Chapter 3 - Linear Regression

September 14, 2023

[97]:
```
!pip install ISLP
```

```
Requirement already satisfied: ISLP in
/Users/barnana/anaconda3/lib/python3.10/site-packages (0.3.16)
Requirement already satisfied: pandas>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (1.5.3)
Requirement already satisfied: numpy>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (1.25.1)
Requirement already satisfied: pygam>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (0.9.0)
Requirement already satisfied: scipy>=0.9 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (1.11.1)
Requirement already satisfied: jupyter>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (1.0.0)
Requirement already satisfied: statsmodels>=0.13 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (0.13.5)
Requirement already satisfied: lifelines>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (0.27.7)
Requirement already satisfied: scikit-learn>=1.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (1.3.0)
Requirement already satisfied: joblib>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (1.1.1)
Requirement already satisfied: matplotlib>=3.3.3 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (3.7.0)
Requirement already satisfied: lxml>=0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from ISLP) (4.9.1)
Requirement already satisfied: ipykernel in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter>=0.0->ISLP)
(6.19.2)
Requirement already satisfied: nbconvert in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter>=0.0->ISLP)
(6.5.4)
Requirement already satisfied: qtconsole in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter>=0.0->ISLP)
(5.4.0)
Requirement already satisfied: jupyter-console in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter>=0.0->ISLP)
(6.6.3)
```

```
Requirement already satisfied: ipywidgets in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter>=0.0->ISLP)
(8.0.7)
Requirement already satisfied: notebook in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter>=0.0->ISLP)
(6.5.2)
Requirement already satisfied: autograd>=1.5 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
lifelines>=0.0->ISLP) (1.6.2)
Requirement already satisfied: formulaic>=0.2.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
lifelines>=0.0->ISLP) (0.6.4)
Requirement already satisfied: autograd-gamma>=0.3 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
lifelines>=0.0->ISLP) (0.5.0)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (2.8.2)
Requirement already satisfied: cycler>=0.10 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (9.4.0)
Requirement already satisfied: packaging>=20.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (22.0)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (1.0.5)
Requirement already satisfied: kiwisolver>=1.0.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (1.4.4)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (3.0.9)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
matplotlib>=3.3.3->ISLP) (4.25.0)
Requirement already satisfied: pytz>=2020.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pandas>=0.0->ISLP)
(2022.7)
Requirement already satisfied: progressbar2<5.0.0,>=4.2.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pygam>=0.0->ISLP)
(4.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from scikit-
learn>=1.2->ISLP) (3.2.0)
```

```
Requirement already satisfied: patsy>=0.5.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
statsmodels>=0.13->ISLP) (0.5.3)
Requirement already satisfied: future>=0.15.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
autograd>=1.5->lifelines>=0.0->ISLP) (0.18.3)
Requirement already satisfied: wrapt>=1.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
formulaic>=0.2.2->lifelines>=0.0->ISLP) (1.14.1)
Requirement already satisfied: typing-extensions>=4.2.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
formulaic>=0.2.2->lifelines>=0.0->ISLP) (4.4.0)
Requirement already satisfied: astor>=0.8 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
formulaic>=0.2.2->lifelines>=0.0->ISLP) (0.8.1)
Requirement already satisfied: interface-meta>=1.2.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
formulaic>=0.2.2->lifelines>=0.0->ISLP) (1.3.0)
Requirement already satisfied: six in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
patsy>=0.5.2->statsmodels>=0.13->ISLP) (1.16.0)
Requirement already satisfied: python-utils>=3.0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
progressbar2<5.0.0,>=4.2.0->pygam>=0.0->ISLP) (3.7.0)
Requirement already satisfied: comm>=0.1.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (0.1.2)
Requirement already satisfied: ipython>=7.23.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (8.10.0)
Requirement already satisfied: pyzmq>=17 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (23.2.0)
Requirement already satisfied: psutil in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (5.9.0)
Requirement already satisfied: traitlets>=5.4.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (5.7.1)
Requirement already satisfied: tornado>=6.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (6.1)
Requirement already satisfied: matplotlib-inline>=0.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (0.1.6)
Requirement already satisfied: nest-asyncio in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (1.5.6)
```

```
Requirement already satisfied: jupyter-client>=6.1.12 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (7.3.4)
Requirement already satisfied: debugpy>=1.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (1.5.1)
Requirement already satisfied: appnope in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipykernel->jupyter>=0.0->ISLP) (0.1.2)
Requirement already satisfied: widgetsnbextension~=4.0.7 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipywidgets->jupyter>=0.0->ISLP) (4.0.8)
Requirement already satisfied: jupyterlab-widgets~=3.0.7 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipywidgets->jupyter>=0.0->ISLP) (3.0.8)
Requirement already satisfied: prompt-toolkit>=3.0.30 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter-
console->jupyter>=0.0->ISLP) (3.0.36)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter-
console->jupyter>=0.0->ISLP) (5.2.0)
Requirement already satisfied: pygments in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter-
console->jupyter>=0.0->ISLP) (2.11.2)
Requirement already satisfied: pandocfilters>=1.4.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (1.5.0)
Requirement already satisfied: tinycss2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (1.2.1)
Requirement already satisfied: entrypoints>=0.2.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (0.4)
Requirement already satisfied: MarkupSafe>=2.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (2.1.1)
Requirement already satisfied: nbformat>=5.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (5.7.0)
Requirement already satisfied: bleach in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (4.1.0)
Requirement already satisfied: jinja2>=3.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (3.1.2)
Requirement already satisfied: beautifulsoup4 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (4.11.1)
```

```
Requirement already satisfied: mistune<2,>=0.8.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (0.8.4)
Requirement already satisfied: nbclient>=0.5.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (0.5.13)
Requirement already satisfied: jupyterlab-pygments in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (0.1.2)
Requirement already satisfied: defusedxml in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbconvert->jupyter>=0.0->ISLP) (0.7.1)
Requirement already satisfied: ipython-genutils in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
notebook->jupyter>=0.0->ISLP) (0.2.0)
Requirement already satisfied: prometheus-client in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
notebook->jupyter>=0.0->ISLP) (0.14.1)
Requirement already satisfied: Send2Trash>=1.8.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
notebook->jupyter>=0.0->ISLP) (1.8.0)
Requirement already satisfied: terminado>=0.8.3 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
notebook->jupyter>=0.0->ISLP) (0.17.1)
Requirement already satisfied: argon2-cffi in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
notebook->jupyter>=0.0->ISLP) (21.3.0)
Requirement already satisfied: nbclassic>=0.4.7 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
notebook->jupyter>=0.0->ISLP) (0.5.2)
Requirement already satisfied: qtpy>=2.0.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
qtconsole->jupyter>=0.0->ISLP) (2.2.0)
Requirement already satisfied: decorator in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (5.1.1)
Requirement already satisfied: jedi>=0.16 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.18.1)
Requirement already satisfied: stack-data in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.2.0)
Requirement already satisfied: pexpect>4.3 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (4.8.0)
Requirement already satisfied: pickleshare in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.7.5)
```

```
Requirement already satisfied: backcall in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.2.0)
Requirement already satisfied: platformdirs>=2.5 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter-
core!=5.0.*,>=4.12->jupyter-console->jupyter>=0.0->ISLP) (2.5.2)
Requirement already satisfied: notebook-shim>=0.1.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbclassic>=0.4.7->notebook->jupyter>=0.0->ISLP) (0.2.2)
Requirement already satisfied: jupyter-server>=1.8 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbclassic>=0.4.7->notebook->jupyter>=0.0->ISLP) (1.23.4)
Requirement already satisfied: fastjsonschema in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbformat>=5.1->nbconvert->jupyter>=0.0->ISLP) (2.16.2)
Requirement already satisfied: jsonschema>=2.6 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
nbformat>=5.1->nbconvert->jupyter>=0.0->ISLP) (4.17.3)
Requirement already satisfied: wcwidth in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from prompt-
toolkit>=3.0.30->jupyter-console->jupyter>=0.0->ISLP) (0.2.5)
Requirement already satisfied: ptyprocess in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
terminado>=0.8.3->notebook->jupyter>=0.0->ISLP) (0.7.0)
Requirement already satisfied: argon2-cffi-bindings in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
argon2-cffi->notebook->jupyter>=0.0->ISLP) (21.2.0)
Requirement already satisfied: soupsieve>1.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
beautifulsoup4->nbconvert->jupyter>=0.0->ISLP) (2.3.2.post1)
Requirement already satisfied: webencodings in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
bleach->nbconvert->jupyter>=0.0->ISLP) (0.5.1)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
jedi>=0.16->ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.8.3)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
jsonschema>=2.6->nbformat>=5.1->nbconvert->jupyter>=0.0->ISLP) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
jsonschema>=2.6->nbformat>=5.1->nbconvert->jupyter>=0.0->ISLP) (22.1.0)
Requirement already satisfied: websocket-client in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter-
server>=1.8->nbclassic>=0.4.7->notebook->jupyter>=0.0->ISLP) (0.58.0)
Requirement already satisfied: anyio<4,>=3.1.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from jupyter-
server>=1.8->nbclassic>=0.4.7->notebook->jupyter>=0.0->ISLP) (3.5.0)
```

```
Requirement already satisfied: cffi>=1.0.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from argon2-cffi-
bindings->argon2-cffi->notebook->jupyter>=0.0->ISLP) (1.15.1)
Requirement already satisfied: executing in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from stack-
data->ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.8.3)
Requirement already satisfied: pure-eval in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from stack-
data->ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (0.2.2)
Requirement already satisfied: asttokens in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from stack-
data->ipython>=7.23.1->ipykernel->jupyter>=0.0->ISLP) (2.0.5)
Requirement already satisfied: sniffio>=1.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
anyio<4,>=3.1.0->jupyter-
server>=1.8->nbclassic>=0.4.7->notebook->jupyter>=0.0->ISLP) (1.2.0)
Requirement already satisfied: idna>=2.8 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
anyio<4,>=3.1.0->jupyter-
server>=1.8->nbclassic>=0.4.7->notebook->jupyter>=0.0->ISLP) (3.4)
Requirement already satisfied: pycparser in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook->jupyter>=0.0->ISLP)
(2.21)
```

