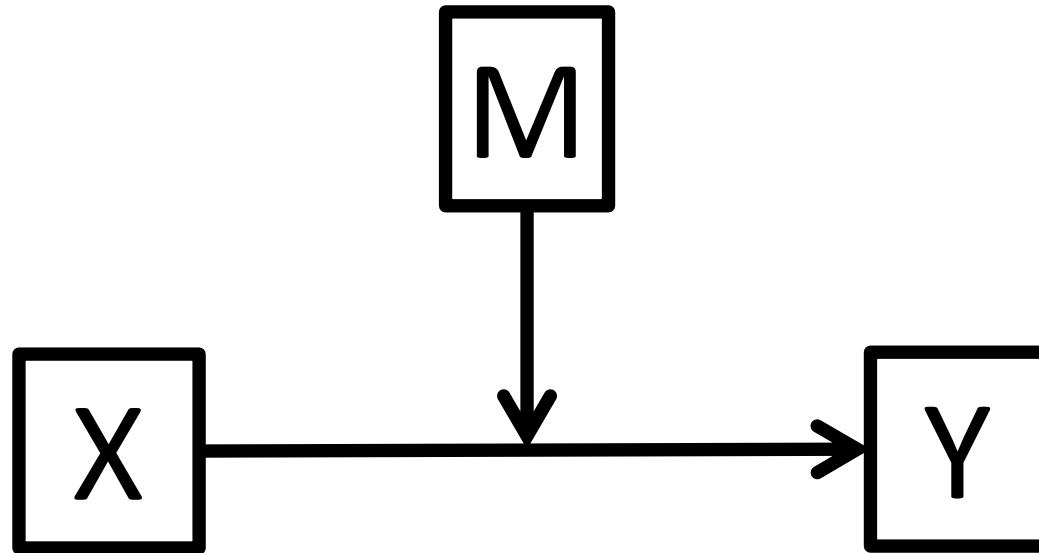


`marginsplot`

# Outline

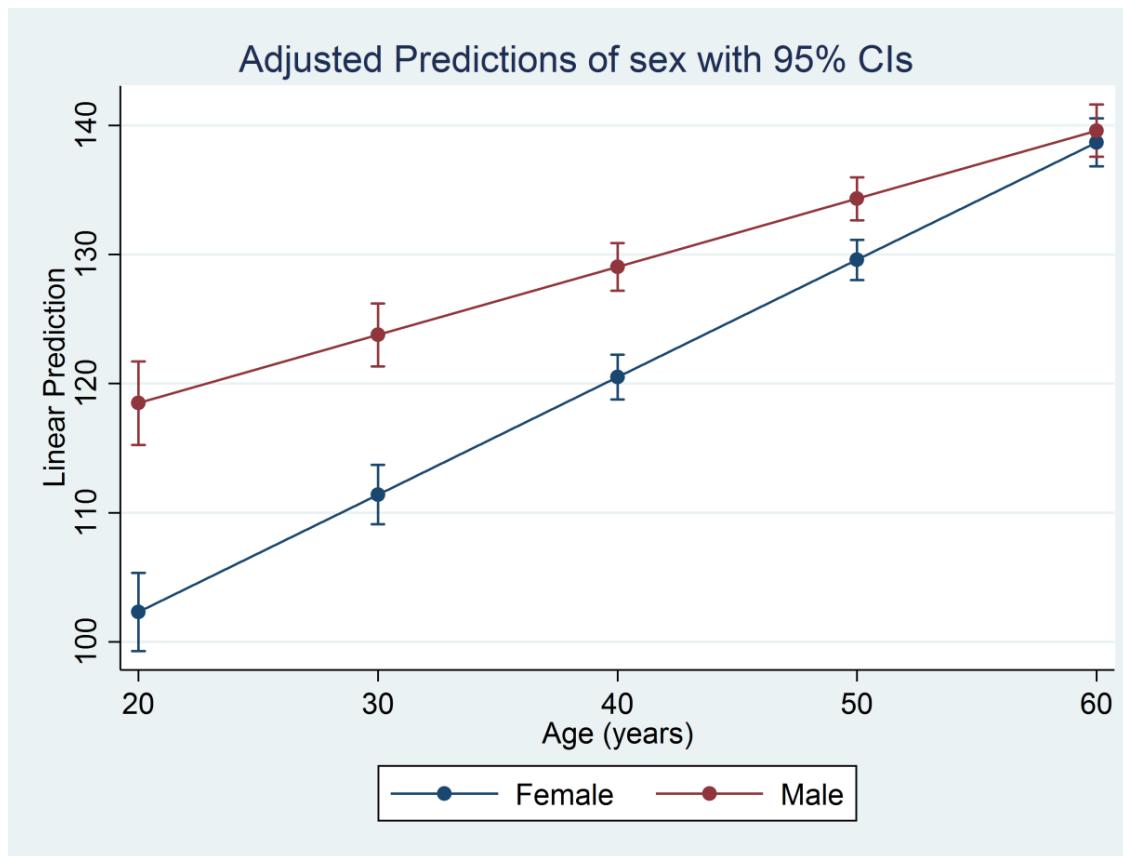
- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - Binary covariates
  - Categorical covariates
- Linear regression with multiple covariates
- **Linear regression with interactions**
- Linear regression with log-transformations

# What is Moderation?



“The effect of X on some variable Y is moderated by M if its size, sign, or strength depends on or can be predicted by M. In that case, M is said to be a moderator of X’s effect on Y, predicted by M.”  
(Hayes, 2013, pg 208).

# Example: Age (X), SBP (Y) and Sex (M)



The effect of age (X) on SBP (Y) is not the same for males and females.  
Thus sex is a moderator for the relationship between age and SBP.

# Factor Variable Notation

- We can use the “#” operator to create an interaction term for two covariates

```
regress sbp c.age i.sex c.age#i.sex
```

- Or we can use the “##” operator to include both main effects and the interaction term

```
regress sbp c.age##i.sex
```

# Factor Variable Notation

```
. list c.age i.sex c.age#i.sex in 1/5
```

|    | 0.<br>age | 1.<br>sex | 0.sex#<br>c.age | 1.sex#<br>c.age |
|----|-----------|-----------|-----------------|-----------------|
| 1. | 30        | 1         | 0               | 30              |
| 2. | 60        | 0         | 1               | 0               |
| 3. | 25        | 0         | 1               | 0               |
| 4. | 28        | 1         | 0               | 28              |
| 5. | 21        | 0         | 1               | 0               |
|    |           |           |                 | 21              |

```
. list c.age##i.sex in 1/5
```

|    | 0.<br>age | 1.<br>sex | 0.sex#<br>c.age | 1.sex#<br>c.age |
|----|-----------|-----------|-----------------|-----------------|
| 1. | 30        | 1         | 0               | 30              |
| 2. | 60        | 0         | 1               | 0               |
| 3. | 25        | 0         | 1               | 0               |
| 4. | 28        | 1         | 0               | 28              |
| 5. | 21        | 0         | 1               | 0               |
|    |           |           |                 | 21              |

