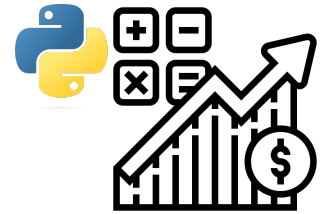


Listas 1 - ME715 2023/2S (Python)



Grupo:

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

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.

Instalando o Pacote: wooldridge

```
In [ ]: # pip install wooldridge
```

Carregando os Pacotes:

```
In [16]: import wooldridge as wd
import pandas as pd
from scipy.stats import pearsonr
import math
```

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 [17]: df = wd.data('meap01')
```

Use esses dados para responder às seguintes perguntas:

item a)

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

```
In [18]: pd.DataFrame({"Minimo":df[["math4"]].min(),  
                      "Maximo":df[["math4"]].max()})
```

```
Out[18]:
```

	Minimo	Maximo
math4	0.0	100.0

item b)

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

```
In [19]: pd.DataFrame({"Quantidade de escolas aprovadas em matemática com exatamente 50%":  
                      [len(df[df.math4 == 50])])})
```

```
Out[19]:
```

Quantidade de escolas aprovadas em matemática com exatamente 50%	
0	17

item c)

Compare as taxas médias de aprovação em matemática e leitura.

```
In [20]: df.loc[:,["math4","read4"]].\  
         agg(["mean"]).\  
         round(3).\  
         rename(index={"mean":"Média"})
```

```
Out[20]:
```

	math4	read4
Média	71.909	60.062

```
In [156... print('O teste com aprovação mais dificil é a leitura com: %.3f' % df["read4"].mean())  
  
O teste com aprovação mais dificil é a leitura com: 60.062
```

item d)

Encontre a correlação entre math4 e read4.

```
In [21]: print('Temos a correlação entre math4 e read4 de: %.4f' %  
              pearsonr(df['math4'],df['read4'])[0])
```

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 [22]: df.loc[:,["exppp"]].\  
         agg(["mean", "std"]).\  
         round(3).\  
         rename(index={"mean":"Médio",  
                      "std":"Desvio padrão"})
```

```
Out[22]:
```

exppp	
Médio	5194.865
Desvio padrão	1091.890

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 [23]: Atual = round(100*((6000-5500)/5500),2)
log_val = round(100*(math.log(6000)-math.log(5500)),2)

print('Temos que os gastos da escola A que superam \
os da escola B em %.2f%%.' %(Atual-log_val))
```

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 [24]: df2 = wd.data('econmath')
```

item a)

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

```
In [25]: def len_econhs_1(x):
        return(
            x.loc[x==1].\
            agg(len)
        )
```

```
df2.loc[:, ["econhs"]].\
    agg([len_econhs_1, len]).\
    rename(index={"len_econhs_1": "Alunos que cursou economia no ensino médio",
                  "len": "Total de estudantes"},
           columns = {"econhs": "Quantidade"})
```

Out[25]:

	Quantidade
Alunos que cursou economia no ensino médio	317
Total de estudantes	856

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 [26]: #df2[df2.econhs == 0].\
#         groupby("econhs")[["score"]].\
#         agg("mean").values[0][0].round(4)

df2.groupby("econhs")[["score"]].\
    agg(["mean"]).\
    round(4).\
    rename(columns = {"score": "Nota",
                      "mean": "Média"},
           index={0: "Não",
                  1: "Sim"})
```

Out[26]:

	Nota	Média
econhs		
Não	72.9079	
Sim	72.0759	

```
In [27]: print('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 q ue 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).

Não, pois para certificar o efeito da causalidade de cursar economia no ensino médio sobre o desempenho no curso universitário, devemos **verificar se existe associação entre elas pelo coeficiente de correlação**.

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?

```
In [35]: print("Como temos %.f" % len(df2),  
            "estudantes e destes, %.f" % len(df2[df2.econhs==1]),  
            "cursarão economia no ensino médio, \  
separaria em duas turmas com %.f" % (len(df2)/2),  
            "alunos para cada turma, das quais aleatoriza a escolha de \  
aproximado 158 alunos que fizeram economia no ensino médio e \  
equivalente 269 alunos que não fizeram economia no ensino médio \  
para cada turma, ao fizar as demais variaveis. Por fim, pelo coefiente \  
de correlação, verificaria entre as variaveis \  
se existe diferenca em relação a média da `score` para cada variaveis.")
```

Como temos 856 estudantes e destes, 317 cursarão economia no ensino médio, separaria em duas turmas com 428 alunos para cada turma, das quais aleatoriza a escolha de aproximado 158 alunos que fizeram economia no ensino médio e equivalente 269 alunos que não fizeram economia no ensino médio para cada turma, ao fizar as demais variaveis. Por fim, pelo coefiente de correlação, verificaria entre as variaveis se existe diferenca em relação a média da `score` para cada variaveis.