Model Interpretation and Visualization using Stata

Iowa Social Research Center (ISRC) Workshop

Desmond D. Wallace

Department of Political Science
The University of Iowa
Iowa City, IA

November 17, 2017

Regression Highlights

- A way to summarize the relationship between variables.
- Assuming there is a relationship between Y and the independent variable(s).
- Relationship may be linear (OLS) or non-linear (CLDV).
- **Regression helps our understanding of how our dependent variable of interest changes when one or more independent variables vary, while holding remaining variables fixed.**

Regression Tables

- Important to report regression results in publication quality
- NEVER USE STATA output
- Multiple ways to create tables that can be featured in Word, PowerPoint, LaTeX documents
- Information table should feature include:
  1. Coefficient Estimate (REQUIRED)
  2. Standard Errors (Could include test statistic or p-value)
  3. Significance Stars
  4. Model Fit Statistics are useful (e.g., R^2)

outreg2

- outreg2 is a user-written Stata program
- Provides a fast and easy to produce regression tables
- Basic Syntax: `outreg2 using` *`filename`*`, replace`
- outreg2 command is executed AFTER regression model is estimated

outreg2 Example

```
reg realrinc age i.female
outreg2 using Tables/model.tex, replace tex(fragment)
```

outreg2 Example

VARIABLES	(1) realrinc
age	255.8***
	(39.38)
Constant	9,558***
	(1,807)
Observations	1,201
R-squared	0.034

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Coefficient Plots

- Sometimes, regression models feature many variables
- Also, showing many numbers and stars can be difficult for some readers
- An alternative to reporting a table is a plot of the regression results

