DEPARTAMENTO DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA DA UNIVERSIDADE DE COIMBRA

*Research Methods*

# Milestone 2 - Pre-registration of hypotheses

Miguel de Oliveira Guerreiro
Luís Espírito Santo
André Carvalho dos Santos

*December 20, 2021*

## 1 INTRODUCTION

We have at our disposal 4 variables:

- the total number of exams, $n \in \{x \in \mathbb{N} : 5 \leq x \leq 18\}$;
- the probability pair of incompatibility for each pair of exams, $p \in [0, 1]$;
- the runtime for "Code 1", $t_1 \in [0, +\infty]$;
- the runtime for "Code 2", $t_2 \in [0, +\infty]$;

**Hypothesis 1** (H1). *For $p = 0.75$ and $n = 15$, we achieve similar runtimes by either using Code 1 or Code 2. We use the following*

$$H_0 : \mu_{t_1} - \mu_{t_2} = 0$$
$$H_1 : \mu_{t_1} - \mu_{t_2} \neq 0$$

**Hypothesis 2** (H2). *For the same problem generated with $p = 0$, Code 1 will take consistently longer time to solve than Code 2. Let $t_1^{(1)}, t_1^{(2)}, t_1^{(3)} \dots t_1^{(n)}$ be measurements for Code 1 and $t_2^{(1)}, t_2^{(2)}, t_2^{(3)} \dots t_2^{(n)}$ measurements for Code 2. Let $d_k = t_1^{(k)} - t_2^{(k)}$,*

$$H_0 : \mu_d \leq 0$$
$$H_1 : \mu_d > 0$$

**Hypothesis 3** (H3). *The linear model we computed from previously collected data is a good model for predicting the runtime of Code 2 for $p = 1$ and given $n$. Our model is $y(n) = -21.89 + 1.92n$. Let $t_1^{(1)}, t_1^{(2)}, t_1^{(3)} \dots t_1^{(m)}$ be measurements for Code 2 with $n^{(1)}, n^{(2)}, n^{(3)} \dots n^{(m)}$ exams. Let $e^k = t^k - y(n^k)$ be the model error, $e \sim N(\mu_e, \sigma_e^2)$:*

$$H_0 : \mu_e = 0$$
$$H_1 : \mu_e \neq 0$$