Spaghetti Plot Demo

Andy Grogan-Kaylor

15 Apr 2021

# Get Data

using data from https://stats.idre.ucla.edu/stata/examples/mlm-imm/introduction-to-multilevel-modeling-by-kreft-and-de-leeuwchapter-4-analyses/

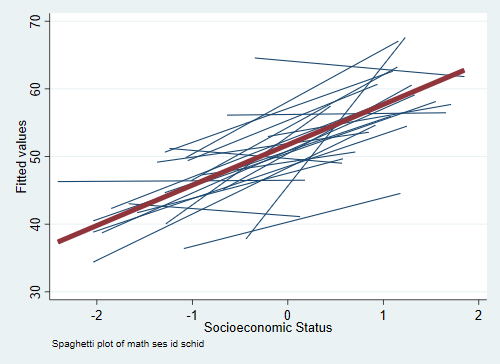
. use https://stats.idre.ucla.edu/stat/examples/imm/imm23, clear

. label variable ses "Socioeconomic Status" // correct spelling of variable label

# Basic Spaghetti Plot

. spagplot math ses, id(schid)

. graph export graph1.png, width(500) replace  
(file graph1.png written in PNG format)



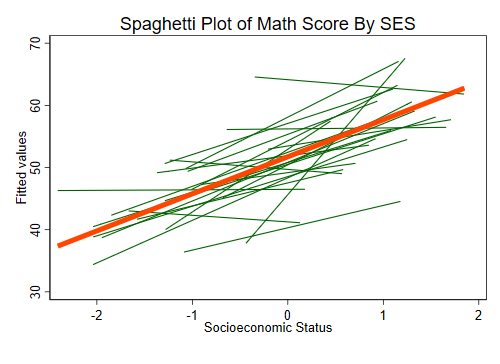
Basic Spaghetti Plot

# Add Better Scheme

Schemes are very helpful in making better looking Stata graphs. A useful Stata scheme is s1color. Useful user written schemes are lean2, plottig (type findit lean2 or findit plottig to install these), and my own [Michigan Stata graph scheme](https://agrogan1.github.io/Stata/michigan-graph-scheme/).

. spagplot math ses, id(schid) ///  
> scheme(s1color) ///  
> title("Spaghetti Plot of Math Score By SES") ///  
> note(" ") // blank "note" since title explains this graph

. graph export graph2.png, width(500) replace  
(file graph2.png written in PNG format)

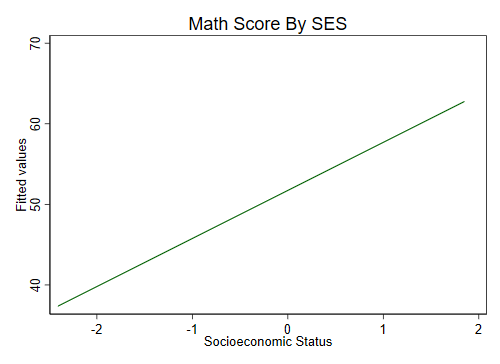


Spaghetti Plot With Better Options

# Now Try twoway Syntax

. twoway lfit math ses, scheme(s1color) title("Math Score By SES")

. graph export graph3.png, width(500) replace  
(file graph3.png written in PNG format)



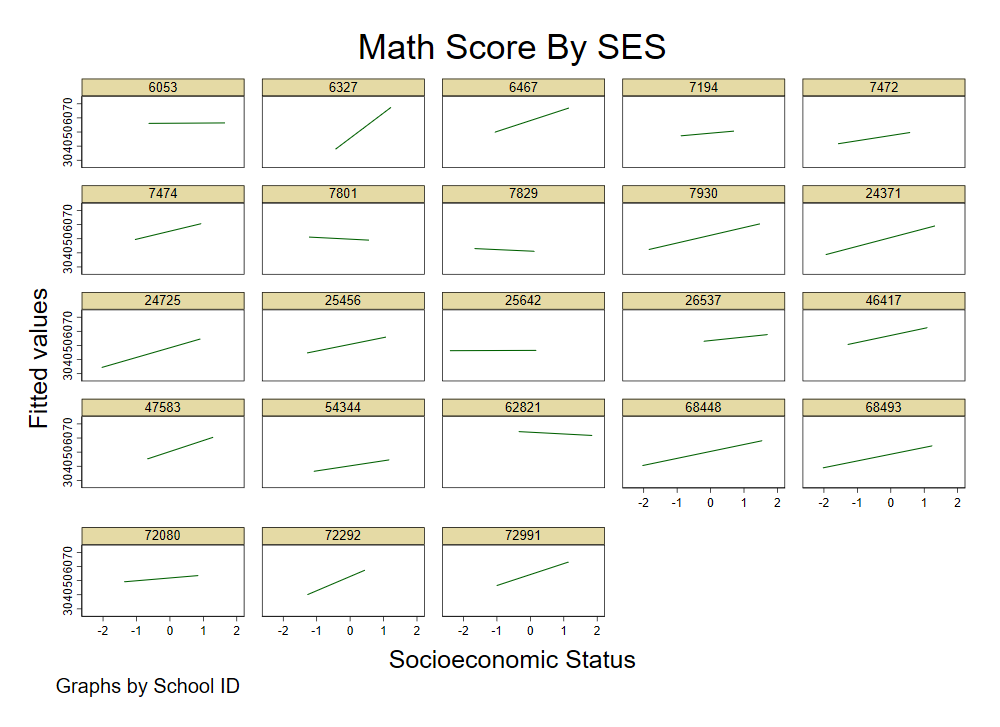
Initial twoway Graph

# Separate Panels For Schools

This ONLY works well with a limited number of schools.

. twoway lfit math ses, scheme(s1color) by(schid, title("Math Score By SES"))

. graph export graph4.png, width(1000) replace  
(file graph4.png written in PNG format)

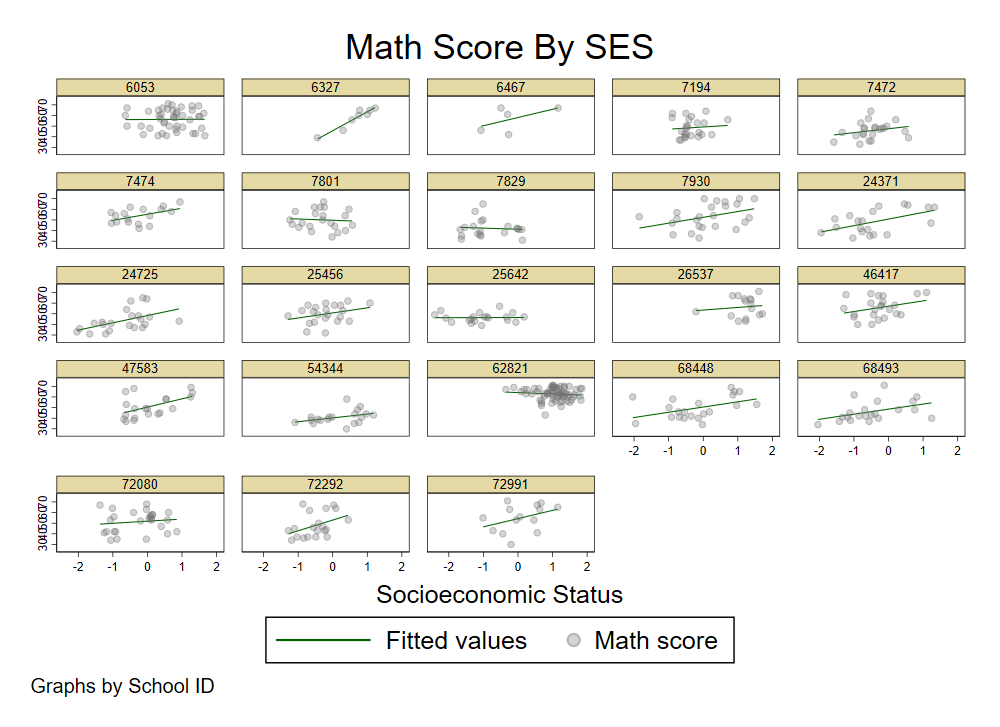


Separate Panels For Schools

# Separate Panels For Schools With Scatterplots

. twoway (lfit math ses) ///  
> (scatter math ses, mcolor(gs7%30)), /// color gs7 @ 30% transparency  
> scheme(s1color) by(schid, title("Math Score By SES"))

. graph export graph5.png, width(1000) replace  
(file graph5.png written in PNG format)



Separate Panels For Schools With Scatterplots

# "Model Based" Spaghetti Plot

A sometimes unacknowledged point is that spaghetti plots--unless we take steps to correct this--reflect *unadjusted*, or *bivariate* associations.

We may sometimes wish to develop a spaghetti plot that reflects the *adjusted* estimates from our models.

To do this we first estimate a multilevel model.

. mixed math ses meanses || schid: // multilevel model; random intercept; no random effects  
  
Performing EM optimization:   
  
Performing gradient-based optimization:   
  
Iteration 0: log likelihood = -1871.9169   
Iteration 1: log likelihood = -1871.9169   
  
Computing standard errors:  
  
Mixed-effects ML regression Number of obs = 519  
Group variable: schid Number of groups = 23  
  
 Obs per group:  
 min = 5  
 avg = 22.6  
 max = 67  
  
 Wald chi2(2) = 69.58  
Log likelihood = -1871.9169 Prob > chi2 = 0.0000  
  
─────────────┬────────────────────────────────────────────────────────────────  
 math │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 ses │ 3.88476 .6096853 6.37 0.000 2.689799 5.079722  
 meanses │ 3.281962 1.464135 2.24 0.025 .4123106 6.151614  
 \_cons │ 51.48904 .7582764 67.90 0.000 50.00284 52.97523  
─────────────┴────────────────────────────────────────────────────────────────  
  
─────────────────────────────┬────────────────────────────────────────────────  
 Random-effects Parameters │ Estimate Std. Err. [95% Conf. Interval]  
─────────────────────────────┼────────────────────────────────────────────────  
schid: Identity │  
 var(\_cons) │ 8.931927 3.813085 3.868681 20.62184  
─────────────────────────────┼────────────────────────────────────────────────  
 var(Residual) │ 75.21885 4.778177 66.41333 85.19187  
─────────────────────────────┴────────────────────────────────────────────────  
LR test vs. linear model: chibar2(01) = 25.58 Prob >= chibar2 = 0.0000

NB that this is a model with only a random intercept, and no random slopes e.g. , etc....

The spaghetti plots so far give an indication of different slopes per school. Below we outline a procedure for (a) developing a spaghetti plot of adjusted estimates; and (b) ensuring that the plot reflects the structure of the model e.g. random intercept only, or random intercept + random slope(s).

To carry out this procedure we employ the \_b notation in Stata. For example, \_b[\_cons] indicates the intercept of the model while \_b[ses] indicates the slope attached to *ses*.

We need to carry out a few preliminary calculations.

1. Estimate (predict) the random effect(s).
2. Estimate the mean values of variables that we are going to hold constant.
3. Estimate predicted values () using the \_b notation.
4. Graph the spaghetti plot.

## Estimate The Random Effects

. predict u0, reffects

## Estimate the Mean Values of Relevant Variables

. summarize meanses  
  
 Variable │ Obs Mean Std. Dev. Min Max  
─────────────┼─────────────────────────────────────────────────────────  
 meanses │ 519 -.0012717 .6206429 -1.0685 1.17625

The mean of meanses is -0.00127.

## Estimate Predicted Values

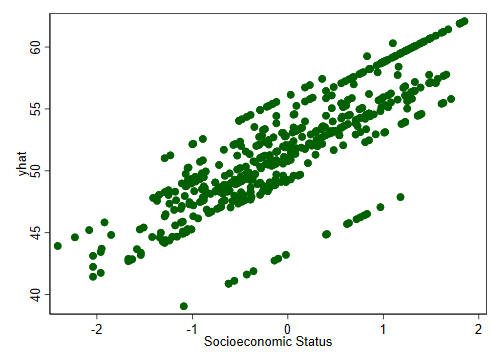
We are using , the random intercept , multiplied by the actual value of ses, and multiplied by the mean of meanses.

. generate yhat = \_b[\_cons] + u0 + \_b[ses] \* ses + \_b[ses] \* -.0012717

## Graph The Spaghetti Plot

. twoway scatter yhat ses, scheme(s1color)

. graph export graph6.png, width(500) replace  
(file graph6.png written in PNG format)



Initial "Model Based" Spaghetti Plot

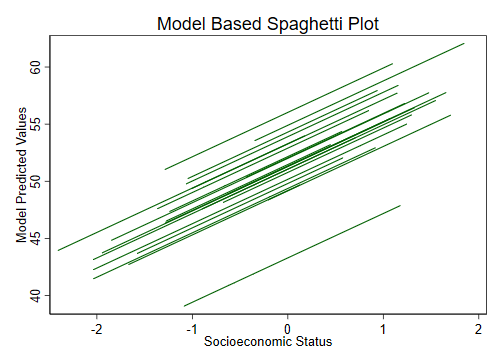
We still have a small amount of work to do to make this look more "spaghetti plot like".

We are going to use twoway connect to create connected line plots. We employ option c(L) to ensure that only ascending values are connected: i.e. each Level 2 unit has their own regression line. For c(L) to work properly we are going to need to sort the data by *school* and *ses*. Lastly, we're going to change the msymbol so that we do not see dots, but only lines.

. sort schid ses // sort on Level 2 units and x values

. twoway connect yhat ses, ///  
> title("Model Based Spaghetti Plot") /// title  
> xtitle("Socioeconomic Status") /// title for x axis  
> ytitle("Model Predicted Values") /// title for y axis  
> c(L) /// connect only ascending values  
> msymbol(none) /// no marker symbol; only lines  
> scheme(s1color) // s1color scheme

. graph export graph7.png, width(500) replace  
(file graph7.png written in PNG format)



Finalized "Model Based" Spaghetti Plot