[98]:
```
!pip install pytorch-lightning
```

```
Requirement already satisfied: pytorch-lightning in
/Users/barnana/anaconda3/lib/python3.10/site-packages (2.0.6)
Requirement already satisfied: typing-extensions>=4.0.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(4.4.0)
Requirement already satisfied: fsspec[http]>2021.06.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(2022.11.0)
Requirement already satisfied: tqdm>=4.57.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(4.64.1)
Requirement already satisfied: lightning-utilities>=0.7.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(0.9.0)
Requirement already satisfied: PyYAML>=5.4 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(6.0)
Requirement already satisfied: torch>=1.11.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(1.12.1)
Requirement already satisfied: torchmetrics>=0.7.0 in
```

```
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(1.0.1)
Requirement already satisfied: numpy>=1.17.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(1.25.1)
Requirement already satisfied: packaging>=17.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from pytorch-lightning)
(22.0)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
fsspec[http]>2021.06.0->pytorch-lightning) (3.8.5)
Requirement already satisfied: requests in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
fsspec[http]>2021.06.0->pytorch-lightning) (2.28.1)
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (4.0.2)
Requirement already satisfied: multidict<7.0,>=4.5 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (6.0.4)
Requirement already satisfied: aiosignal>=1.1.2 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (1.3.1)
Requirement already satisfied: charset-normalizer<4.0,>=2.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (2.0.4)
Requirement already satisfied: attrs>=17.3.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (22.1.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (1.4.0)
Requirement already satisfied: yarl<2.0,>=1.0 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>2021.06.0->pytorch-lightning) (1.9.2)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
requests->fsspec[http]>2021.06.0->pytorch-lightning) (2023.5.7)
Requirement already satisfied: idna<4,>=2.5 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
requests->fsspec[http]>2021.06.0->pytorch-lightning) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/Users/barnana/anaconda3/lib/python3.10/site-packages (from
requests->fsspec[http]>2021.06.0->pytorch-lightning) (1.26.14)
```

[99]:
```
from ISLP import load_data
import pandas as pd
```

```
import numpy as np
import seaborn as sns
import math
import matplotlib.pyplot as plt
import statsmodels.api as sm
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.model_selection import train_test_split, GridSearchCV, KFold,␣
 ↪cross_val_score
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import RidgeCV
from sklearn.linear_model import LassoCV
from sklearn.decomposition import PCA
from sklearn.pipeline import Pipeline
from sklearn.cross_decomposition import PLSRegression
import itertools
from sklearn.tree import plot_tree
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import BaggingRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
import warnings
warnings.filterwarnings("ignore", category=FutureWarning)
warnings.filterwarnings("ignore", category=DeprecationWarning)
import torch
import torch.nn as nn
import torch.nn.functional as F
from sklearn.metrics import classification_report
```

[100]:
```
# Creating a function to print output in green and bold
# ANSI escape code for green color and bold font
GREEN_BOLD = '\033[1;32m'

# ANSI escape code to reset colors and font style
RESET = '\033[0m'

def print_green_bold(*args):
    text = ' '.join(str(arg) for arg in args)
    print(GREEN_BOLD + text + RESET)
```

1 Chapter 3

1.1 Question 15

This problem involves the Boston data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors. (a) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

[121]:
```
boston_c3q15 = load_data('Boston')
```

[122]:
```
##Dependent and Independent variables
y_c3q15 = boston_c3q15['crim']
X_list_c3q15 = boston_c3q15.drop('crim', axis=1)

n=1

##Create a dataframe to store regression results in a concise manner
results_all_var_c3q15 = pd.DataFrame(columns=['Independent Variable',␣
 ↪'Individual Regression Coefficient', 'P-value'])

# For loop to run a simple linear regression for each independent variable
for x in X_list_c3q15.columns:
    X_c3q15 = sm.add_constant(X_list_c3q15[x])
    lr_model_c3q15 = sm.OLS(y_c3q15, X_c3q15)
    results_c3q15 = lr_model_c3q15.fit()

    # Append results to the DataFrame
    results_all_var_c3q15 = results_all_var_c3q15.append({
        'Independent Variable': X_c3q15.columns[1],
        'Individual Regression Coefficient': results_c3q15.params[1],
        'P-value': round(results_c3q15.pvalues[1],4)
    }, ignore_index=True)

    # Print the regression summary
    print_green_bold(f'Regression Output Model number: {n}')
    print_green_bold(f'Independent variable: {X_c3q15.columns[1]}')
    print_green_bold(f'Coefficient: {results_c3q15.params[1]:.4f}')
    print_green_bold(f'P-value: {results_c3q15.pvalues[1]:.4f}')
    print_green_bold('----------------------------------------')
    print_green_bold(results_c3q15.summary())
    print_green_bold('----------------------------------------\n\n\n\n')
    n=n+1
```

```
Regression Output Model number: 1
Independent variable: zn
```

```
Coefficient: -0.0739
P-value: 0.0000
----------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                   crim   R-squared:                       0.040
Model:                            OLS   Adj. R-squared:                  0.038
Method:                 Least Squares   F-statistic:                     21.10
Date:                Sun, 30 Jul 2023   Prob (F-statistic):           5.51e-06
Time:                        15:19:12   Log-Likelihood:                -1796.0
No. Observations:                 506   AIC:                             3596.
Df Residuals:                     504   BIC:                             3604.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.4537      0.417     10.675      0.000       3.634       5.273
zn            -0.0739      0.016     -4.594      0.000      -0.106      -0.042
==============================================================================
Omnibus:                      567.443   Durbin-Watson:                   0.857
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            32753.004
Skew:                           5.257   Prob(JB):                         0.00
Kurtosis:                      40.986   Cond. No.                         28.8
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
----------------------------------------


Regression Output Model number: 2
Independent variable: indus
Coefficient: 0.5098
P-value: 0.0000
----------------------------------------
```

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.165
Model:	OLS	Adj. R-squared:	0.164
Method:	Least Squares	F-statistic:	99.82
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	1.45e-21
Time:	15:19:12	Log-Likelihood:	-1760.6
No. Observations:	506	AIC:	3525.
Df Residuals:	504	BIC:	3534.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.0637	0.667	-3.093	0.002	-3.375	-0.753
indus	0.5098	0.051	9.991	0.000	0.410	0.610

Omnibus:	585.118	Durbin-Watson:	0.986
Prob(Omnibus):	0.000	Jarque-Bera (JB):	41418.938
Skew:	5.449	Prob(JB):	0.00
Kurtosis:	45.962	Cond. No.	25.1

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 3
Independent variable: chas
Coefficient: -1.8928
P-value: 0.2094

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.003
Model:	OLS	Adj. R-squared:	0.001
Method:	Least Squares	F-statistic:	1.579
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	0.209
Time:	15:19:12	Log-Likelihood:	-1805.6
No. Observations:	506	AIC:	3615.
Df Residuals:	504	BIC:	3624.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.7444	0.396	9.453	0.000	2.966	4.523
chas	-1.8928	1.506	-1.257	0.209	-4.852	1.066

Omnibus:	561.663	Durbin-Watson:	0.817
Prob(Omnibus):	0.000	Jarque-Bera (JB):	30645.429
Skew:	5.191	Prob(JB):	0.00
Kurtosis:	39.685	Cond. No.	3.96

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 4
Independent variable: nox
Coefficient: 31.2485
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.177
Model:	OLS	Adj. R-squared:	0.176
Method:	Least Squares	F-statistic:	108.6
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	3.75e-23
Time:	15:19:12	Log-Likelihood:	-1757.0
No. Observations:	506	AIC:	3518.
Df Residuals:	504	BIC:	3526.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-13.7199	1.699	-8.073	0.000	-17.059	-10.381
nox	31.2485	2.999	10.419	0.000	25.356	37.141

Omnibus:	591.712	Durbin-Watson:	0.992
Prob(Omnibus):	0.000	Jarque-Bera (JB):	43138.106
Skew:	5.546	Prob(JB):	0.00
Kurtosis:	46.852	Cond. No.	11.3

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 5
Independent variable: rm
Coefficient: -2.6841
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.048
Model:	OLS	Adj. R-squared:	0.046
Method:	Least Squares	F-statistic:	25.45
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	6.35e-07
Time:	15:19:12	Log-Likelihood:	-1793.9
No. Observations:	506	AIC:	3592.
Df Residuals:	504	BIC:	3600.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	20.4818	3.364	6.088	0.000	13.872	27.092
rm	-2.6841	0.532	-5.045	0.000	-3.729	-1.639

Omnibus:	575.717	Durbin-Watson:	0.879
Prob(Omnibus):	0.000	Jarque-Bera (JB):	36658.093
Skew:	5.345	Prob(JB):	0.00
Kurtosis:	43.305	Cond. No.	58.4

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 6
Independent variable: age
Coefficient: 0.1078
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.124
Model:	OLS	Adj. R-squared:	0.123
Method:	Least Squares	F-statistic:	71.62
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.85e-16
Time:	15:19:12	Log-Likelihood:	-1772.7
No. Observations:	506	AIC:	3549.
Df Residuals:	504	BIC:	3558.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.7779	0.944	-4.002	0.000	-5.633	-1.923
age	0.1078	0.013	8.463	0.000	0.083	0.133

Omnibus:	574.509	Durbin-Watson:	0.956
Prob(Omnibus):	0.000	Jarque-Bera (JB):	36741.903
Skew:	5.322	Prob(JB):	0.00
Kurtosis:	43.366	Cond. No.	195.

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 7
Independent variable: dis
Coefficient: -1.5509
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.144
Model:	OLS	Adj. R-squared:	0.142
Method:	Least Squares	F-statistic:	84.89
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	8.52e-19
Time:	15:19:12	Log-Likelihood:	-1767.0
No. Observations:	506	AIC:	3538.
Df Residuals:	504	BIC:	3546.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	9.4993	0.730	13.006	0.000	8.064	10.934
dis	-1.5509	0.168	-9.213	0.000	-1.882	-1.220

Omnibus:	576.519	Durbin-Watson:	0.952
Prob(Omnibus):	0.000	Jarque-Bera (JB):	37426.729
Skew:	5.348	Prob(JB):	0.00
Kurtosis:	43.753	Cond. No.	9.32

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 8
Independent variable: rad
Coefficient: 0.6179
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.391
Model:	OLS	Adj. R-squared:	0.390
Method:	Least Squares	F-statistic:	323.9
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.69e-56
Time:	15:19:12	Log-Likelihood:	-1680.8
No. Observations:	506	AIC:	3366.
Df Residuals:	504	BIC:	3374.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.2872	0.443	-5.157	0.000	-3.158	-1.416
rad	0.6179	0.034	17.998	0.000	0.550	0.685

Omnibus:	656.459	Durbin-Watson:	1.337
Prob(Omnibus):	0.000	Jarque-Bera (JB):	75417.007
Skew:	6.478	Prob(JB):	0.00
Kurtosis:	61.389	Cond. No.	19.2

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 9
Independent variable: tax
Coefficient: 0.0297
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.340
Model:	OLS	Adj. R-squared:	0.338
Method:	Least Squares	F-statistic:	259.2
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.36e-47
Time:	15:19:12	Log-Likelihood:	-1701.4
No. Observations:	506	AIC:	3407.
Df Residuals:	504	BIC:	3415.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-8.5284	0.816	-10.454	0.000	-10.131	-6.926
tax	0.0297	0.002	16.099	0.000	0.026	0.033

Omnibus:	635.377	Durbin-Watson:	1.252
Prob(Omnibus):	0.000	Jarque-Bera (JB):	63763.835
Skew:	6.156	Prob(JB):	0.00
Kurtosis:	56.599	Cond. No.	1.16e+03

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.16e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Regression Output Model number: 10
Independent variable: ptratio
Coefficient: 1.1520
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.084
Model:	OLS	Adj. R-squared:	0.082
Method:	Least Squares	F-statistic:	46.26
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.94e-11
Time:	15:19:12	Log-Likelihood:	-1784.1
No. Observations:	506	AIC:	3572.
Df Residuals:	504	BIC:	3581.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-17.6469	3.147	-5.607	0.000	-23.830	-11.464
ptratio	1.1520	0.169	6.801	0.000	0.819	1.485

Omnibus:	568.053	Durbin-Watson:	0.905
Prob(Omnibus):	0.000	Jarque-Bera (JB):	34221.853
Skew:	5.245	Prob(JB):	0.00
Kurtosis:	41.899	Cond. No.	160.

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 11
Independent variable: lstat
Coefficient: 0.5488
P-value: 0.0000

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.208
Model:	OLS	Adj. R-squared:	0.206
Method:	Least Squares	F-statistic:	132.0
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.65e-27
Time:	15:19:12	Log-Likelihood:	-1747.5
No. Observations:	506	AIC:	3499.
Df Residuals:	504	BIC:	3507.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.3305	0.694	-4.801	0.000	-4.694	-1.968
lstat	0.5488	0.048	11.491	0.000	0.455	0.643

Omnibus:	601.306	Durbin-Watson:	1.182
Prob(Omnibus):	0.000	Jarque-Bera (JB):	49918.826
Skew:	5.645	Prob(JB):	0.00
Kurtosis:	50.331	Cond. No.	29.7

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Regression Output Model number: 12
Independent variable: medv
Coefficient: -0.3632
P-value: 0.0000