# Regression for Age (X) and Sex (M)

```
. regress sbp c.age i.sex c.age#i.sex
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 217559.889 | 3     | 72519.9629 | F(3, 1263)    | = | 171.96 |
| Residual | 532649.829 | 1,263 | 421.733831 | Prob > F      | = | 0.0000 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.2900 |

|           | sbp | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-----------|-----|-----------|-----------|-------|-------|----------------------|
| age       |     | .9091169  | .0462905  | 19.64 | 0.000 | .8183022 .9999316    |
| sex       |     |           |           |       |       |                      |
| Male      |     | 23.81985  | 3.498187  | 6.81  | 0.000 | 16.95696 30.68275    |
| sex#c.age |     |           |           |       |       |                      |
| Male      |     | -.3819001 | .068185   | -5.60 | 0.000 | -.5156684 -.2481318  |
| _cons     |     | 84.12891  | 2.384564  | 35.28 | 0.000 | 79.45077 88.80705    |

# Regression for Age (X) and Sex (M)

```
. regress sbp c.age##i.sex
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,267  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 217559.889 | 3     | 72519.9629 | F(3, 1263)    | = | 171.96 |
| Residual | 532649.829 | 1,263 | 421.733831 | Prob > F      | = | 0.0000 |
| Total    | 750209.717 | 1,266 | 592.582715 | R-squared     | = | 0.2900 |

|           | sbp       | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-----------|-----------|----------|-----------|-------|-------|----------------------|
| age       | .9091169  | .0462905 |           | 19.64 | 0.000 | .8183022 .9999316    |
| sex       |           |          |           |       |       |                      |
| Male      | 23.81985  | 3.498187 |           | 6.81  | 0.000 | 16.95696 30.68275    |
| sex#c.age |           |          |           |       |       |                      |
| Male      | -.3819001 | .068185  |           | -5.60 | 0.000 | -.5156684 -.2481318  |
| _cons     | 84.12891  | 2.384564 |           | 35.28 | 0.000 | 79.45077 88.80705    |

# Margins for Age (X) and Sex (M)

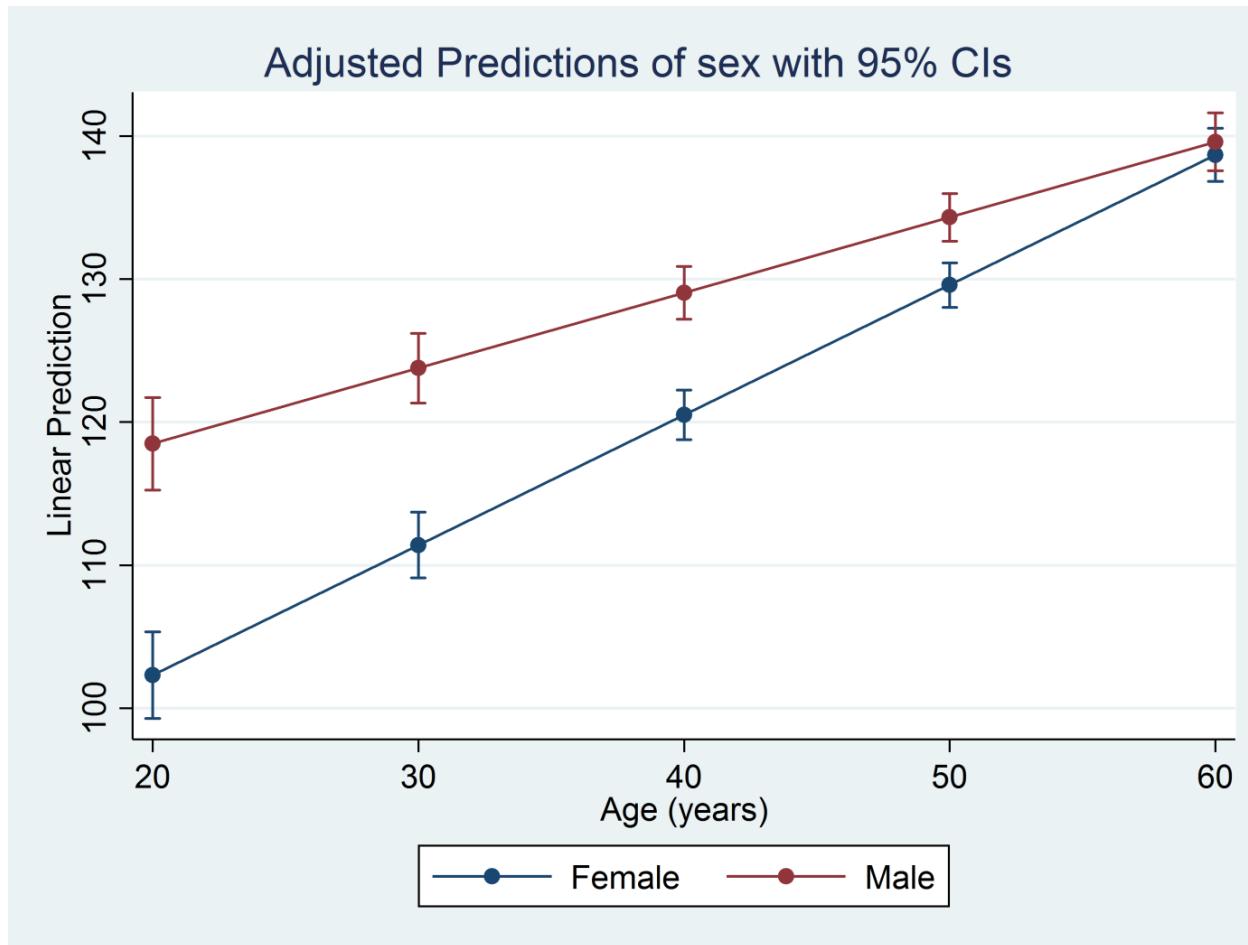
```
. margins sex, at(age=(20(10)60)) vsquish
```

```
Adjusted predictions                                         Number of obs      =      1,267
Model VCE       : OLS

Expression     : Linear prediction, predict()
1._at          : age            =        20
2._at          : age            =        30
3._at          : age            =        40
4._at          : age            =        50
5._at          : age            =        60
```