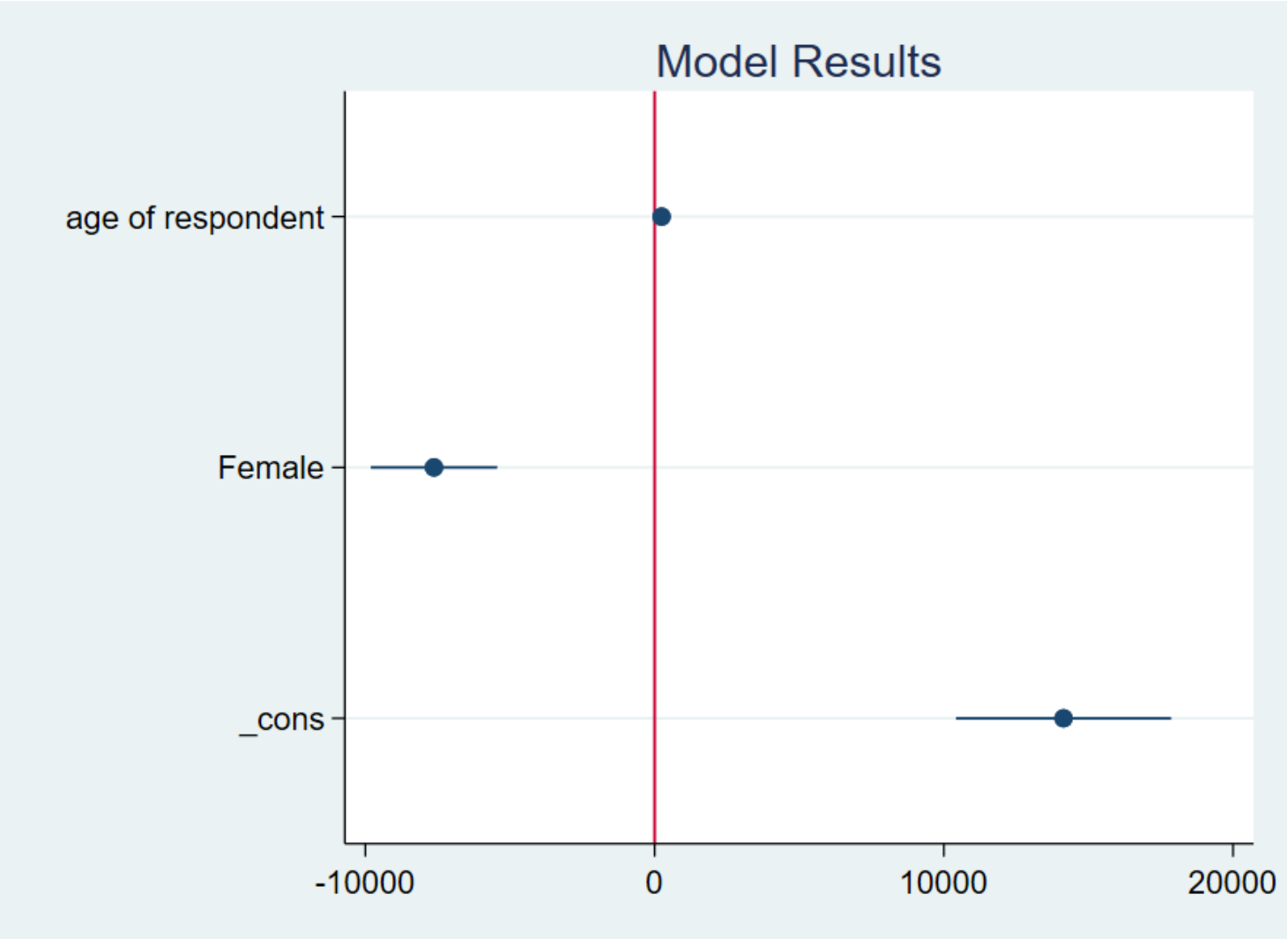
coefplot

- coefplot is another user-written Stata program
- Plots regression results in "dot-whisker" format
  - "Dot" – Coefficient Estimate
  - "Whisker" – Confidence Interval
- Basic Syntax: coefplot
- coefplot command is executed AFTER regression model is estimated

coefplot Example

```
reg realrinc age i.female
coefplot, title("Model Results") xline(0)
```