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   crim   R-squared:                       0.151
Model:                            OLS   Adj. R-squared:                  0.149
Method:                 Least Squares   F-statistic:                     89.49
Date:                Sun, 30 Jul 2023   Prob (F-statistic):           1.17e-19
Time:                        15:19:12   Log-Likelihood:                -1765.0
No. Observations:                 506   AIC:                             3534.
Df Residuals:                     504   BIC:                             3542.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         11.7965      0.934     12.628      0.000       9.961      13.632
medv          -0.3632      0.038     -9.460      0.000      -0.439      -0.288
==============================================================================
Omnibus:                      558.880   Durbin-Watson:                   0.996
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            32740.044
Skew:                           5.108   Prob(JB):                         0.00
Kurtosis:                      41.059   Cond. No.                         64.5
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
----------------------------------------
```

Let's take a look at the results_all_var dataframe.

[123]:
```
results_all_var_c3q15
```

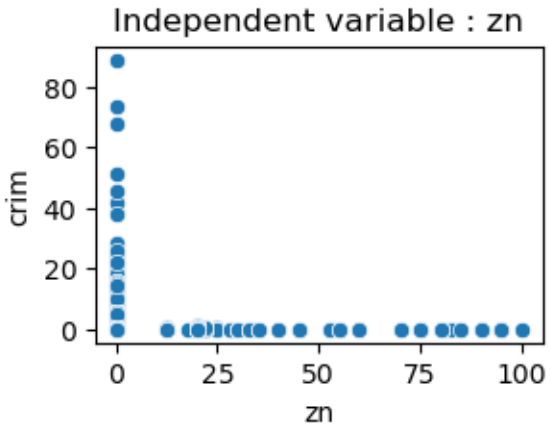
[123]:

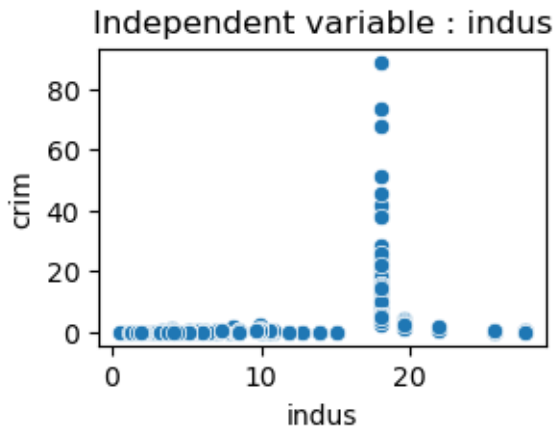
	Independent Variable	Individual Regression Coefficient	P-value
0	zn	-0.073935	0.0000
1	indus	0.509776	0.0000

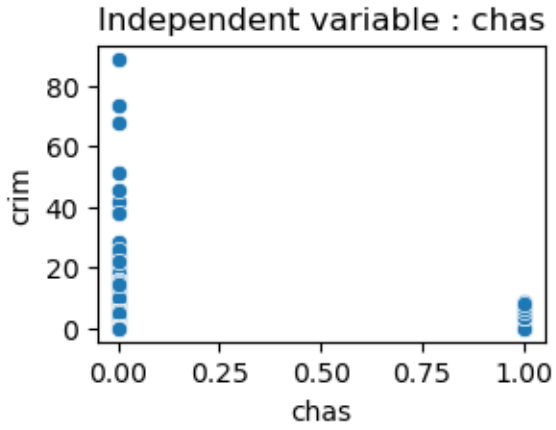
2	chas	-1.892777	0.2094
3	nox	31.248531	0.0000
4	rm	-2.684051	0.0000
5	age	0.107786	0.0000
6	dis	-1.550902	0.0000
7	rad	0.617911	0.0000
8	tax	0.029742	0.0000
9	ptratio	1.151983	0.0000
10	lstat	0.548805	0.0000
11	medv	-0.363160	0.0000

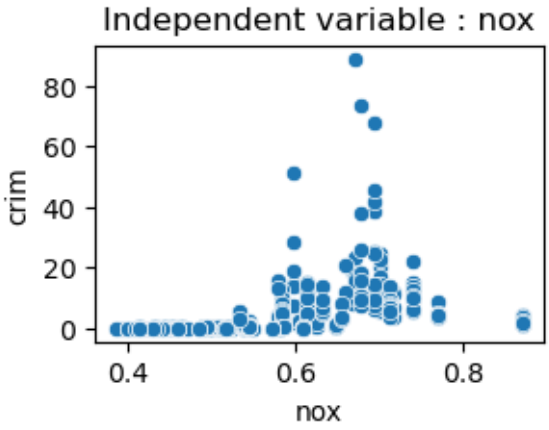
All independent variables except chas have very low p-values. This means that all of them except chas are individually statistically significant (at both 5% and 1% level of significance) with regards to their impact on crim. Let's look at pairplots between crim and all the independent variables. We expect to see a pattern in all plots except the one with chas. We'll look at a boxplot to visualize the distribution of crim for different categories of chas. Since chas is categorical, we expect to see two boxes representing the distribution of crim for each category. If there is a notable difference in the median or spread of crim between the two categories, it could indicate a potential impact of chas on crim.

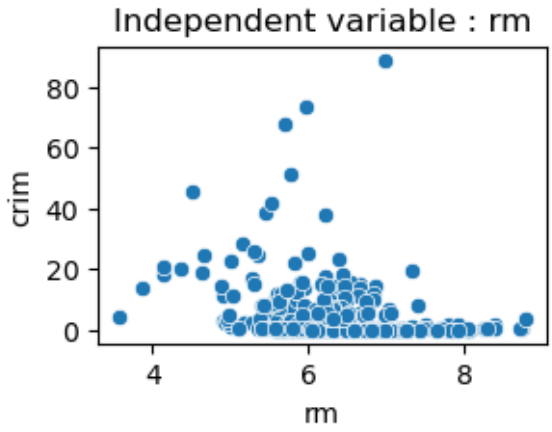
[124]:
```
for x in X_list_c3q15.columns:
    plt.figure(figsize=(3, 2))
    pplot= sns.scatterplot(x=x, y='crim', data=boston_c3q15)
    pplot.set_title(f'Independent variable : {x}')
    plt.show()
```

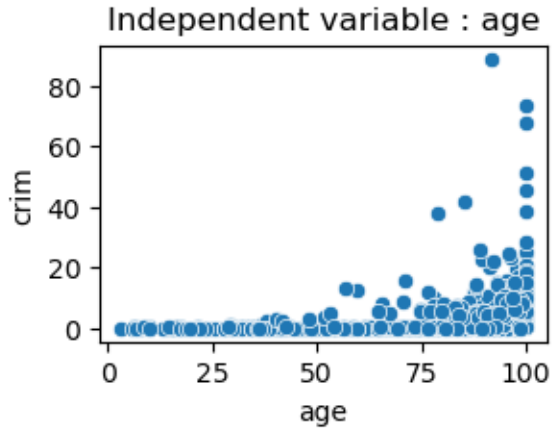


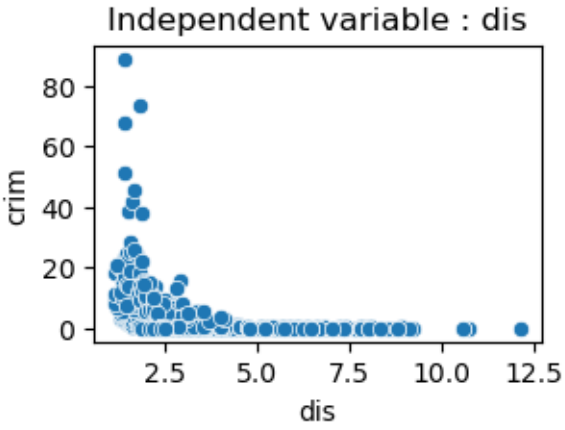

Independent variable : indus
crim
indus
0
20
40
60
80
0
10
20

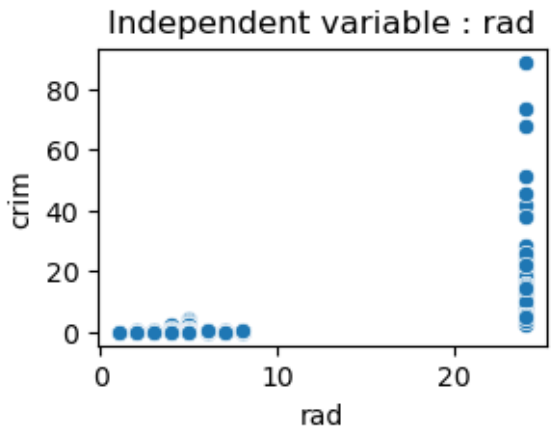

Independent variable : chas
crim
chas
0
20
40
60
80
0.00
0.25
0.50
0.75
1.00

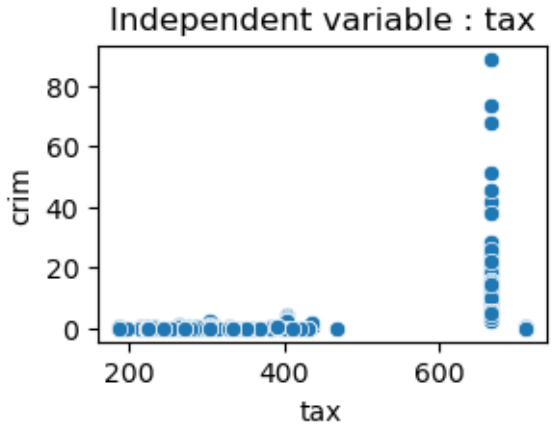

Independent variable : nox
crim
nox
0
20
40
60
80
0.4
0.6
0.8

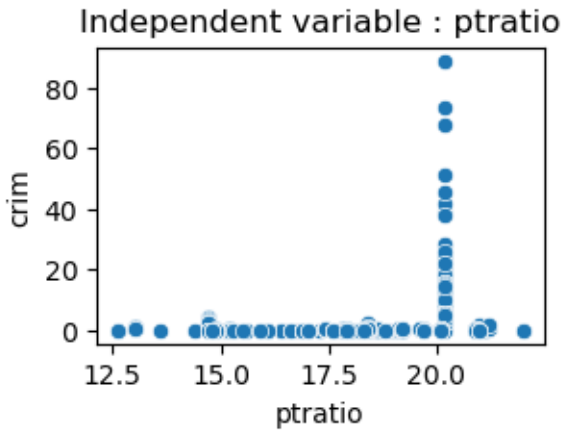

Independent variable : rm
crim
rm
0
20
40
60
80
4
6
8

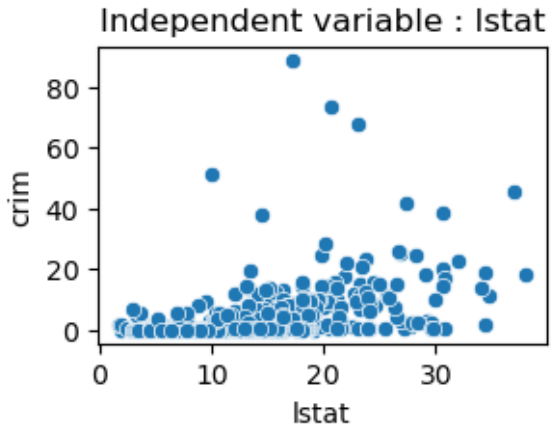

Independent variable : age
crim
age
0
20
40
60
80
0
25
50
75
100

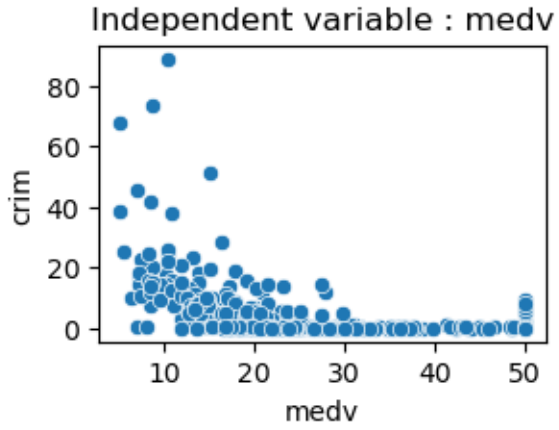

Independent variable : dis
crim
dis
0
20
40
60
80
2.5
5.0
7.5
10.0
12.5


Independent variable : rad
crim
rad
0
20
40
60
80
0
10
20


Independent variable : tax
crim
tax
0
20
40
60
80
200
400
600

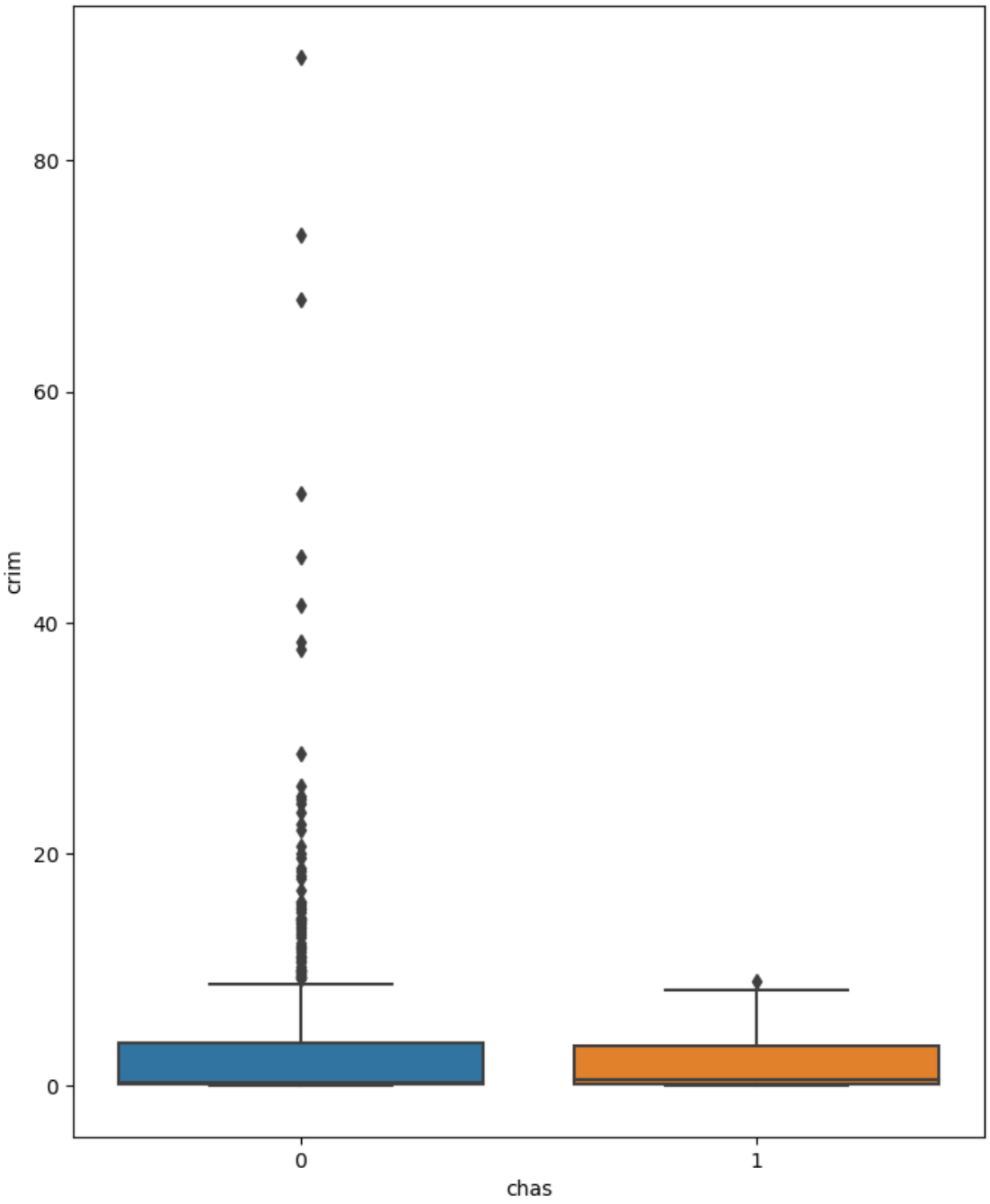

Independent variable : ptratio
crim
ptratio
0
20
40
60
80
12.5
15.0
17.5
20.0





As expected, except for chas, variables such as medv, lstat, nox, age, etc. appear to show a relationship with the crim variable.

[125]:
```
# Box plot for chas
plt.figure(figsize=(8, 10))
sns.boxplot(x='chas', y='crim', data=boston_c3q15)
plt.xlabel('chas')
plt.ylabel('crim')
plt.show()
```



The boxes for both categories (0 and 1) of chas are similar which suggests that the median and the interquartile range (IQR) of crim are comparable for both categories. In other words, the central tendency of crim variable is not significantly different between the two chas categories. There are also lots of points outside the whiskers for category=0, which means that there are many outliers for crim when chas=0. Based on the data at hand, it appears that chas does not have a significant impact on crim.

(b) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis H0 : j = 0?

[126]:
```
X_all_c3q15=X_list_c3q15.copy()