|               | Delta-method |           |        |       |                      |          |
|---------------|--------------|-----------|--------|-------|----------------------|----------|
|               | Margin       | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
| <u>at#sex</u> |              |           |        |       |                      |          |
| 1#Female      | 102.3113     | 1.541822  | 66.36  | 0.000 | 99.28644             | 105.3361 |
| 1#Male        | 118.4931     | 1.64783   | 71.91  | 0.000 | 115.2603             | 121.7259 |
| 2#Female      | 111.4024     | 1.168586  | 95.33  | 0.000 | 109.1098             | 113.695  |
| 2#Male        | 123.7653     | 1.244726  | 99.43  | 0.000 | 121.3233             | 126.2072 |
| 3#Female      | 120.4936     | .8846096  | 136.21 | 0.000 | 118.7581             | 122.2291 |
| 3#Male        | 129.0374     | .9405435  | 137.19 | 0.000 | 127.1922             | 130.8826 |
| 4#Female      | 129.5848     | .7924875  | 163.52 | 0.000 | 128.03               | 131.1395 |
| 4#Male        | 134.3096     | .8492214  | 158.16 | 0.000 | 132.6436             | 135.9756 |
| 5#Female      | 138.6759     | .9497897  | 146.01 | 0.000 | 136.8126             | 140.5393 |
| 5#Male        | 139.5818     | 1.029081  | 135.64 | 0.000 | 137.5629             | 141.6007 |

# Marginsplot for Age (X) and Sex (M)



marginsplot

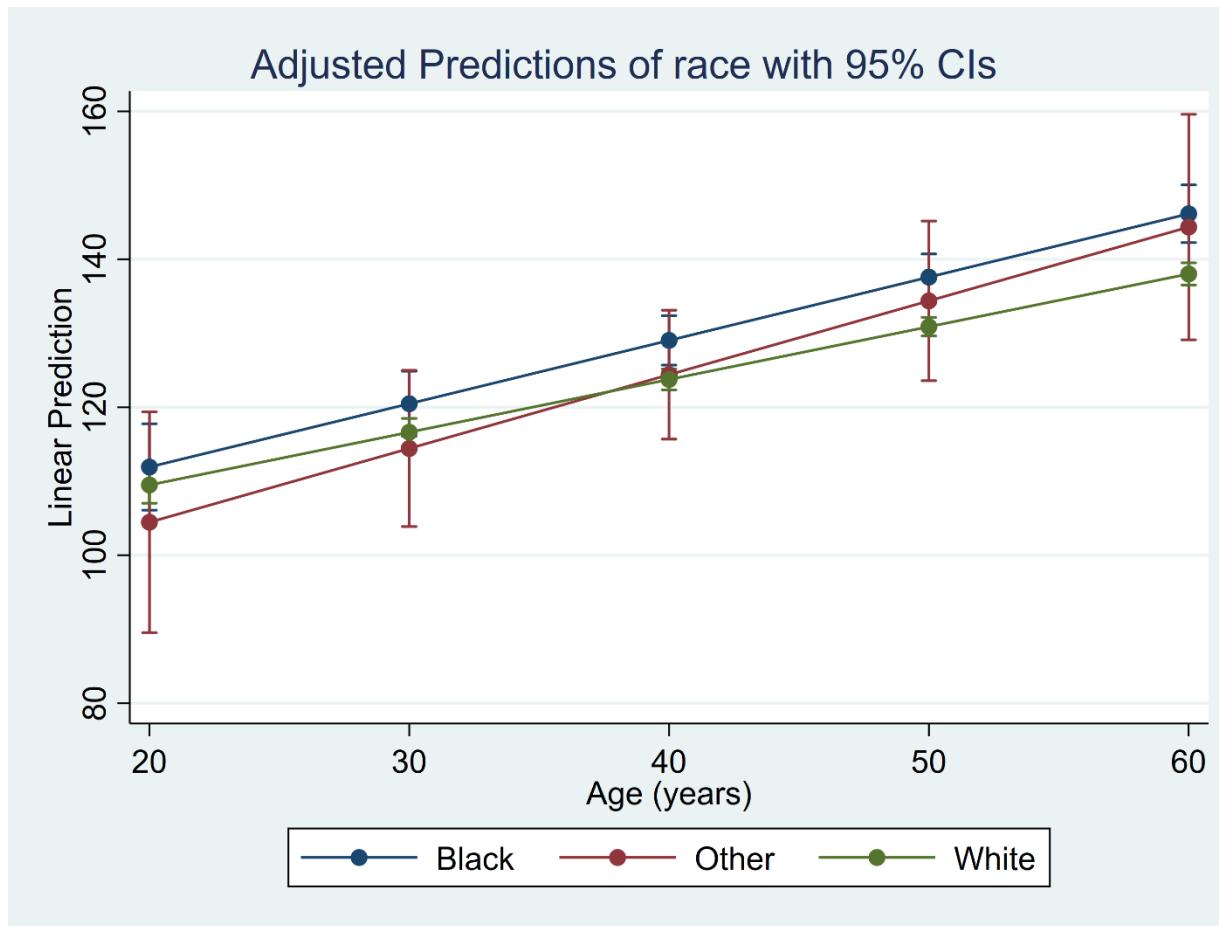
# Regression for Age (X) and Race (M)

. regress sbp c.age##i.race

| Source   | SS         | df    | MS         | Number of obs | = | 1,266  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 202409.031 | 5     | 40481.8062 | F(5, 1260)    | = | 93.24  |
| Residual | 547053.949 | 1,260 | 434.169801 | Prob > F      | = | 0.0000 |
| Total    | 749462.98  | 1,265 | 592.460854 | R-squared     | = | 0.2701 |

|            | sbp       | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|------------|-----------|----------|-----------|-------|-------|----------------------|
| age        | .8555534  | .0935956 |           | 9.14  | 0.000 | .671933 1.039174     |
| race       |           |          |           |       |       |                      |
| Other      | -10.3098  | 14.01752 |           | -0.74 | 0.462 | -37.81005 17.19044   |
| White      | .4203122  | 5.05856  |           | 0.08  | 0.934 | -9.503817 10.34444   |
| race#c.age |           |          |           |       |       |                      |
| Other      | .1419229  | .327337  |           | 0.43  | 0.665 | -.5002626 .7841084   |
| White      | -.1425008 | .1008253 |           | -1.41 | 0.158 | -.3403047 .0553031   |
| _cons      | 94.81904  | 4.671978 |           | 20.30 | 0.000 | 85.65333 103.9848    |

# Marginsplot for Age (X) and Race (M)



```
margins race, at(age=(20(10)60)) vsquish  
marginsplot, legend(rows(1))
```

# F-test for Interaction

```
. testparm c.age#i.race
```

```
( 1) 2.race#c.age = 0
```

```
( 2) 3.race#c.age = 0
```

```
F(  2,  1260) =      1.35
```

```
Prob > F =    0.2595
```