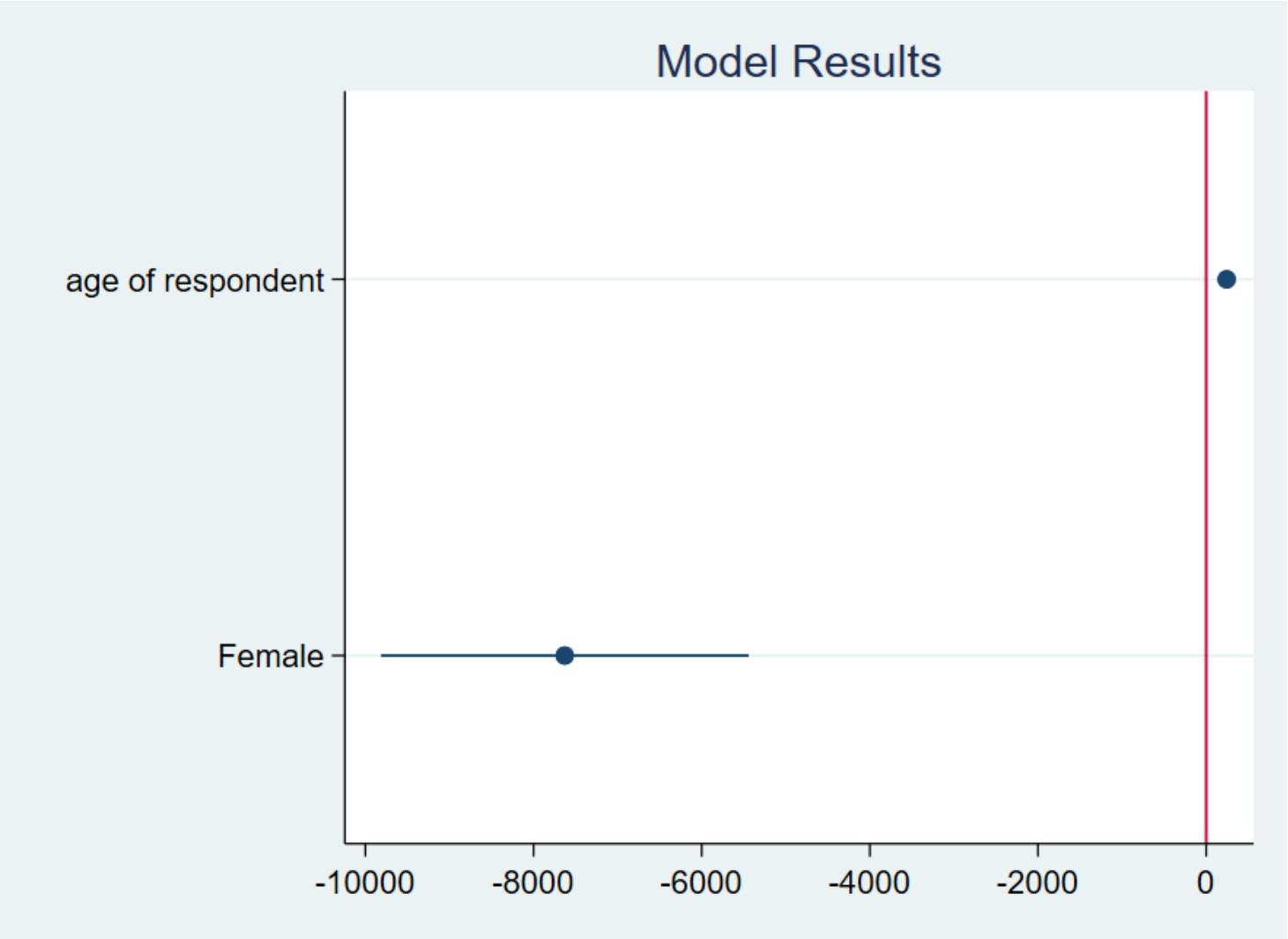
coefplot Example



coefplot Example

```
reg realrinc age i.female
coefplot, title("Model Results") xline(0) drop(_cons)
```

coefplot Example



Interpreting Coefficients

- Can directly interpret coefficient estimates.
- *A one unit change in X_k leads to a β_k change in Y (holding all other variables constant).*
- Assumes X_k is not a constituent term for an interaction variable.

Predicted (Fitted) Values

- The result of substituting values of interest for the independent variable(s).
- $E[Y|X] = X\hat{\beta}$
- Can calculate standard errors to determine if $E[Y|X = x]$ is statistically significantly different from zero.
- Multiple ways to calculate fitted values in Stata.

Marginal and Discrete Change

- Measuring the change in the dependent variable for a change in one independent variable, holding remaining independent variables constant.
  - *Marginal Change* is the partial derivative, or instantaneous rate of change, in the dependent variable w.r.t. an independent variable, holding remaining variables constant.
  - *Discrete Change* or *First Difference* is the difference in the prediction from one specified value of an independent variable to another specified value, holding remaining variables constant.

