# STAT 547T Homework 2

**Instructions**: Typeset your solutions and submit them as a PDF via Canvas.

**Grading**: One problem will be fully graded, and the other will be graded on a binary scale (full points or no points) based on effort.

## Problem 1

Let $\mathcal{H}_0$ be a set of candidate null hypotheses that could be selected from some data $X \sim F_X$.

Let $\mathcal{A}$ be an arbitrary subset of the support of $F_X$. Suppose that we have a test with the following property:

$$\sup_{F_X \in \mathcal{F}_{H_0}} \mathbb{P}_{F_X}(\text{ Reject } H_0 \text{ based on } X \mid \text{Select } H_0 \text{ from } X, X \in \mathcal{A}) = \alpha, \quad \text{for all } 0 \leq \alpha \leq 1, \;\; \text{for all } H_0 \in \mathcal{H}_0,$$

where $\mathcal{F}\_H_0$ denotes the set of all distributions that satisfy $H_0$. Show that this test controls the selective type I error rate. That is, show that:

$$\sup_{F_X \in \mathcal{F}_{H_0}} \mathbb{P}_{F_X}(\text{ Reject } H_0 \text{ based on } X \mid \text{Select } H_0 \text{ from } X) = \alpha, \quad \text{for all } 0 \leq \alpha \leq 1, \;\; \text{for all } H_0 \in \mathcal{H}_0.$$

## Problem 2

Let $\tilde{M}(\cdot)$ denote the variable selection procedure that uniformly selects a subset of variables in $\{1, 2, \ldots, p\}$ at random without looking at the data. To put it more formally, for the response variable $Y \sim F_Y$ with support on $\mathbb{R}^n$, we have that $\tilde{M}(Y)$ is independent of $Y$, and that:

$$\mathbb{P}_{F_Y}(\tilde{M}(Y) = M) = 1/|2^{\{1,2,\ldots,p\}}|, \quad \text{for any } M \subseteq \{1, 2, \ldots, p\},$$

where $2^{\{1,2,\ldots,p\}} = \{\emptyset, \{1\}, \ldots, \{p\}, \{1,2\}, \{1,3\}, \ldots, \{p-1, p\}, \ldots, \{1, 2, \ldots, p\}\}$ denotes the set of all possible subsets of $\{1, 2, \ldots, p\}$

Given any realization $y$ from $Y$ and a fixed covariate matrix $X \in \mathbb{R}^{n\times p}$, suppose that we fit a regression model with $y$ as the response and the variables in $\tilde{M}(y)$ as covariates, and report the p-values and rejections without adjusting.

Does this procedure control the selective type I error rate? Specifically, it would control the selective type I error rate if for all $0 \leq \alpha \leq 1$ , and for all $F_Y$ such that $[\beta^*(M)]_j = [(X^TX)^{-1}X^T\mathbb{E}_{F_Y}[Y]]_j = 0$, we have:

$$\mathbb{P}_{F_Y}(\text{Reject } H_0 : [\beta^*(\tilde{M}(Y))]_j = 0 \mid \tilde{M}(Y) = M) \leq \alpha, \quad \text{for all } M \subseteq \{1, 2, \ldots, p\},\ j \in \{1, 2, \ldots, |M|\}.$$

You can assume without argument that for any fixed subset $M \subseteq \{1, 2, \ldots, p\}$, and for any $F_Y$ such that $[\beta^*(M)]_j = [(X^TX)^{-1}X^T\mathbb{E}_{F_Y}[Y]]_j = 0$, we have:

$$\mathbb{P}_{F_Y}(\text{Reject } H_0 : H_0 : [\beta^*(M)]_j = 0) = \alpha, \quad \text{for all } 0 \leq \alpha \leq 1,\ j = 1, 2, \ldots, |M|.$$