**Homework 4 General Feedback**

**Conceptual Questions**

For question 1, a few of you answered that the mean is the most efficient estimator for a given sample based on the core assumption of normal distribution of DATA. This is true for the special case of a mean only model. However, we were hoping you would answer “normal distribution of ERRORS”. Again, this distinction doesn’t matter much for the special case of just estimating the mean. However, note that this matters a lot as soon as you introduce one or more continuous predictors into the linear model. In that case, the assumption is about the distribution of errors of y and not the distribution of the scores of y. The invalid notion that the linear model assumption is about normal distribution of the scores is left over from the ANOVA framework and doesn’t apply in the general linear model.

**Data Analysis**

In your scripts, make sure you tell R to show your plot by typing the name of the object after you create the plot. This is necessary in ggplot but would not be in base graphing (if you don’t store your plot in an object).

Most of you included both vertical and horizontal jitter on the data points you put over your bar graphs. I assume you did this because our example code included this. However, in many cases (such as this one), horizontal jitter is enough. The added vertical jitter ended up making the figures look weirder than they had to.

**Write-ups**

I think it is worth pasting again what Mitch said about HW3 :). In his example results, notice that he includes the *b*s.

“It’s worth looking at the key to get an idea of roughly what we were imagining. You’ll also notice I likely made some edits to the way you reported (or failed to report) statistics. You will come to resent what a stickler I am about this. See example:

Students get more annoyed with Mitch every time he corrects the way they report statistics, *F*(1, 17) = 6.53, *p* < .05. Every correction predicts a 0.2 point increase in annoyance on a 7-point Likert scale, *b* = 0.211. Error corrections explain 33% of students’ annoyance with Mitch, = .332.

Two important notes: always italicize the name of statistics (e.g., *R2*) and only include a “0” before a decimal point if that number *can* exceed one (e.g., = .31 but *F*(X, XX) = 0.31).”

When writing up your results sections, be sure to include all tests that you did. In this homework, you first tested for baseline differences but most of you did not report this.

**General**

Before emailing your r script, make sure it runs from top to bottom with no errors.

Again, Mitch already asked for this: “Pleeeease delete existing question text when you turn in homework, leaving only the question number and your response. Having all those words on the page makes it difficult to grade.”