# Add a constant to X_all_c3q15 for multiple regression
X_all_c3q15= sm.add_constant(X_all_c3q15)

mlr_model_c3q15 = sm.OLS(y_c3q15, X_all_c3q15)
results_mlr_c3q15 = mlr_model_c3q15.fit()

mlr_coefficients_c3q15 = pd.DataFrame({'Independent Variable': X_all_c3q15.
 ↪columns,
                                       'Multiple Regression Coefficient':␣
 ↪results_mlr_c3q15.params,
                                       'P-value_mlr': results_mlr_c3q15.
 ↪pvalues})

# Print the model summary
print_green_bold(results_mlr_c3q15.summary())
```

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.449
Model:	OLS	Adj. R-squared:	0.436
Method:	Least Squares	F-statistic:	33.52
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.03e-56
Time:	15:19:34	Log-Likelihood:	-1655.4
No. Observations:	506	AIC:	3337.
Df Residuals:	493	BIC:	3392.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	13.7784	7.082	1.946	0.052	-0.136	27.693
zn	0.0457	0.019	2.433	0.015	0.009	0.083
indus	-0.0584	0.084	-0.698	0.486	-0.223	0.106
chas	-0.8254	1.183	-0.697	0.486	-3.150	1.500
nox	-9.9576	5.290	-1.882	0.060	-20.351	0.436
rm	0.6289	0.607	1.036	0.301	-0.564	1.822
age	-0.0008	0.018	-0.047	0.962	-0.036	0.034
dis	-1.0122	0.282	-3.584	0.000	-1.567	-0.457
rad	0.6125	0.088	6.997	0.000	0.440	0.784
tax	-0.0038	0.005	-0.730	0.466	-0.014	0.006
ptratio	-0.3041	0.186	-1.632	0.103	-0.670	0.062
lstat	0.1388	0.076	1.833	0.067	-0.010	0.288
medv	-0.2201	0.060	-3.678	0.000	-0.338	-0.103

Omnibus:	663.436	Durbin-Watson:	1.516
Prob(Omnibus):	0.000	Jarque-Bera (JB):	80856.852
Skew:	6.579	Prob(JB):	0.00
Kurtosis:	63.514	Cond. No.	1.24e+04

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Only zn, dis, rad and medv have p-values less than 0.05 and are therefore, statistically significant at a 5% level. This means at a 5% level of signficance, we can reject the null hypothesis for these variables. Out of these, dis, rad and medv are also significant at a 1% level of significance and we can reject the null hypothesis for them at 1%.

[127]:
```
# Checking if the results change if I do feature scaling
X_all_t_c3q15 = X_list_c3q15.copy()
scaler = StandardScaler()
X_all_scaled_c3q15 = scaler.fit_transform(X_all_t_c3q15)
X_all_scaled_c3q15 = pd.DataFrame(X_all_scaled_c3q15, columns=X_all_t_c3q15.
 ↪columns)

