**Why does treating solo practices as a group and imputing the same value k for all group variables for these practices, work the way it does?**

Observation: the coefficients for all group variables except grpsize\_1 stay the same no matter what k! Why?

Basically, if we just run a regression of outcome on ONLY the group-variables, and we only run it on observations grpsize\_1 == 0 (basically treating all grp variables as missing if it’s a solo practice), these coefficients will be the same as if we impute any value of k. Why? Draw out a scatter graph of outcome vs one group variable, and use dots (o) for observations tht are not imputed and crosses (x) for observations that are imputed with k. We will see that the x’s will stack in a vertical line where k is on the grp\_variable-axis. A multivariate regression of outcome on group\_variable controlling for grpsize\_1 tries to force the same slope across the group labeled by (o) and the solo-practice group labeled by (x), in a way that minimizes squared errors. But the latter group here is a vertical line no matter what k is, so any best fit line crossing the vertical stack at the average outcome value E[outcome] (like all best fit lines do) will inevitably yield the same sum of squared errors (draw it!) – because the predicted value is always E[Y] for this group! Since this solo-practice group does not matter for computing the slope for the grp\_variabe, this explains the above observation.

Observation: the coefficient for grpsize\_1 changes depending on the value of k. Why?

Again, draw a scatter plot of outcome versus grpsize\_1. The slope in a univariate regression is just the level shift in average outcome of going from grpsize\_1== 0 to grpsize\_1 == 1. When also controlling for grp\_variable, this is the E[outcome|grp\_variable = 0.1], E[outcome|grp\_variable=0.2],…, E[outcome|grp\_variable=k].