

# Listas 1 - ME715 2023/2S (Julia)



## Grupo:

- Tiago Henrique da Cruz, RA:206333.
- Marcelo Marcelo Henrique de Jesus, RA: 183168.
- João Vitor Mantovani, RA: 199910.

## Instruções

- A seguinte lista deve ser **resolvida em grupo antes da próxima aula**.
- A **resolução da lista será discutida em sala de aula por algum dos grupos** (a participação será avaliada).
- O grupo/aluno que se negar a participar, terá pontos descontados.
- Os exercícios computacionais **deverão ser resolvidos em R, Python e Julia**. Se os resultados não baterem, o grupo deverá investigar o motivo das diferenças.

**Pacote:** wooldridge

In [188...

```
#Instalação
#using Pkg
#Pkg.add("DataFrames")
#Pkg.add("WooldridgeDatasets")
#Pkg.add("Statistics")
```

In [189...

```
#Carregando
using WooldridgeDatasets
using DataFrames
using Statistics
```

## Questão 1

[C3] Os dados existentes no arquivo MEAP01 são do estado de Michigan no ano de 2001:

O Banco de dados contém 1823 observações e 11 variáveis, desses:

- **dcode**: código do distrito;
- **bcode**: código do edifício;
- **math4**: percentual de estudantes satisfeitos com a matemática da quarta série;
- **read4**: percentual de estudantes satisfeitos com a leitura da quarta série;
- **lunch**: percentual de estudantes com almoço grátis ou desconto;
- **enroll**: matrícula escolar;
- **expend**: gasto total, '\$;
- **exppp**: despesas por aluno, gasto/matricula;
- **lenroll**: log(enroll);
- **lexpend**: log(expend);
- **lexppp**: log(exppp);

In [190...

```
meap01 = DataFrame(wooldridge("meap01"));
```

Use esses dados para responder às seguintes perguntas:

item a)

Encontre os **maiores e menores valores de math4**.

```
In [191... DataFrame(Minima = minimum(meap01.math4),
              Maxima = maximum(meap01.math4))
```

```
Out[191]: 1x2 DataFrame
           Row  Minima  Maxima
           Float64  Float64
1           0.0    100.0
```

item b)

Quantas escolas têm a **taxa de aprovação em matemática de exatamente 50%**?

```
In [192... DataFrame(Quantidade_de_escolas_aprovadas_em_matemática_com_exatamente_50 =
           size(filter(row -> row.math4 == 50, meap01),1))
```

```
Out[192... 1×1 DataFrame
  Row  Quantidade_de_escolas_aprovadas_em_matemática_com_exatamente_50
  Int64
1                                           17
```

item c)

Compare as taxas médias de aprovação em matemática e leitura.

```
In [193... describe(select(meap01, [:math4, :read4]), :mean)
```

```
Out[193... 2x2 DataFrame
```

Row	variable	mean
	Symbol	Float64
1	math4	71.909
2	read4	60.0619

O teste com aprovação mais difícil é a leitura com: 60.062

item d)

Encontre a correlação entre `math4` e `read4`.

```
In [194...   println("Temos a correlação entre math4 e read4 de ",
            round(cor(meap01.math4, meap01.read4), digits = 4))
```

Temos a correlação entre math4 e read4 de 0.8427

## item e)

A variável exppp são os gastos por aluno. **Encontre o exppp médio e seu desvio padrão.**

```
In [195... #describe(select(meap01, :exppp), :mean)
#std(meap01.exppp)

insertcols!(describe(select(meap01, :exppp), :mean),
              3,
              :desvio => std(meap01.exppp)
            )
```

```
Out[195... 1×3 DataFrame
  Row  variable  mean  desvio
  ---  ---
  1    exppp    5194.87 1091.89
```

## item f)

Suponha que a escola A gaste USD\$6.000 por estudante e a escola B gaste USD\$5.500 por estudante. Com que percentual os gastos da escola A superam os da escola B? Compare isso a  $100 \times [\log(6.000) - \log(5.500)]$ , que é a diferença percentual aproximada baseada na diferença dos logaritmos.

```
In [196... Atual = round(100*((6000-5500)/5500), digits = 4)
log_val = round(100*(log(6000)-log(5500)), digits = 4)

println("Temos que os gastos da escola A que superam os da escola B em ",
        round((Atual - log_val), digits=2), "%")
```

