

Listas 1 - ME715 2023/2S (Python)



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.

Instalando a biblioteca: wooldridge

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

Carregando as bibliotecas:

```
In [2]: 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 [6]: df = wd.data('meap01')
```

```
Out[6]:
```

	dcode	bcode	math4	read4	lunch	enroll	expend	exppp	lenroll	le
0	1010.0	4937	83.300003	77.800003	40.599998	468	2747475.0	5870.672852	6.148468	14.0
1	2070.0	597	90.300003	82.300003	27.100000	679	1505772.0	2217.631836	6.520621	14.0
2	2080.0	4860	61.900002	71.400002	41.750000	400	2121871.0	5304.677734	5.991465	14.0
3	3010.0	790	85.699997	60.000000	12.750000	251	1211034.0	4824.836426	5.525453	14.0
4	3010.0	1403	77.300003	59.099998	17.080000	439	1913501.0	4358.771973	6.084499	14.0
...
1818	83010.0	2024	93.599998	57.400002	50.529999	285	1484497.0	5208.761230	5.652489	14.0
1819	83010.0	2210	69.000000	48.799999	53.670002	259	1279763.0	4941.169922	5.556828	14.0
1820	83010.0	2480	72.500000	63.400002	46.849998	254	1241369.0	4887.279297	5.537334	14.0
1821	83010.0	5101	81.000000	73.800003	34.299999	277	1312697.0	4738.978516	5.624018	14.0
1822	83070.0	2527	88.099998	56.700001	52.920002	650	1660206.0	2554.163086	6.476973	14.0

1823 rows × 11 columns

Use esses dados para responder às seguintes perguntas:

item a)

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

```
In [9]: pd.DataFrame({"Minimo":df[["math4"]].min(),
                    "Maximo":df[["math4"]].max()})
```

```
Out[9]:
```

	Minimo	Maximo
math4	0.0	100.0

item b)

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

```
In [11]: pd.DataFrame({"Quantidade de escolas aprovadas em matemática com exatamente 50%":
                    [len(df[df.math4 == 50])])})
```

```
Out[11]:
```

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

item c)

Compare as taxas médias de aprovação em matemática e leitura.

```
In [14]: df.loc[:,["math4","read4"]].\
        agg(["mean"]).\
        round(3).\
        rename(index={"mean":"Média"})
```

Out[14]:

	math4	read4
Média	71.909	60.062

```
In [15]: print('O teste com aprovação mais difícil é a leitura com: %.3f' % df["read4"].mean())
```

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

item d)

Encontre a correlação entre math4 e read4.

```
In [16]: 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

Out[16]: (0.8427281457721156, 0.0)

item e)

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

```
In [18]: df.loc[:, ["exppp"]].\
         agg(["mean", "std"]).\
         round(3).\
         rename(index={"mean": "Médio",
                       "std": "Desvio padrão"})
```

Out[18]:

	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 [36]: Atual = 100*((6000-5500)/5500)
log_val = 100*(math.log(6000)-math.log(5500))

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 attndce 'excelente';
- **attgood**: se for 1, se passado attndce '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 [37]: df2 = wd.data('econmath')
```

item a)

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

```
In [42]: 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[42]:
```

	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 [45]: #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[45]:

	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 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.

In []: