



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

Carlos J. Costa

UTILIZAÇÃO DE INTERNET

Passos a seguir

- Importar bibliotecas relevantes
- Importar dados para df
- Converter os dados considerados relevantes
- Fazer tratamento adequado de nulos (ex. Eliminar)
- Identificar VD e VI
- Fazer modelo
- Ajustar e mostrar resultados
- Melhorar modelo

Importar bibliotecas relevantes

- `import pandas as pd`
- `import statsmodels.api as sm`
- `import matplotlib.pyplot as plt`



Importar dados para df

- # Obtido do Banco Mundial
- `df=pd.read_csv('WordBank1.csv')`

Converter os dados considerados relevantes

- `df.dtypes`



Converter os dados considerados relevantes

```
df.loc[:, 'Literacy'] = pd.to_numeric(df['Literacy'], errors='coerce')  
df.loc[:, 'Internetuse'] = pd.to_numeric(df['Internetuse'], errors='coerce')  
df.loc[:, 'SecureServers'] = pd.to_numeric(df['SecureServers'], errors='coerce')  
df.loc[:, 'GDPCapita'] = pd.to_numeric(df['GDPCapita'], errors='coerce')
```

Fazer tratamento adequado de nulos

- Eliminar:

```
XY=df.dropna()
```

- Substituir pela media:

```
XY= df.fillna(df.mean())
```

- Substituir por outro valor

```
XY= df.fillna(valor)
```

Identificar VD e VI

- $y = XY[\text{'Internetuse'}]$
- $X = XY[[\text{'Literacy'}, \text{'SecureServers'}, \text{'GDPCapita'}]]$

Fazer e ajustar Modelo

- `X1=sm.add_constant(X)`
- `model = sm.OLS(y,X1).fit()`
- `model.summary()`

Melhorar Modelo

- Talvez escolher só uma V_i
- Por exemplo logaritmizar V_i e V_d .

Bibliografia

- <https://pandas.pydata.org/>
- https://pandas.pydata.org/pandas-docs/stable/getting_started/10min.html
- <https://scikit-learn.org/>
- <https://scikit-learn.org/stable/index.html>
- <https://www.statsmodels.org/stable/index.html>