Temos que os gastos da escola A que superam os da escola B em 0.39%

## Questão 2

[C8] Os dados em ECONMATH foram obtidos de **estudantes de um grande curso universitário em introdução à microeconomia**. Para este problema, estamos interessados em duas variáveis: **score**, que é a nota do final do curso, e, **econhs**, que é uma variável binária que indica se um estudante fez um curso de economia no ensino médio.

Seja as seguintes variáveis:

- **age**: idade em anos;
- **work**: horas de trabalho por semana;
- **study**: horas de estudo por semana;
- **econhs**: se for 1, estudou economia no ensino médio;
- **colgpa**: faculdade GPA, início do semestre;
- **hsgpa**: nota média do ensino médio;
- **acteng**: pontuação do inglês ACT;
- **actmth**: pontuação da ACT;
- **act**: composto ACT;

- **mathscr**: pontuação do teste de matemática, 0-10;
- **male**: se for 1, masculino;
- **calculus**: se for 1, cursou cálculo;
- **attexc**: se for 1, se passado atndce 'excelente';
- **attgood**: se for 1, se passado atndce 'bom';
- **fathcoll**: se for 1, o pai tem BA;
- **mothcoll**: se for 1, a mãe tem BA;
- **score**: nota do final do curso, em porcentual;

In [197... `econmath = DataFrame(wooldridge("econmath"));`

## item a)

Quantos estudantes estão na amostra? Quantos estudantes declaram ter frequentado um curso de economia no ensino médio?

In [198... 

```
#nrow(econmath)
#nrow(filter(row -> row.econhs == 1, econmath))

DataFrame(Total_de_estudantes = nrow(econmath),
  Alunos_que_cursou_economia_no_ensino_médio = nrow(filter(row -> row.econhs == 1,
```

Out[198... 1×2 DataFrame

Row	Total_de_estudantes	Alunos_que_cursou_economia_no_ensino_médio
	Int64	Int64
1	856	317

## item b)

Encontre a nota média dos alunos que frequentaram um curso de economia do ensino médio. Como se compara com a nota média daqueles que não o fizeram?

In [199... `combine(groupby(econmath, :econhs), :score => mean)`

Out[199... 2×2 DataFrame

Row	econhs	score_mean
	Int64	Float64
1	0	72.9079
2	1	72.0759

In [200... 

```
println("Os estudantes que não fizeram economia no ensino médio \
  teve um desempenho maior daqueles que realizaram, \
  com um diferencial de 0.832.")
```

Os estudantes que não fizeram economia no ensino médio teve um desempenho maior daqueles que realizaram, com um diferencial de 0.832.

## item c)

Os resultados encontrados dizem necessariamente alguma coisa sobre o efeito causal de cursar economia no ensino médio sobre o desempenho no curso universitário? (explique).

Através dos resultados obtidos, não é possível afirmar a existência de um efeito causal em cursar economia no ensino médio sobre o desempenho no curso universitário. Isso se dá devido ao modo em que o experimento foi realizado. O experimento em questão é observacional, o que dificulta o estudo de fatores de confundimento e assim portanto impossibilitando a análise de efeito causal entre os fatores.

### **item d)**

Se quiser obter uma boa estimativa causal do efeito de fazer um curso de economia no ensino médio utilizando a diferença de médias, que experiência faria?

Para o estudo dos efeitos de causalidade, podemos melhorar o experimento de modo a balancear as classes do tratamento de interesse adicionar possíveis fatores de confundimento no experimento. De modo a realizar um experimento balanceado, podemos obter uma amostra com um número igual de alunos que estudaram economia no ensino médio e de alunos que não estudaram. Além da realização de um experimento balanceado, para o estudo de causalidade, é interessante adicionar no conjunto de dados possíveis fatores de confundimento, artigos científico já realizados e relacionados ao assunto seriam consultados.