X_all_scaled_c3q15 = sm.add_constant(X_all_scaled_c3q15)

mlr_model_t_c3q15 = sm.OLS(y_c3q15, X_all_scaled_c3q15)
results_mlr_t_c3q15 = mlr_model_t_c3q15.fit()

print_green_bold(results_mlr_t_c3q15.summary())
```

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.449
Model:	OLS	Adj. R-squared:	0.436
Method:	Least Squares	F-statistic:	33.52
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.03e-56
Time:	15:19:38	Log-Likelihood:	-1655.4
No. Observations:	506	AIC:	3337.
Df Residuals:	493	BIC:	3392.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.6135	0.287	12.582	0.000	3.049	4.178
zn	1.0650	0.438	2.433	0.015	0.205	1.925
indus	-0.3999	0.573	-0.698	0.486	-1.526	0.726
chas	-0.2094	0.300	-0.697	0.486	-0.799	0.381
nox	-1.1527	0.612	-1.882	0.060	-2.356	0.050
rm	0.4414	0.426	1.036	0.301	-0.396	1.279
age	-0.0239	0.505	-0.047	0.962	-1.016	0.968
dis	-2.1294	0.594	-3.584	0.000	-3.297	-0.962
rad	5.3276	0.761	6.997	0.000	3.832	6.824
tax	-0.6357	0.871	-0.730	0.466	-2.347	1.075
ptratio	-0.6577	0.403	-1.632	0.103	-1.450	0.134
lstat	0.9902	0.540	1.833	0.067	-0.071	2.052
medv	-2.0219	0.550	-3.678	0.000	-3.102	-0.942

Omnibus:	663.436	Durbin-Watson:	1.516
Prob(Omnibus):	0.000	Jarque-Bera (JB):	80856.852
Skew:	6.579	Prob(JB):	0.00
Kurtosis:	63.514	Cond. No.	9.81

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Note that on using scaled features, the results remain the same with regards to statistical significance.

(c) How do your results from (a) compare to your results from (b)? Create a plot displaying the univariate regression coefficients from (a) on the x-axis, and the multiple regression coefficients from (b) on the y-axis. That is, each predictor is displayed as a single point in the plot. Its coefficient in a simple linear regres- sion model is shown on the x-axis, and its coefficient estimate in the multiple linear regression model is shown on the y-axis.

[128]:
```
# Individual regression coefficients
lr_coefficients_c3q15 = results_all_var_c3q15.copy()
lr_coefficients_c3q15 = lr_coefficients_c3q15.rename(columns={'P-value':␣
 ↪'P-value_lr'})


# Getting both individual and multiple regression coefficients (unscaled) in␣
 ↪the same table
both_coefficients_c3q15 = pd.merge(lr_coefficients_c3q15,␣
 ↪mlr_coefficients_c3q15, on='Independent Variable')

# Print the final merged table
both_coefficients_c3q15
```

[128]:

	Independent Variable	Individual Regression Coefficient	P-value_lr	\
0	zn	-0.073935	0.0000	
1	indus	0.509776	0.0000	
2	chas	-1.892777	0.2094	
3	nox	31.248531	0.0000	
4	rm	-2.684051	0.0000	
5	age	0.107786	0.0000	
6	dis	-1.550902	0.0000	
7	rad	0.617911	0.0000	
8	tax	0.029742	0.0000	
9	ptratio	1.151983	0.0000	
10	lstat	0.548805	0.0000	
11	medv	-0.363160	0.0000	

	Multiple Regression Coefficient	P-value_mlr
0	0.045710	1.534403e-02
1	-0.058350	4.857094e-01
2	-0.825378	4.858406e-01
3	-9.957587	6.036986e-02
4	0.628911	3.007385e-01
5	-0.000848	9.623231e-01
6	-1.012247	3.725942e-04
7	0.612465	8.588123e-12
8	-0.003776	4.657565e-01

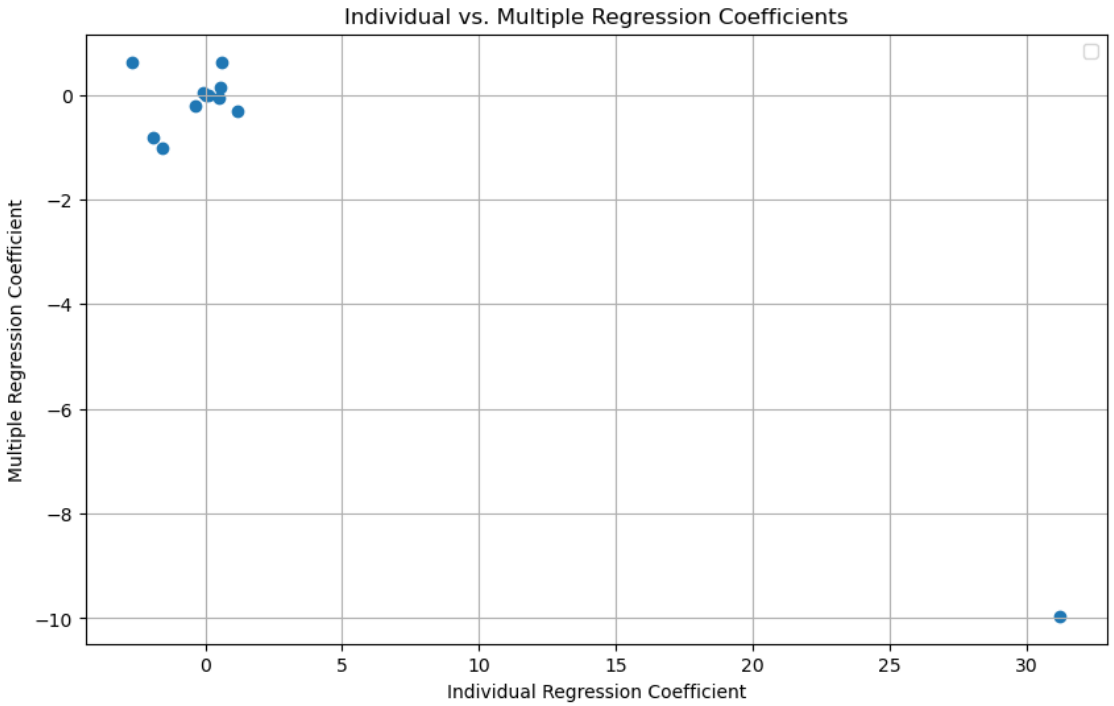
```
9                          -0.304073  1.033932e-01
10                          0.138801  6.739844e-02
11                         -0.220056  2.605302e-04
```

Observations:

- For some variables (chas, dis, rad, medv), the sign of the regression coefficient remains the same between individual and multiple regression analysis, while the magnitude may change.
- For other variables (zn, indus, nox, rm, age, tax, ptratio), the sign of the regression coefficient changes between individual and multiple regression analysis.
- The p-values also change between individual and multiple regression analyses. The variables indus, chas, nox, rm, age, tax, ptratio, stat were no longer statistically significant in a multiple regression. Only zn, dis, rad and medv continued to be significant when we ran a multiple regresison model. The sign of the coefficient for zn changed but that for dis, rad and medv remained the same.

[129]:
```
plt.figure(figsize=(10, 6))
plt.scatter(both_coefficients_c3q15['Individual Regression Coefficient'],␣
 ↪both_coefficients_c3q15['Multiple Regression Coefficient'])
plt.xlabel('Individual Regression Coefficient')
plt.ylabel('Multiple Regression Coefficient')
plt.title('Individual vs. Multiple Regression Coefficients')
plt.legend()
plt.grid(True)
plt.show()
```