# Likelihood Ratio Test for Interaction

```
. quietly regress sbp c.age i.race c.age#i.race  
  
. estimates store w_int  
  
. quietly regress sbp c.age i.race  
  
. estimates store no_int  
  
. lrtest w_int no_int
```

Likelihood-ratio test  
(Assumption: no\_int nested in w\_int)

LR chi2(2) = 2.71  
Prob > chi2 = 0.2578

# Continuous-by-Continuous Interactions

```
. regress sbp c.age##c.weight
```

| Source   | SS         | df    | MS         | Number of obs | = | 1,265  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 262182.286 | 3     | 87394.0953 | F(3, 1261)    | = | 226.17 |
| Residual | 487252.185 | 1,261 | 386.401416 | Prob > F      | = | 0.0000 |
| Total    | 749434.471 | 1,264 | 592.907018 | R-squared     | = | 0.3498 |
|          |            |       |            | Adj R-squared | = | 0.3483 |
|          |            |       |            | Root MSE      | = | 19.657 |

| sbp            | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|----------------|-----------|-----------|-------|-------|----------------------|----------|
| age            | .9478535  | .1509158  | 6.28  | 0.000 | .6517798             | 1.243927 |
| weight         | .593942   | .1014167  | 5.86  | 0.000 | .394978              | .792906  |
| c.age#c.weight | -.0032773 | .0020846  | -1.57 | 0.116 | -.007367             | .0008124 |
| _cons          | 53.38597  | 7.331376  | 7.28  | 0.000 | 39.00294             | 67.76901 |

# Continuous-by-Continuous Interactions

```
. quietly margins, at(age=(20(5)60) weight=(40(5)160))      ///
>                               saving(predictions.dta, replace) vsquish

. use predictions.dta, clear
(Created by command margins; also see char list)

. describe _margin _at1 _at2
```

| variable | name | storage | display | value | variable                     | label |
|----------|------|---------|---------|-------|------------------------------|-------|
|          |      | type    | format  | label |                              |       |
| _margin  |      | float   | %9.0g   |       | Linear prediction, predict() |       |
| _at1     |      | byte    | %10.0g  |       | Age (years)                  |       |
| _at2     |      | int     | %9.0g   |       | weight (kg)                  |       |

# Continuous-by-Continuous Interactions

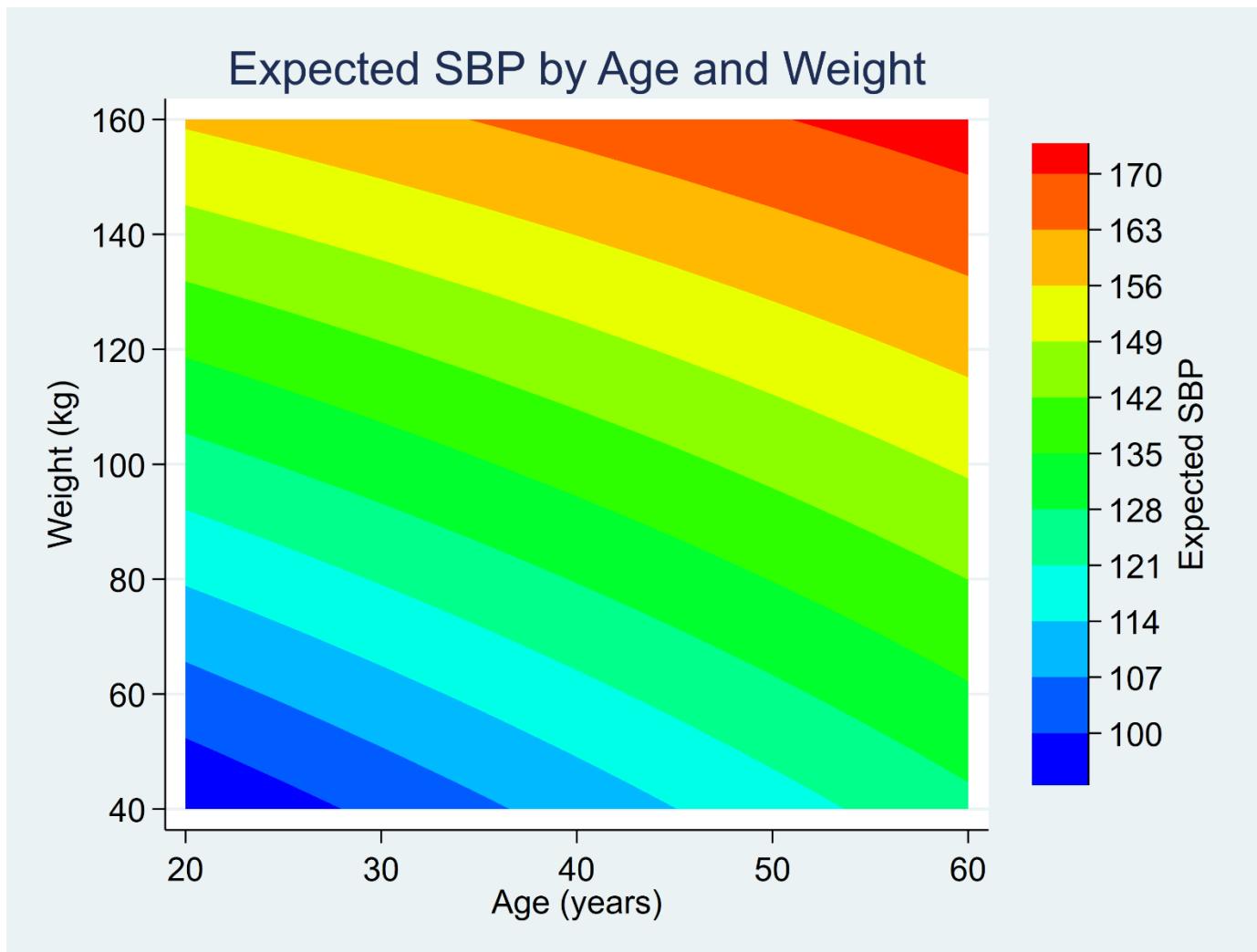
- . rename \_margin pr\_sbp
- . rename \_at1 age
- . rename \_at2 weight
- . describe pr\_sbp age weight

| variable | storage<br>name | display<br>type | value<br>format | label | variable label               |
|----------|-----------------|-----------------|-----------------|-------|------------------------------|
| pr_sbp   |                 | float           | %9.0g           |       | Linear prediction, predict() |
| age      |                 | byte            | %10.0g          |       | Age (years)                  |
| weight   |                 | int             | %9.0g           |       | weight (kg)                  |

# Continuous-by-Continuous Interactions

```
. twoway (contour pr_sbp weight age, ccuts(100(7)170)), ///
> xlabel(20(10)60) ///
> ylabel(40(20)160, angle(horizontal)) ///
> xtitle("Age (years)") ///
> ytitle("Weight (kg)") ///
> ztitle("Expected SBP") ///
> title("Expected SBP by Age and Weight")
```

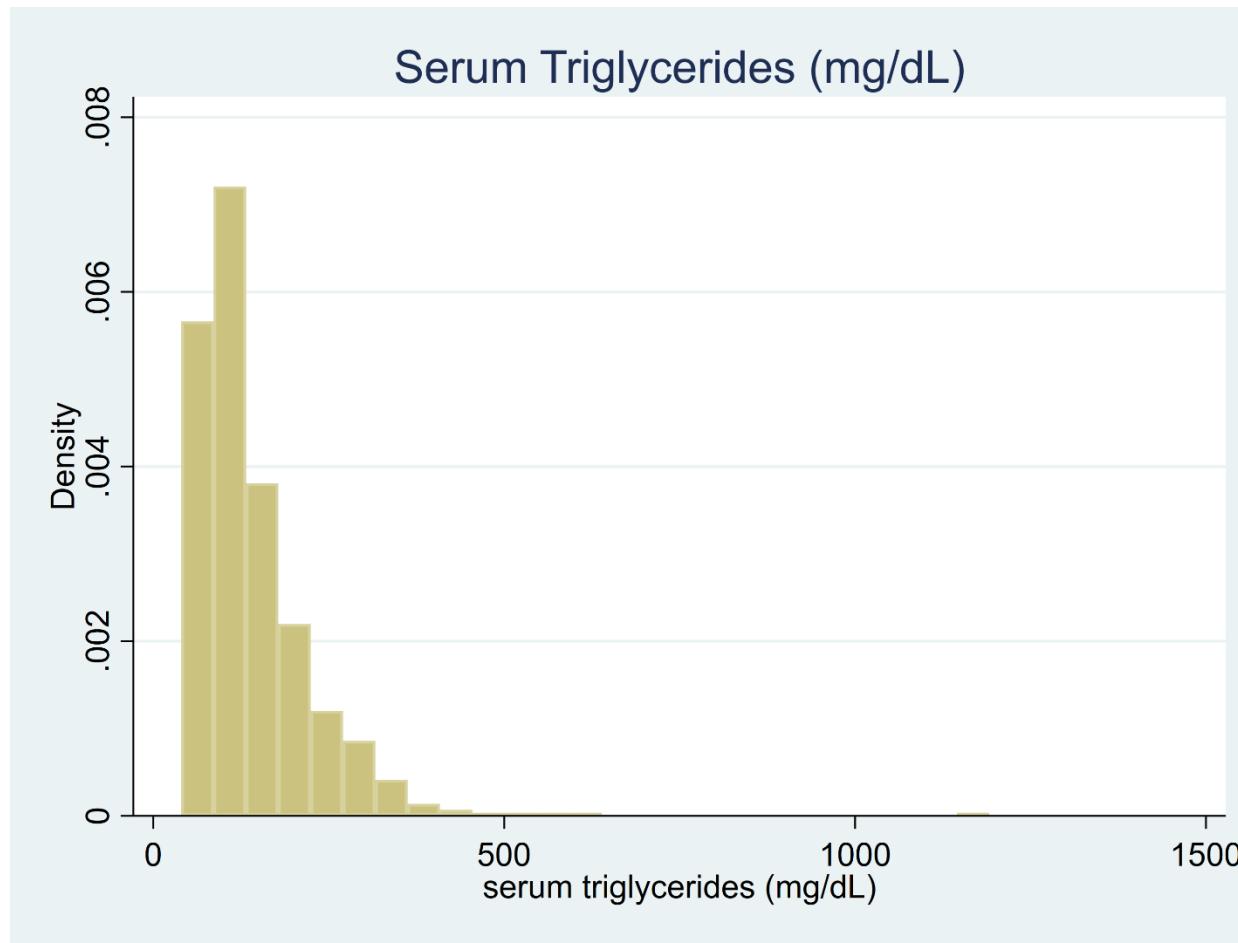
# Continuous-by-Continuous Interactions



# Outline

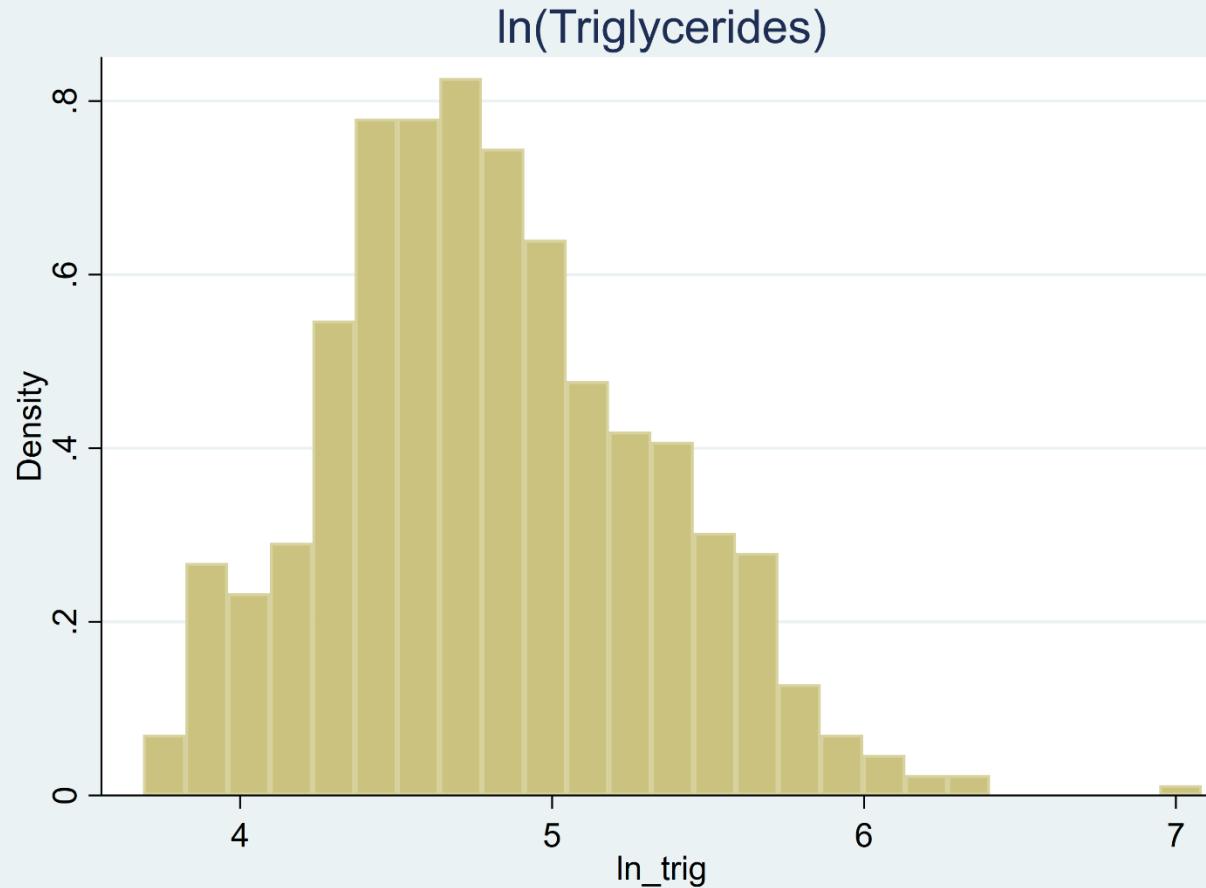
- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - Binary covariates
  - Categorical covariates
- Linear regression with multiple covariates
- Linear regression with interactions
- **Linear regression with log-transformations**

