Discussion: Multiple hypothesis testing (MHT) in experimental economics

Michael Menietti

2020/03/06

Front matter

Links

- This presentation and sample code can be found at:
  `https://github.com/mmenietti/list_mht`

- Sample code from List, Shaikh, and Xu (2019) can be found at:
  `https://github.com/seidelj/mht`

Overview

Terminology

- Note, herein a hypothesis is a *null* hypothesis.

$$\beta_1 = \beta_2$$

- Typically the "goal" is to *reject* the hypothesis.

$$\beta_1 \neq \beta_2$$

- The chance of rejecting, if the hypothesis is true, is α.
- α is type I error.

The Problem with MHT I

- An experiment can generate several hypotheses
  - Multiple outcomes
  - Heterogeneous treatment effects
- Incorrectly rejecting a single hypothesis is controlled
- Incorrectly rejecting one of multiple tests is *not*

The Problem with MHT II

- Assume n hypotheses are independent.
- A single hypothesis is not rejected by chance with $p_i = 1 - \alpha$.
- All n hypotheses are not rejected by chance with $p = (1 - \alpha)^n$.
- The chance at least 1 type I error occurs is $p = 1 - (1 - \alpha)^n$
- For $\alpha = 0.05, n = 5$ the chance of a type I error is ≈ 0.23

Classic Correction Approaches

- Bonferroni correction, $\tilde{\alpha} = \frac{\alpha}{n}$
- Holm correction, stepwise algorithm, more powerful
- Valid for all error distributions
- Assume worst-case, independent distributions
- Low-power, high chance of type II error

Nonparametric Approaches

- Estimate the distribution of the test statistics
  - Usually bootstrapping, nonparametric method
- Apply a stepwise procedure
- List, Shaikh, and Xu (2019), White (2000), and Romano and Wolf (2005)

MHT Error Types

- Multiple ways to measure MHT errors.
- Non-exhaustively:

Familywise at least 1 type I error

m-familywise m or more type I errors

False discovery rate expected proportion of type I errors

List (2019) Method

Data

- $Y_i \in \mathbb{R}^K$ outcomes for the i^{th} observation
- $D_i \in J_T$ treatment index for the i^{th} observation
- $Z_i \in J_G$ subgroup index for the i^{th} observation
- Note: Only discrete covariates

Hypothesis Tests

$$H\left(d, d', k, z\right) = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} Y_{i,k} - \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} Y_{i,k} = 0$$

$$\mathcal{A} = \left\{ i \in J_N \,\middle|\, D_i = d, Z_i = z \right\}$$

$$\mathcal{B} = \left\{ i \in J_N \,\middle|\, D_i = d', Z_i = z \right\}$$

Test Statistic

$$T_h(Y', D', Z') = \sqrt{n} \left| \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} Y'_{i,k} - \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} Y'_{i,k} \right|$$

$$\tilde{T}_h(Y', D', Z') = \sqrt{n} \left| \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} Y'_{i,k} - \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} Y'_{i,k} - \left(\mu_{\mathcal{A}} - \mu_{\mathcal{B}}\right) \right|$$

$$\mu_{\mathcal{A}} = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} Y_{i,k}, \qquad \mu_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} Y_{i,k}$$

Studentized Test Statistic

$$\lambda_h = \sqrt{n\left(\frac{\sigma_{\mathcal{A}}^2}{n_{\mathcal{A}}} + \frac{\sigma_{\mathcal{B}}^2}{n_{\mathcal{B}}}\right)}$$

$$T_h^{\text{stud}} = \frac{T_h}{\lambda_h}$$

$$\tilde{T}_h^{\text{stud}} = \frac{\tilde{T}_h}{\lambda_h}$$

Algorithm 3.1

1. Calculate test statistics for original data $T_h^0(Y, D, Z)$.
2. Bootstrap the data, generating $T_h^b(T_h^0, Y, D, Z)$.
3. Calculate the empirical distribution values for all test statistics.

$$P_h^0 = \Pr\left(T_h^b \leq T_h^0\right)$$
$$P_h^b = \Pr\left(T_h^{b'} \leq T_h^b\right)$$

4. The modified p-values are calculated against the distribution of the maximum over the single hypothesis distributions.

$$\tilde{P}_{\mathcal{A}}^b = \max_{h \in \mathcal{A}} P_h^b, \qquad \mathcal{A} \subset \mathcal{S}$$
$$\tilde{P}_h^0 = \Pr\left(\tilde{P}_{\mathcal{A}}^b \leq P_h^0\right)$$

Calculating the modified p-values

1. Order hypotheses by the single hypothesis distributions,

$$h_1 < h_2 \iff P_1^0 \leq P_2^0$$

2. The modified p-values are

$$\tilde{P}_1^0 = \Pr\left(\max_{h \in \mathcal{A}} P_h^b < P_1^0\right), \quad \mathcal{A} = \{1, \dots, S\}$$

$$\tilde{P}_2^0 = \Pr\left(\max_{h \in \mathcal{A}} P_h^b < P_2^0\right), \quad \mathcal{A} = \{2, \dots, S\}$$

$$\tilde{P}_i^0 = \Pr\left(\max_{h \in \mathcal{A}} P_h^b < P_i^0\right), \quad \mathcal{A} = \{i, \dots, S\}$$

Remark 3.7

1. There is a more complex procedure (exploiting transitivity)
2. The procedure is not clear
3. Based on the code, it involves enumerating all subsets of hypotheses

References

List, John A., Azeem M. Shaikh, and Yang Xu (Jan. 29, 2019). “Multiple hypothesis testing in experimental economics.” In: *Experimental Economics*.

Romano, Joseph P. and Michael Wolf (Mar. 5, 2005). “Stepwise Multiple Testing as Formalized Data Snooping.” In: *Econometrica*.

White, Halbert (Sept. 1, 2000). “A Reality Check for Data Snooping.” In: *Econometrica*.