

# Sensitivity Analysis

- Simple sensitivity analysis:  $E[T] = \beta_0 + \beta_1 X$ . If you're missing some  $X$ , how would you fill them in to get really different estimates of  $\beta$  than you get from a CCA?

You can impute values of  $X$  that are missing to see how different your estimates would be.

- Stochastic: you implicitly posit some model for missing values of  $X_M$ , impute them according to that model, get parameter estimates from the completed datasets and plot the distribution of those.

Note that this implicitly sets some model for missing  $X$ , so the “sensitivity” really has to do with the model you specify.

- Optimized: If you want to get a sense of how bad things could be, you would want to find values of  $X_M$  such that  $\hat{\beta}^*$  on the completed data would be really different from your  $\hat{\beta}$ .

Essentially, we have this problem: Let  $\hat{\beta}$  be the obtained estimate from the analysis. We want to find  $X_M$  such that  $\mathbf{c}'\hat{\beta}^* - \mathbf{c}'\hat{\beta}$  is large, where  $c$  is some contrast vector. Thus, we have

$$\begin{aligned} \min_{X_M} & -(\mathbf{c}'\hat{\beta}^* - \mathbf{c}'\hat{\beta})^2 \\ \text{subject to } & X_M \in \mathcal{X} \end{aligned}$$

It will be difficult to solve this analytically, but we could try gradient descent. This may be sensitive to starting point, so we may want to try several random initializations. Finally, we should keep track of imputed  $X_M$  and the squared difference at each step, so that we can plot observed vs. imputed  $X$  and how different the imputed parameter estimates are.

Note, if the math proves unwieldy to update all missing  $X_i$  at once, we could do a series of stochastic updates, where we take a subset of the  $i$  for which there is missingness, update those  $X_i$  and then turn around and update another set of  $X_i$ . It would make sense to sort by missingness pattern to do this.

- Paper would propose the idea of sensitivity analysis. Describe the two algorithms (stochastic search vs. gradient descent). Demonstrate them on a data set or two. Close with recommendations.
- Other idea: Find solutions to

$$\mathbf{c}^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{T} = \mathbf{c}^T\mathbf{0}$$

for  $\mathbf{X}_M$ . This will have a large number of solutions because it is a nonlinear system of equations with more unknowns than knowns.

Perhaps a series of numerical solution?