

# Estatística Multivariada

## Trabalho No.2

Entregar até o dia 23 de maio de 2023.

- 1- Dispomos de um conjunto de dados real com 197 indivíduos e covariáveis. Os pacientes eram uma amostra aleatória de 50% dos pacientes com retinopatia diabética de "alto risco", conforme definido pelo Diabetic Retinopathy Study (DRS), Huster (1989).

Cada paciente teve um olho randomizado para tratamento a laser e o outro olho não recebeu nenhum tratamento. Para cada olho, o evento de interesse foi o tempo desde o início do tratamento até o momento em que a acuidade visual caiu abaixo de 5/200; chamemos de "cegueira". Os tempos neste conjunto de dados são o tempo real até a cegueira em meses. A saída do estudo foi causada por morte, desistência ou término do estudo.

Referência: Huster WJ, Brookmeyer R, Self SG (1989). Modeling paired survival data with covariates. *Biometrics* 45, 145-156.

Dados obtidos como:

```
> dados = read.csv("/home/fernando/Dropbox/Sitio/CE090/20231S/DRS.dat",  
header = TRUE, sep = ",")
```

Um quadro de dados contém 394 observações nas 8 variáveis a seguir.

**Obs** registro de cada linha de observação sem interesse neste estudo

**id** identificador do indivíduo

**ind** indicador de margem, 1 para olho direito, 2 para olho esquerdo

**obs\_time** tempo de cegueira (em meses) desde o tratamento

**status** indicador de censura, 1 = cegueira, 0 = não cegueira

**treat** tipo de tratamento a laser, 0 = sem tratamento, 1 = xenônio, 2 = argônio

**age** idade no diagnóstico de diabetes

**type** tipo de diabetes, 1 = juvenil (idade no tratamento  $\leq$  20), 2 = adulto

Utilize modelos de regressão gaussiana copula considerando que as observações de cada indivíduo, variável **id**, estão correlacionadas.

A resposta neste estudo é **status**, o indicador de censura e queremos saber se o tempo de cegueira desde o tratamento, o tipo de tratamento, a idade de diagnóstico e o tipo de diabetes influenciam a cegueira.

- 2- Queremos analisar as variações no consumo de pacotes de cigarros por mês  $y$ , a variável resposta, em função da marca: A ou B (Brand), do preço (Price) por pacote e em função do gênero do fumante: M ou F (Gender).

Os dados estão em:

```
read.csv(file = "http://leg.ufpr.br/~lucambio/MSM/cigarettes.txt",  
header = TRUE, sep = "")
```

- a) Além dos efeitos da marca, preço e gênero, teste se houver uma interação entre a marca e o preço.
- b) Gostaríamos de introduzir um termo de curvatura para a variável de preço. Como podemos proceder? Teste se esse coeficiente é significativo.
- c) Encontre o valor ótimo da constante  $\lambda$  na regressão ridge aplicada a estes dados.