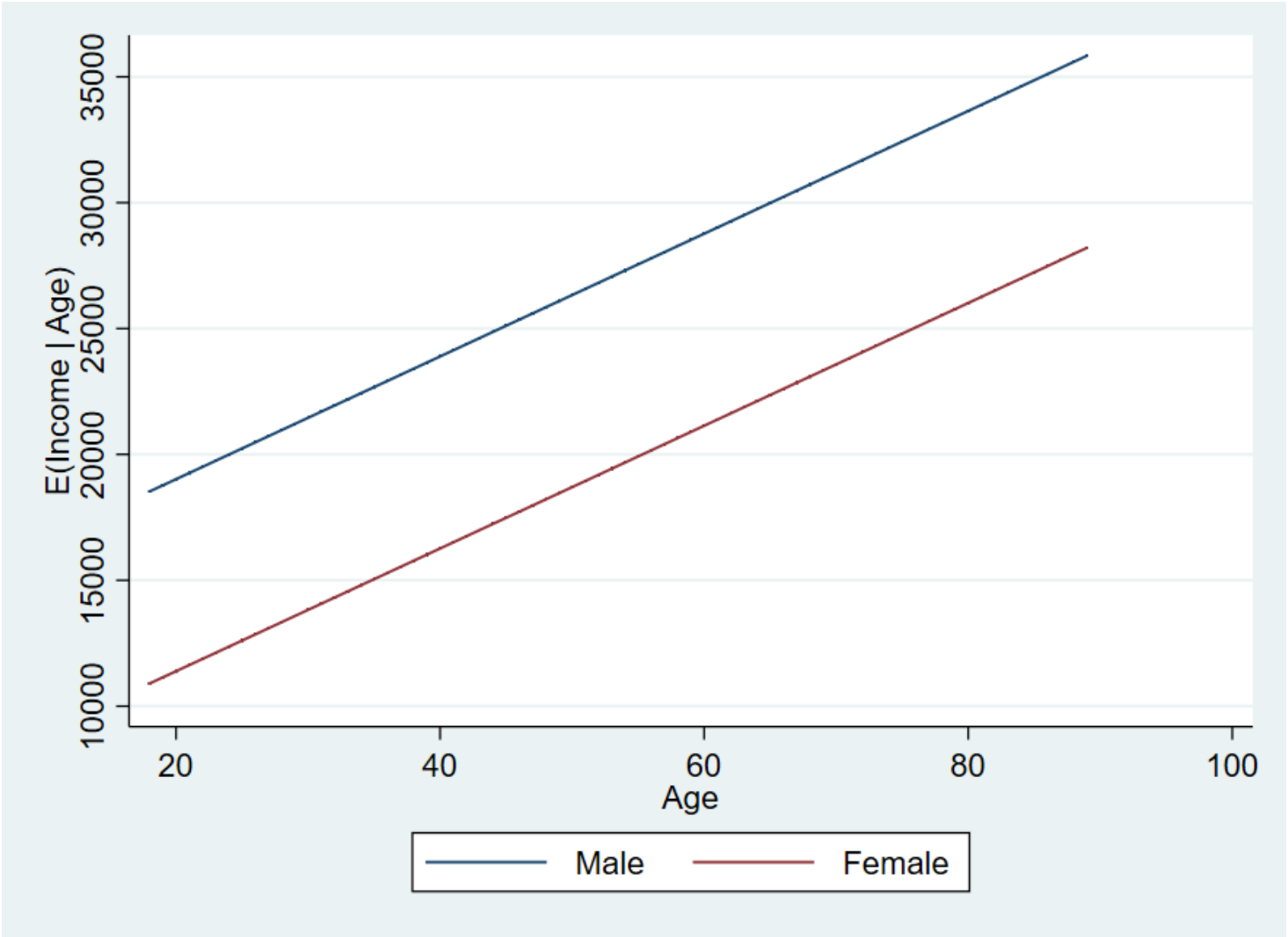
Marginal and Discrete Change

- Marginal Change: $\frac{\partial E[Y|X]}{\partial x_k} = \frac{\partial X\beta}{\partial x_k} = \beta_k$
- Discrete Change: $\frac{\Delta E[Y|X]}{\Delta x_k} = E[Y|X, x_k + 1] - E[Y|X, x_k] = \beta_k$

Marginal and Discrete Change

- $\frac{\partial E[Y|X]}{\partial x_k} = \frac{\Delta E[Y|X]}{\Delta x_k} = \beta_k$, assuming there is no interaction terms.
- The standard error of the marginal effect is the same as the standard error of the estimated beta coefficient.
- *For a unit increase in x_k, the expected change in Y equals β_k, holding all other variables constant.*
- *Having characteristic x_k (as opposed to not having the characteristic) results in an expected change of β_k in Y, holding all other variables constant.*
- When there is no interaction term present,
  Marginal Change = Discrete Change

Marginal Effects



margins

- Computes predicted values and marginal effects from last estimated regression model
- Reports computed statistic, standard error, test statistic, *p*-value and 95% CI.
- `at(atspec)` option allows for the calculation of predicted values and marginal effects at specific values of independent variable(s).
- `dydx()` option allows for calculating marginal effects.
- Factor variables (`i.varname`) can go after the `margins` command or within the `at(atspec)` option.
- Continuous variables can only be specified within the `at(atspec)` option.
- `atmeans` option sets variables not specified to be held at their mean value.

Predicted (Fitted) Values – `margins` Syntax

- `margins` – Overall predicted value with all independent variables held at their mean value.
- `margins, at(varname=#)` – Predicted value when one or more independent variables are fixed to a specific value and remaining independent variables held at their mean value.
- `margins, at(varname=numlist)` – Predicted value(s) when one or more independent variables are fixed to multiple values and remaining independent variables held at their mean value.
- `margins varname` – Overall predicted value(s) for categories of `varname` with remaining independent variables held at their mean value.

Marginal Change – `margins` Syntax

- `margins, dydx(varname)` – Average marginal effect a one-unit increase in `varname` has on the dependent variable, holding all other variables constant.

Discrete Change – margins Syntax

- `margins, at(varname=(start end)) post` – Calculates predicted values at specified values, and treats results as estimation results.
- `lincom 2._at - 1._at` – Calculates the difference between the prediction of the ending value and the prediction of the starting value.

marginsplot

- Graphs the results of last estimated `margins` command
- **Needs to be executed immediately after** `margins`
- Resulting graph includes an overall title, a title for the y-axis, x-axis features the name of the variable (variable label if one is included).
- The featured values on the x-axis are the values specified from the `margins` command.
- Can use the `recast` and `recastci` options to change how results are graphed.

Email: desmond-wallace@uiowa.edu
Any Questions?