# Log-Transformations



```
histogram triglycerides, title("Serum Triglycerides (mg/dL)")
```

# Log-Transformations



```
generate ln_trig = ln(triglycerides)
histogram ln_trig, title("ln(Triglycerides)")
```

# Log-Transformations

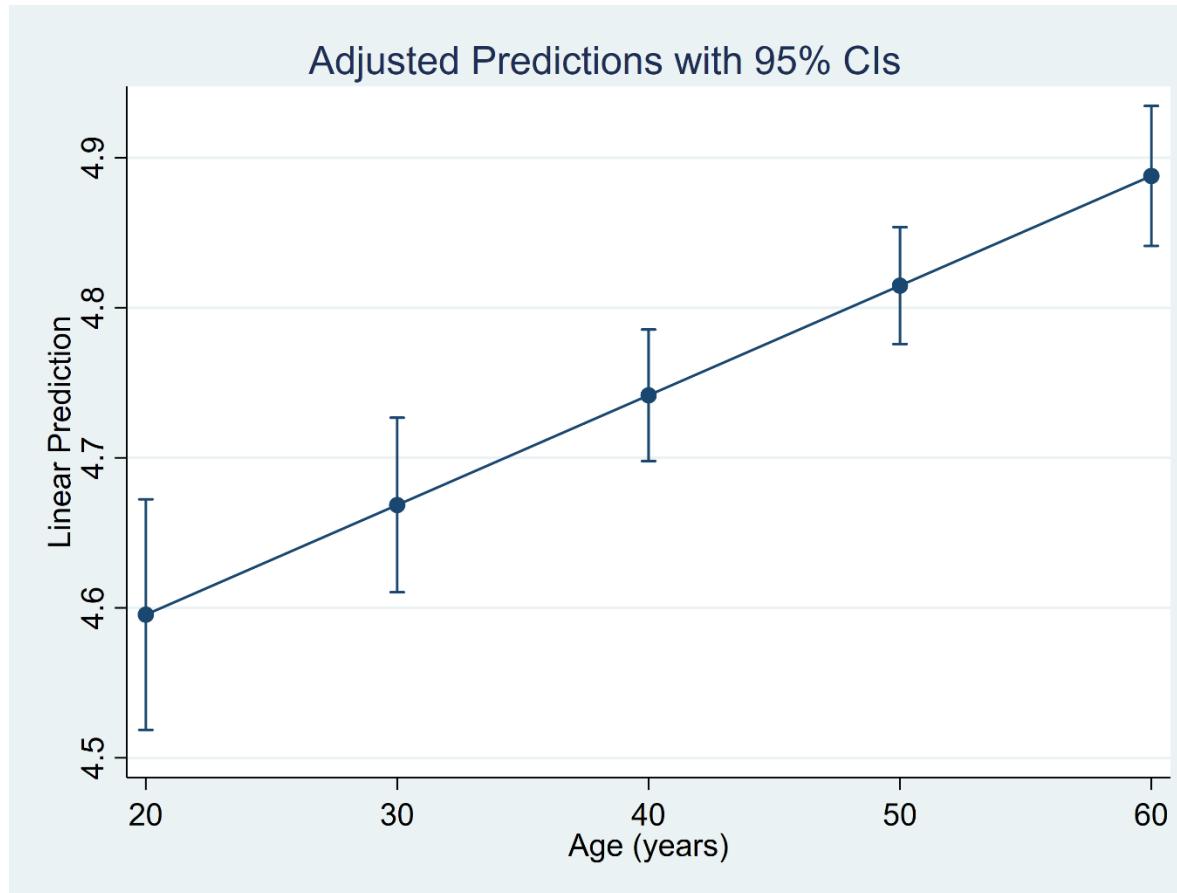
```
. regress ln_trig c.age
```

| Source   | SS         | df  | MS         | Number of obs | = | 633    |
|----------|------------|-----|------------|---------------|---|--------|
| Model    | 9.63606636 | 1   | 9.63606636 | F(1, 631)     | = | 38.86  |
| Residual | 156.454307 | 631 | .247946603 | Prob > F      | = | 0.0000 |
| Total    | 166.090373 | 632 | .262801223 | R-squared     | = | 0.0580 |
|          |            |     |            | Adj R-squared | = | 0.0565 |
|          |            |     |            | Root MSE      | = | .49794 |

| ln_trig | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|---------|----------|-----------|-------|-------|----------------------|----------|
| age     | .0073099 | .0011726  | 6.23  | 0.000 | .0050073             | .0096126 |
| _cons   | 4.44925  | .0605402  | 73.49 | 0.000 | 4.330365             | 4.568134 |

