

# Stata Workshop 4:

## Simple and Multiple Regressions

Tao Wang, SSQL @ Swarthmore, [swarthmore.edu/ssql](http://swarthmore.edu/ssql)

---

### 1. Workshop Objectives

Welcome! In this workshop, we will cover selected topics on regression that will help you with PS 4 and 5.

By the end of this session, you will be able to:

- **Define a DGP:** Create a "true" relationship between variables to test OLS performance.
  - **Diagnose Models:** Generate residuals and fitted values to check for model violations.
  - **Identify Outliers:** Visualize how extreme observations skew results.
  - **Understand Controls:** Grasp Omitted Variable Bias (OVB) using correlated predictors.
  - **Master the FWL Theorem:** See how multiple regression "partials out" confounding variables.
- 

### 2. New Commands to be Mastered

OLS regression: `regress`

Set sample size: `set obs #`

Generate normal data: `drawnorm`, `rnormal()`

Post-estimation: `predict`, `rvfplot`, `rvpplot`, `avplot`

---

### 3. Workshop Exercises

Exercise 1:

With the same generated X's and epsilons, multiple epsilon by 10 in the DGP. How does that affect the regression results?

Exercise 2:

In addition to setting  $Y = 100$  in the first row, set  $X = 2$ . How does that affect the regression results?

Exercise 3:

Run a regression of  $Y$  on  $X$  and  $Z$ . Then run a regression of  $Y$  on  $X$  only. Compare the estimated coefficient on  $X$ .