

Multiple Outcomes

A 'Reverse Regression' Approach

Permutation Test Overview

Last Updated: October 6, 2017

1 Permutation Tests

The below tests assume a set-up in which there are (i) outcome covariates (Y), (ii) control covariates (X) and (iii) missingness indicators ($X0$). *Unless otherwise stated $X0$ is composed of missingness indicators for the outcome covariates ($X0_Y$) and missingness indicators for the control covariates ($X0_X$).* Where missingness indicators ($X0$) are included in a model *all* observations are used (median imputation), where missingness indicators are not included only the subset of complete observations is used.

1.1 Baseline Balance

The below permutation tests are performed only once, i.e are not dependent on the outcome covariates

- **Baseline Balance (I)** ($H1$: $(X, X0_X)$ jointly predict T better than chance)
 - *Ensemble / Sample Size*: $T = F(X, X0_X)$ / All observations
 - *Permutation*: Permute T against $(X, X0_X)$
 - *Motivation*: Fail to reject $H0 \Rightarrow$ Fail to reject baseline balance (of the control covariates (X) and their missingness ($X0_X$))
- **Baseline Balance (II)** ($H1$: X predicts T better than chance)
 - *Ensemble / Sample Size*: $T = F(X)$ / Complete observations only, i.e. observations with values for *all* control covariates
 - *Permutation*: Permute T against X
 - *Motivation*: Fail to reject $H0 \Rightarrow$ Fail to reject baseline balance (of the control covariates (X))

1.2 Signal

The below permutation tests are performed at the outcome family level, i.e. Tasks #1-#17

- **Signal (I-a)** (*H1: $(Y, X, X0)$ jointly predict T better than chance*)
 - *Ensemble / Sample Size:* $T = F(Y, X, X0)$ / All observations
 - *Permutation:* Permute T against $(Y, X, X0)$
 - *Motivation:* Reject $H0 \Rightarrow$ Predictability (driven by the outcome predictors (Y), the control covariates (X), and/or missingness $X0$ in the controls/outcome covariates (i.e. $X0_X$ or $X0_Y$))
- **Signal (I-b)** (*H1: $(Y, X0_Y)$ jointly predict T better than chance*)
 - *Ensemble / Sample Size:* $T = F(Y, X0_Y)$ / All observations
 - *Permutation:* Permute T against $(Y, X0_Y)$
 - *Motivation:* Reject $H0 \Rightarrow$ Predictability (driven by the outcome predictors (Y) and/or missingness in the outcome covariates ($X0_Y$))
- **Signal (II-a)** (*H1: $(X, X0)$ jointly predict T better than chance*)
 - *Ensemble / Sample Size:* $T = F(X, X0)$ / All observations
 - *Permutation:* Permute T against $(X, X0)$
 - *Motivation:* Reject $H0 \Rightarrow$ Predictability (driven by the control covariates X and/or missingness $X0$ in the controls/outcome covariates (i.e. $X0_X$ or $X0_Y$))
- **Signal (III)** (*H1: $X0_Y$ predicts T better than chance*)
 - *Ensemble / Sample Size:* $T = F(X0_Y)$ / All observations
 - *Permutation:* Permute T against $(X0_Y)$
 - *Motivation:* Reject $H0 \Rightarrow$ Predictability (driven by missingness in the outcome covariates ($X0_Y$))

Note that the idea is to compare (a) *Signal Test (I-b)* vs. *Signal Test (III)* [Basic Analysis] OR (b) *Signal Test (I-a)* vs. *Signal Test (II-a)* vs. *Signal Test (III)* [Extended Analysis]