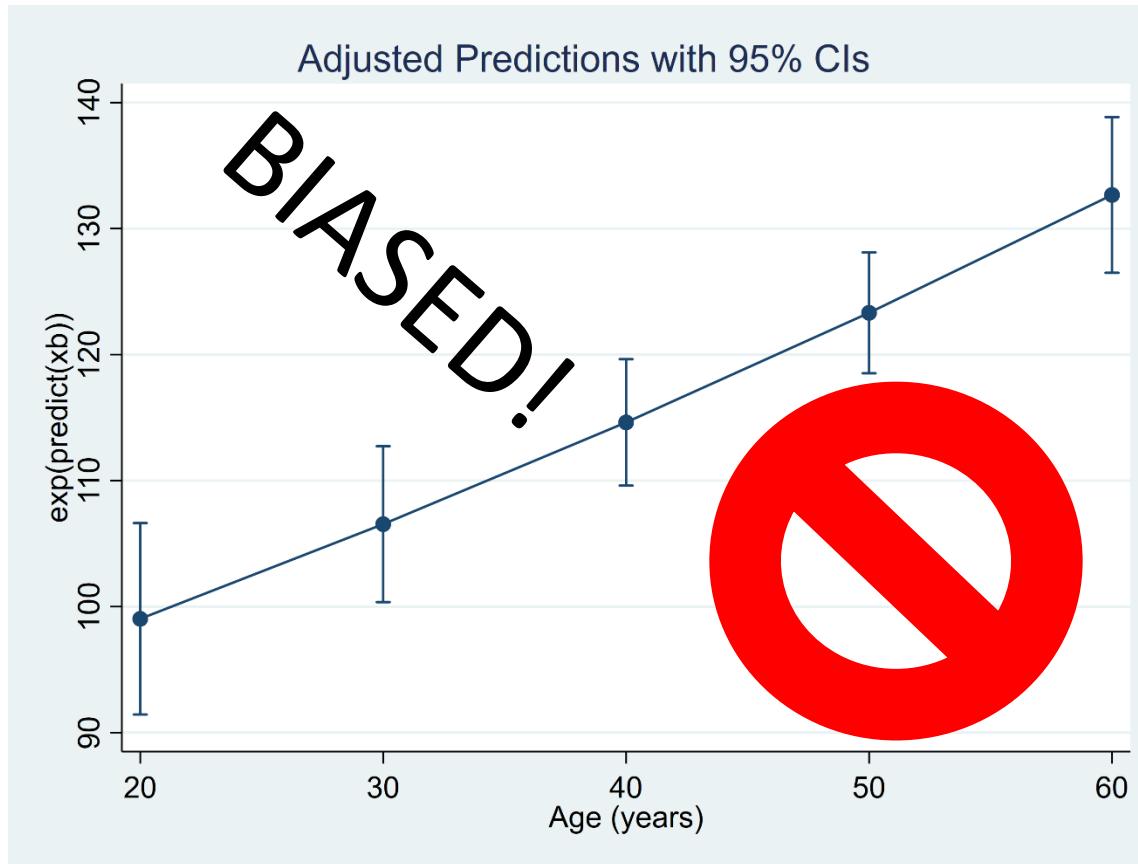
How do we interpret the coefficient for age?

# Log-Transformations



```
 margins, at(age=(20(10)60))  
 marginsplot
```

# Log-Transformations



```
margins, expression(exp(predict(xb))) at(age=(20(10)60))  
marginsplot
```

# Log-Transformations

```
. poisson triglycerides c.age, vce(robust)
```

Iteration 0: log pseudolikelihood = -15239.099

Iteration 1: log pseudolikelihood = -15239.099

Poisson regression

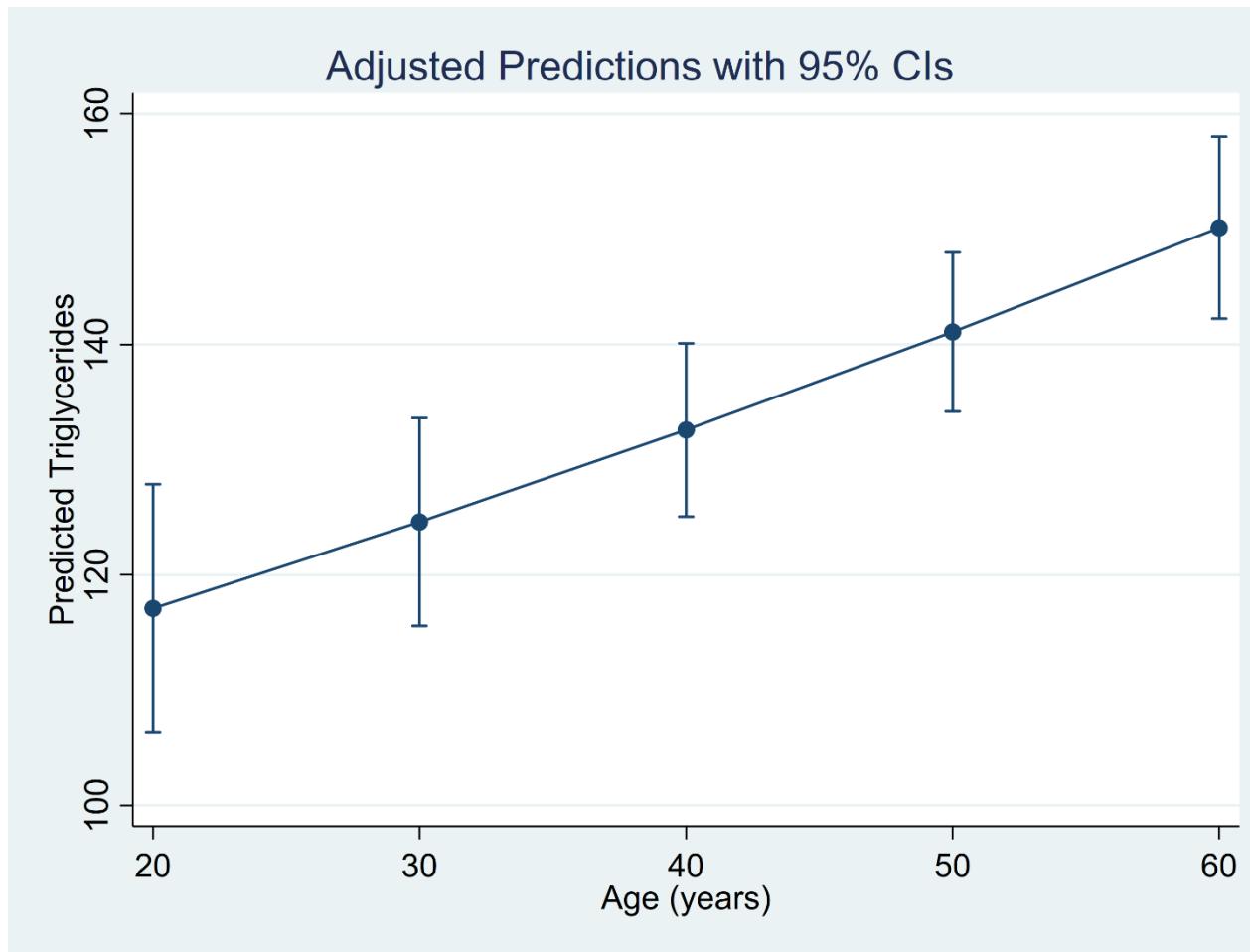
|               |   |        |
|---------------|---|--------|
| Number of obs | = | 633    |
| Wald chi2(1)  | = | 24.76  |
| Prob > chi2   | = | 0.0000 |
| Pseudo R2     | = | 0.0306 |