```
No artists with labels found to put in legend.  Note that artists whose label
start with an underscore are ignored when legend() is called with no argument.
```



(d) Is there evidence of non-linear association between any of the predictors and the response? To answer this question, for each predictor X, fit a model of the form Y = 0 + 1X+ 2X2 + 3X3 + .

[130]:
```
# Fitting the model for each predictor
n=1
for x in X_list_c3q15.columns:
    X_c3q15 = pd.DataFrame({
        x: X_list_c3q15[x],
         f'{x}^2': X_list_c3q15[x] ** 2,
         f'{x}^3': X_list_c3q15[x] ** 3
    })

    X_c3q15 = sm.add_constant(X_c3q15)


    non_linear_model_c3q15 = sm.OLS(y_c3q15, X_c3q15)
    non_linear_results_c3q15 = non_linear_model_c3q15.fit()
    print_green_bold(f'Non Linear Regression Output Model number: {n}')
    print_green_bold(f'Independent variable: {X_c3q15.columns[1]}')
    print_green_bold('----------------------------------------')
    print_green_bold(non_linear_results_c3q15.summary())
    print_green_bold('----------------------------------------\n\n\n\n')
    n=n+1
```

Non Linear Regression Output Model number: 1
Independent variable: zn

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.058
Model:	OLS	Adj. R-squared:	0.053
Method:	Least Squares	F-statistic:	10.35
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	1.28e-06
Time:	15:19:49	Log-Likelihood:	-1791.2
No. Observations:	506	AIC:	3590.
Df Residuals:	502	BIC:	3607.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.8461	0.433	11.192	0.000	3.995	5.697
zn	-0.3322	0.110	-3.025	0.003	-0.548	-0.116
zn^2	0.0065	0.004	1.679	0.094	-0.001	0.014
zn^3	-3.776e-05	3.14e-05	-1.203	0.230	-9.94e-05	2.39e-05

Omnibus:	569.133	Durbin-Watson:	0.875
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33700.991
Skew:	5.272	Prob(JB):	0.00
Kurtosis:	41.565	Cond. No.	1.89e+05

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.89e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
----------------------------------------


Non Linear Regression Output Model number: 2
Independent variable: indus
----------------------------------------
```

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.260
Model:	OLS	Adj. R-squared:	0.255
Method:	Least Squares	F-statistic:	58.69
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	1.55e-32
Time:	15:19:49	Log-Likelihood:	-1730.3
No. Observations:	506	AIC:	3469.
Df Residuals:	502	BIC:	3486.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.6626	1.574	2.327	0.020	0.570	6.755
indus	-1.9652	0.482	-4.077	0.000	-2.912	-1.018
indus^2	0.2519	0.039	6.407	0.000	0.175	0.329
indus^3	-0.0070	0.001	-7.292	0.000	-0.009	-0.005

Omnibus:	611.788	Durbin-Watson:	1.116
Prob(Omnibus):	0.000	Jarque-Bera (JB):	51742.286
Skew:	5.820	Prob(JB):	0.00
Kurtosis:	51.153	Cond. No.	2.47e+04

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.47e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Non Linear Regression Output Model number: 3
Independent variable: chas

```
----------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                    crim   R-squared:                       0.003
Model:                             OLS   Adj. R-squared:                 -0.001
Method:                  Least Squares   F-statistic:                    0.7881
Date:                 Sun, 30 Jul 2023   Prob (F-statistic):              0.455
Time:                         15:19:49   Log-Likelihood:                -1805.6
No. Observations:                  506   AIC:                             3617.
Df Residuals:                      503   BIC:                             3630.
Df Model:                            2
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.7444      0.397      9.444      0.000       2.965       4.523
chas        -4.77e+13   1.15e+14     -0.414      0.679   -2.74e+14    1.79e+14
chas^2      2.385e+13   5.77e+13      0.414      0.679   -8.95e+13    1.37e+14
chas^3      2.385e+13   5.77e+13      0.414      0.679   -8.95e+13    1.37e+14
==============================================================================
Omnibus:                      561.675   Durbin-Watson:                   0.817
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            30647.554
Skew:                           5.191   Prob(JB):                         0.00
Kurtosis:                      39.686   Cond. No.                     5.48e+28
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.71e-55. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
----------------------------------------

Non Linear Regression Output Model number: 4
```

```
Independent variable: nox
---------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                   crim   R-squared:                       0.297
Model:                            OLS   Adj. R-squared:                  0.293
Method:                 Least Squares   F-statistic:                     70.69
Date:                Sun, 30 Jul 2023   Prob (F-statistic):           3.81e-38
Time:                        15:19:49   Log-Likelihood:                -1717.2
No. Observations:                 506   AIC:                             3442.
Df Residuals:                     502   BIC:                             3459.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        233.0866     33.643      6.928      0.000     166.988     299.185
nox        -1279.3713    170.397     -7.508      0.000   -1614.151    -944.591
nox^2       2248.5441    279.899      8.033      0.000    1698.626    2798.462
nox^3      -1245.7029    149.282     -8.345      0.000   -1538.997    -952.409
==============================================================================
Omnibus:                      614.412   Durbin-Watson:                   1.159
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            53523.997
Skew:                           5.851   Prob(JB):                         0.00
Kurtosis:                      52.008   Cond. No.                     1.36e+03
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.36e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
---------------------------------------
```

Non Linear Regression Output Model number: 5
Independent variable: rm

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.068
Model:	OLS	Adj. R-squared:	0.062
Method:	Least Squares	F-statistic:	12.17
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	1.07e-07
Time:	15:19:49	Log-Likelihood:	-1788.6
No. Observations:	506	AIC:	3585.
Df Residuals:	502	BIC:	3602.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	112.6246	64.517	1.746	0.081	-14.132	239.382
rm	-39.1501	31.311	-1.250	0.212	-100.668	22.368
rm^2	4.5509	5.010	0.908	0.364	-5.292	14.394
rm^3	-0.1745	0.264	-0.662	0.509	-0.693	0.344

Omnibus:	585.097	Durbin-Watson:	0.913
Prob(Omnibus):	0.000	Jarque-Bera (JB):	40144.207
Skew:	5.465	Prob(JB):	0.00
Kurtosis:	45.245	Cond. No.	5.36e+04

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 5.36e+04. This might indicate that there are strong multicollinearity or other numerical problems.

```
----------------------------------------


Non Linear Regression Output Model number: 6
Independent variable: age
----------------------------------------
```

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.174
Model:	OLS	Adj. R-squared:	0.169
Method:	Least Squares	F-statistic:	35.31
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	1.02e-20
Time:	15:19:49	Log-Likelihood:	-1757.9
No. Observations:	506	AIC:	3524.
Df Residuals:	502	BIC:	3541.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.5488	2.769	-0.920	0.358	-7.989	2.892
age	0.2737	0.186	1.468	0.143	-0.093	0.640
age^2	-0.0072	0.004	-1.988	0.047	-0.014	-8.4e-05
age^3	5.745e-05	2.11e-05	2.724	0.007	1.6e-05	9.89e-05

Omnibus:	577.477	Durbin-Watson:	1.025
Prob(Omnibus):	0.000	Jarque-Bera (JB):	39586.670
Skew:	5.336	Prob(JB):	0.00
Kurtosis:	44.997	Cond. No.	4.74e+06

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 4.74e+06. This might indicate that there are strong multicollinearity or other numerical problems.

Non Linear Regression Output Model number: 7
Independent variable: dis

```
----------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                   crim   R-squared:                       0.278
Model:                            OLS   Adj. R-squared:                  0.274
Method:                 Least Squares   F-statistic:                     64.37
Date:                Sun, 30 Jul 2023   Prob (F-statistic):           3.14e-35
Time:                        15:19:49   Log-Likelihood:                -1724.0
No. Observations:                 506   AIC:                             3456.
Df Residuals:                     502   BIC:                             3473.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         30.0476      2.446     12.285      0.000      25.242      34.853
dis          -15.5544      1.736     -8.960      0.000     -18.965     -12.144
dis^2          2.4521      0.346      7.078      0.000       1.771       3.133
dis^3         -0.1186      0.020     -5.814      0.000      -0.159      -0.079
==============================================================================
Omnibus:                      577.742   Durbin-Watson:                   1.129
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            42444.706
Skew:                           5.305   Prob(JB):                         0.00
Kurtosis:                      46.596   Cond. No.                     2.10e+03
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.1e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
----------------------------------------

Non Linear Regression Output Model number: 8
```

Independent variable: rad

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.400
Model:	OLS	Adj. R-squared:	0.396
Method:	Least Squares	F-statistic:	111.6
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	2.31e-55
Time:	15:19:49	Log-Likelihood:	-1677.1
No. Observations:	506	AIC:	3362.
Df Residuals:	502	BIC:	3379.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.6055	2.050	-0.295	0.768	-4.633	3.422
rad	0.5127	1.044	0.491	0.623	-1.538	2.563
rad^2	-0.0752	0.149	-0.506	0.613	-0.367	0.217
rad^3	0.0032	0.005	0.703	0.482	-0.006	0.012

