f

Diahmin Hawkins dlh2166@columbia.edu

12/6/2024

Methods

The methods used in this the ADEMP framework. The ADEMP Framework is a structured approach used in simulation models that stands for Aims, Data-generating mechanisms, Methods, Estimands, Performance measures. The *Aims* of this study consists of : *Aim 1*: Design a simulation study using the ADEMP framework from class to evaluate potential study designs. *Aim 2*: Explore relationships between the underlying data generation mechanism parameters and the relative costs (c_2/c_1) and how these impact the optimal study design. *Aim 3*: Extend your simulation study to the setting in which Y follows a Poisson distribution with mean μ_i and explore how this impacts the results. The hierarchical model for this setting is given below. The *Data-generating mechanism* used in this study is a randomized cluster trial with assigned treatment groups using simulated data. To start this analysis, we will consider Y to be normally distributed.For the observation r(r=1,. . . ,R for repeated observations) in cluster g(g=1,. . . ,G groups). The X_i be a binary indicator of whether or not cluster g is assigned to treatment group (0= control, 1= treatment) and let Y be the observed outcome. To estimate the treatment effect, we will assume a hierarchical model for Y where $\mu_{i0} = \alpha + \beta X_i$ fixed effect, μ_{i0}=α+ β $\mu_i|\epsilon_i$= μ_{i0} + ϵ_i with ϵ_i~N(0,γ^2), or in other words μ_i ~ N(μ_{i0},γ^2) $Y_{rg}|\mu_i$ + ϵ_{rg} with ϵ_{rg}~iid N(0,σ^2).

This means that the marginal mean of Y_{rg} is E($Y_{rg}|$ X_i)= α + β X_i and the conditional mean given ϵ_i is E($Y_{rj}|X_i$,ϵ_i)=α + β+ X_i+ ϵ_i. The estimate of β will be our estimate of the average treatment effect and is our parameter of interest.

For *Aim 3* , each cluster g, we have log(μ_i)~ N (α + β X_I,γ^2). We observe the conditionally independent units (r=1,. . . R) withinh the cluster $Y_{rg}|\mu_i$ ~ Poisson(μ_i). The sum of iid Poisson random variables is still Poisson therefore we have the simplified model $Y_r|\mu_r$~Poisson(Rμ_i).

Methods: The methods used in this simulation model is a normally distributed linear regression model for Aim 1 with varying factors that varies different parameters(γ and σ). We will vary (γ and σ) because these parameters directly influence the behavior of the clustering structure, precision of estimates, and design considerations. In Aim 2, we will vary the ratios of (c_2/c_1) to see how the total cost is impacted. Varying $c1$ (cost per cluster) and $c2$ (cost per additional individual within a cluster) in this simulation modelbecause it assists with resource allocations, optimizing study design, and help maxing out our constrained budget. These costs directly impact the total cost of the study and guide decisions on whether to prioritize increasing the number of clusters (G) or the number of individuals per cluster (R). In Aim 3, we will use a poisson regression model to represent an extension of the hierarchical model to handle the outcomes Y_rj in a count base way, rather than countinous

Performance Measures: The performance measures we will be evaluating in this study is the ICC,variance, and cost efficiency. We will also be observing the correlation between observations using the ICC,cost efficieny, and the design efficiency by finding the optimal design.