Log pseudolikelihood = -15239.099

| triglycerides | Robust   |           |       |       |                      |          |
|---------------|----------|-----------|-------|-------|----------------------|----------|
|               | Coef.    | Std. Err. | z     | P> z  | [95% Conf. Interval] |          |
| age           | .0062146 | .0012488  | 4.98  | 0.000 | .003767              | .0086622 |
| _cons         | 4.63863  | .0694436  | 66.80 | 0.000 | 4.502523             | 4.774737 |

Use poisson rather than regress; tell a friend

# Log-Transformations

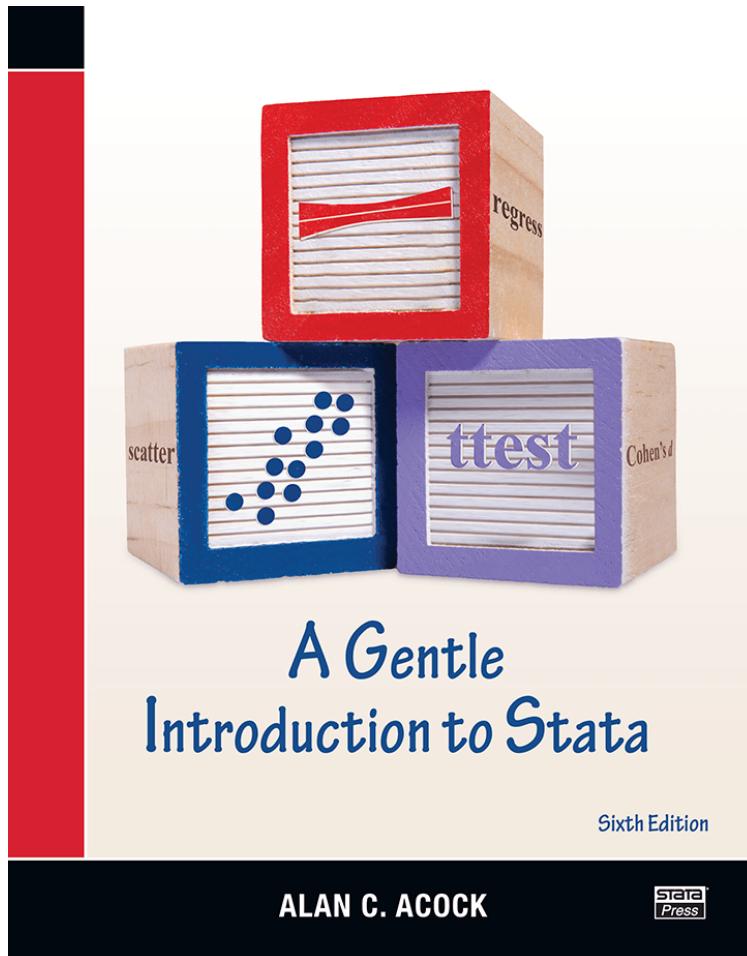


```
 margins, at(age=(20(10)60))  
 marginsplot
```

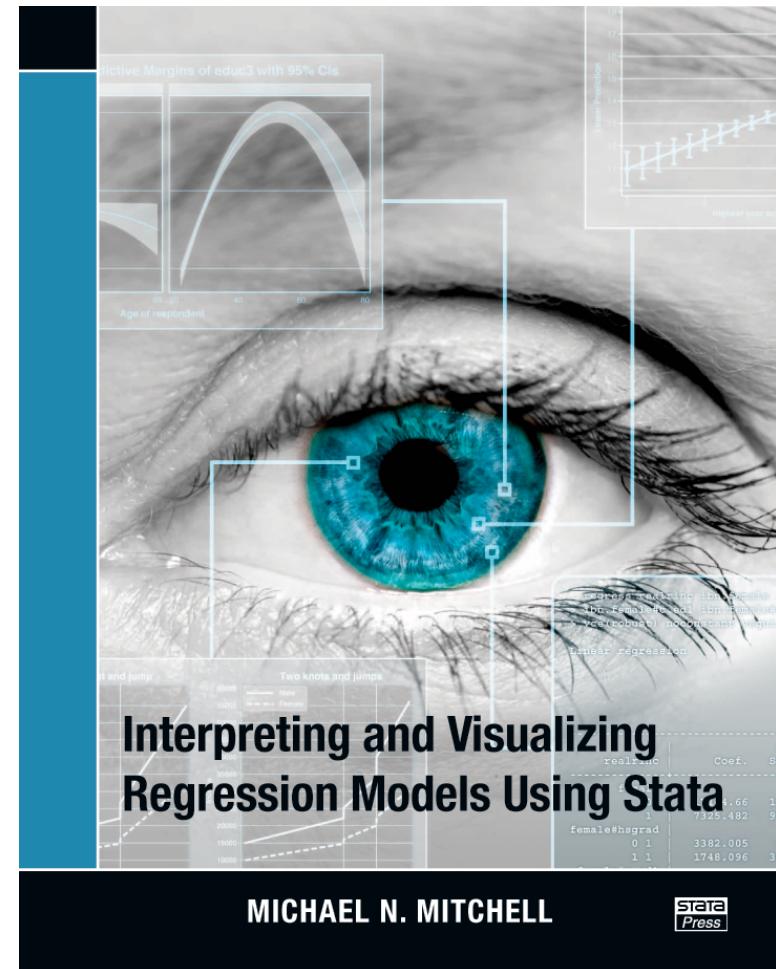
# Outline

- The dataset
- Linear regression with one covariate
  - Continuous covariates
  - Binary covariates
  - Categorical covariates
- Linear regression with multiple covariates
- Linear regression with interactions
- Linear regression with log-transformations

# Other Resources



<https://www.stata.com/bookstore/gentle-introduction-to-stata/>



<https://www.stata.com/bookstore/interpreting-visualizing-regression-models/>

# Other Resources

- [Stata Manual: margins](#)
- [Stata Manual: marginsplot](#)
- [Stata Manual: margins, contrast](#)
- [Stata Manual: margins, pwcompare](#)
- [Stata Manual: graph twoway contour](#)
- [Stata Blog: Use poisson rather than regress; Tell a friend](#)
- [In the spotlight: Interpreting models for log-transformed outcomes](#)
- [In the spotlight: Visualizing continuous-by-continuous interactions with margins and twoway contour](#)
- [YouTube: Factor Variables Playlist](#)
- [YouTube: Margins Playlist](#)
- [YouTube: Profile Plots and Interaction Plots Playlist](#)

# Thanks for coming!

## Questions?

[chuber@stata.com](mailto:chuber@stata.com)

You can download the slides, datasets, and do-files here:

<https://tinyurl.com/statareg>