**Homework 6 General Feedback**

**Reading questions**

These went quite well overall, it seems as though you all pretty much understand these concepts. The one thing I’d like to point out is that Judd et al. actually say it is *unethical* to report results that are unduly influenced by outliers. This is strong language you don’t need to agree with by any means, but it’s worth thinking about why they say this. Note that they suggest reporting results without outliers (not both ways), if testing both ways reveals outlier effects and you satisfactorily explain how you identified outliers in your results section / preregistration.

**Data analysis**

Overall you all did quite well with this. There was some variability in terms of which points you drew out as being influential on different metrics, but that’s to be expected. Almost everyone made the same choice as we did in the key about who to remove.

Many of you used varScore to calculate the mean post-test IAT score. This works just fine, but it a bit complicated given what you want to do. When you create scores, be mindful of what the benefit of varScore is: it’s especially useful in dealing with multiple-item scales where you have some degree of missingness. With postIAT, we probably don’t want to include anyone who’s missing one of the two measurements (4 and 8 weeks), so rowMeans or simple math would work just fine (it won’t compute a score for anyone with NAs on either of these variables). Obviously varScore isn’t wrong, but you can save yourself some typing by being mindful of whether you need to use it.

A note on interpreting correlations: make sure you’re clear on the distinction between *r*, *p*, and *b.* The first, *r*, indicates essentially how close the data are, in general, to the regression line: the tightness of fit. It doesn’t tell you the magnitude of the relationship, but rather the strength of the relationship. It is relatively closely related to *p*, which additionally incorporates information about sample size: holding sample size constant, higher *r* = more significant *p*. The parameter estimate, *b*, is what will tell you about the actual magnitude of the relationship: that is, the slope of the regression line. Make sure this distinction is clear. A few of you used what I’d consider to be slightly incorrect terminology when referring to the correlations in the dataset.

* *r* = strength, tightness around best fit line
* *p* = significance, likelihood we’d observe a relationship of that strength by chance, given our sample size
* *b* = magnitude, how steep the slope of the regression line is