Omnibus:	659.751	Durbin-Watson:	1.351
Prob(Omnibus):	0.000	Jarque-Bera (JB):	77838.247
Skew:	6.526	Prob(JB):	0.00
Kurtosis:	62.343	Cond. No.	5.43e+04

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 5.43e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Non Linear Regression Output Model number: 9
Independent variable: tax

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.369
Model:	OLS	Adj. R-squared:	0.365
Method:	Least Squares	F-statistic:	97.80
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	7.34e-50
Time:	15:19:49	Log-Likelihood:	-1689.9
No. Observations:	506	AIC:	3388.
Df Residuals:	502	BIC:	3405.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	19.1836	11.796	1.626	0.105	-3.991	42.358
tax	-0.1533	0.096	-1.602	0.110	-0.341	0.035
tax^2	0.0004	0.000	1.488	0.137	-0.000	0.001
tax^3	-2.204e-07	1.89e-07	-1.167	0.244	-5.91e-07	1.51e-07

Omnibus:	644.161	Durbin-Watson:	1.293
Prob(Omnibus):	0.000	Jarque-Bera (JB):	69773.212
Skew:	6.278	Prob(JB):	0.00
Kurtosis:	59.141	Cond. No.	6.16e+09

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 6.16e+09. This might indicate that there are strong multicollinearity or other numerical problems.

```
---------------------------------------


Non Linear Regression Output Model number: 10
Independent variable: ptratio
---------------------------------------
```

OLS Regression Results

Dep. Variable:	crim	R-squared:	0.114
Model:	OLS	Adj. R-squared:	0.108
Method:	Least Squares	F-statistic:	21.48
Date:	Sun, 30 Jul 2023	Prob (F-statistic):	4.17e-13
Time:	15:19:49	Log-Likelihood:	-1775.8
No. Observations:	506	AIC:	3560.
Df Residuals:	502	BIC:	3577.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	477.1840	156.795	3.043	0.002	169.129	785.239
ptratio	-82.3605	27.644	-2.979	0.003	-136.673	-28.048
ptratio^2	4.6353	1.608	2.882	0.004	1.475	7.795
ptratio^3	-0.0848	0.031	-2.743	0.006	-0.145	-0.024

Omnibus:	572.356	Durbin-Watson:	0.945
Prob(Omnibus):	0.000	Jarque-Bera (JB):	36070.763
Skew:	5.294	Prob(JB):	0.00
Kurtosis:	42.985	Cond. No.	3.02e+06

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.02e+06. This might indicate that there are strong multicollinearity or other numerical problems.

Non Linear Regression Output Model number: 11
Independent variable: lstat

```
----------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                    crim   R-squared:                       0.218
Model:                             OLS   Adj. R-squared:                  0.213
Method:                  Least Squares   F-statistic:                     46.63
Date:                 Sun, 30 Jul 2023   Prob (F-statistic):           1.35e-26
Time:                         15:19:49   Log-Likelihood:                -1744.2
No. Observations:                 506   AIC:                             3496.
Df Residuals:                     502   BIC:                             3513.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.2010      2.029      0.592      0.554      -2.785       5.187
lstat         -0.4491      0.465     -0.966      0.335      -1.362       0.464
lstat^2        0.0558      0.030      1.852      0.065      -0.003       0.115
lstat^3       -0.0009      0.001     -1.517      0.130      -0.002       0.000
==============================================================================
Omnibus:                      607.734   Durbin-Watson:                   1.239
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            53621.219
Skew:                           5.726   Prob(JB):                         0.00
Kurtosis:                      52.114   Cond. No.                     5.20e+04
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.2e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
----------------------------------------

Non Linear Regression Output Model number: 12
```

```
Independent variable: medv
---------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                   crim   R-squared:                       0.420
Model:                            OLS   Adj. R-squared:                  0.417
Method:                 Least Squares   F-statistic:                     121.3
Date:                Sun, 30 Jul 2023   Prob (F-statistic):           4.45e-59
Time:                        15:19:49   Log-Likelihood:                -1668.5
No. Observations:                 506   AIC:                             3345.
Df Residuals:                     502   BIC:                             3362.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         53.1655      3.356     15.840      0.000      46.571      59.760
medv          -5.0948      0.434    -11.744      0.000      -5.947      -4.242
medv^2         0.1555      0.017      9.046      0.000       0.122       0.189
medv^3        -0.0015      0.000     -7.312      0.000      -0.002      -0.001
==============================================================================
Omnibus:                      569.730   Durbin-Watson:                   1.359
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            47929.717
Skew:                           5.106   Prob(JB):                         0.00
Kurtosis:                      49.573   Cond. No.                     3.67e+05
==============================================================================
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.67e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
---------------------------------------
```

[131]:
```
# Creating a final results df with each model, coefficient and p value
# Create an empty list to store the results for each dependent variable
results_dfs_c3q15 = []

n = 1
for x in X_list_c3q15.columns:
    X_c3q15 = pd.DataFrame({
        x: X_list_c3q15[x],
        f'{x}^2': X_list_c3q15[x] ** 2,
        f'{x}^3': X_list_c3q15[x] ** 3
    })

    X_c3q15 = sm.add_constant(X_c3q15)

    non_linear_model_c3q15 = sm.OLS(y_c3q15, X_c3q15)
    non_linear_results_c3q15 = non_linear_model_c3q15.fit()

#     # Print the details for each iteration
#     print_green_bold(f'Non Linear Regression Output Model number: {n}')
#     print_green_bold(f'Independent variable: {x}')
#     print_green_bold('----------------------------------------')
#     print_green_bold(non_linear_results.summary())
#     print_green_bold('----------------------------------------\n\n\n\n')

    # Extract and store the coefficients and p-values for the original␣
 ↪predictor, squared, and cubed terms
    feature_names_c3q15 = X_c3q15.columns[1:]
    coefficients_c3q15 = non_linear_results_c3q15.params[1:]
    p_values_c3q15 = non_linear_results_c3q15.pvalues[1:]

    # Create a DataFrame for the current dependent variable and store the␣
 ↪results
    result_df_c3q15 = pd.DataFrame({
        'Dependent Variable': [f'Model number {n}'] * len(feature_names_c3q15),
        'Feature': feature_names_c3q15,
        'Coefficient': coefficients_c3q15,
        'P-Value': p_values_c3q15
    })

    # Append the DataFrame to the list of results for each dependent variable
    results_dfs_c3q15.append(result_df_c3q15)


    n=n+1

final_results_df_c3q15 = pd.concat(results_dfs_c3q15, ignore_index=True)
```

```
print_green_bold("\nFinal Results DataFrame:")
##Printing wherever P-value<0.05
print_green_bold(final_results_df_c3q15[final_results_df_c3q15['P-Value']<0.05])
```

```
Final Results DataFrame:
   Dependent Variable    Feature  Coefficient       P-Value
0       Model number 1         zn    -0.332188  2.612296e-03
3       Model number 2      indus    -1.965213  5.297064e-05
4       Model number 2    indus^2     0.251937  3.420187e-10
5       Model number 2    indus^3    -0.006976  1.196405e-12
9       Model number 4        nox -1279.371252  2.758372e-13
10      Model number 4      nox^2  2248.544053  6.811300e-15
11      Model number 4      nox^3 -1245.702874  6.961110e-16
16      Model number 6      age^2    -0.007230  4.737733e-02
17      Model number 6      age^3     0.000057  6.679915e-03
18      Model number 7        dis   -15.554353  6.374792e-18
19      Model number 7      dis^2     2.452072  4.941214e-12
20      Model number 7      dis^3    -0.118599  1.088832e-08
27     Model number 10    ptratio   -82.360538  3.028663e-03
28     Model number 10  ptratio^2     4.635347  4.119552e-03
29     Model number 10  ptratio^3    -0.084760  6.300514e-03
33     Model number 12       medv    -5.094831  2.637707e-28
34     Model number 12     medv^2     0.155496  3.260523e-18
35     Model number 12     medv^3    -0.001490  1.046510e-12
```

We find that the following predictors are statistically significant at 5% level of significance : - zn - indus - indus^2 - indus^3 - nox - nox^2 - nox^3 - age^2 - age^3 - dis - dis^2 - dis^3 - ptratio - ptratio^2 - ptratio^3 - med - medv^2 - medv^3 Thus for indus, nox, age, dis, ptratio and medv, there is evidence of a non-linear relationship with crim as the squared/cubed predictors are statistically significant.

[ ]: