

**numpy** - библиотека за „scientific computing”, слична је машинаму

`import numpy as np`

**matplotlib** - библиотека за употребите

`import matplotlib.pyplot as plt`

**pandas** - библиотека за манипулацију података

`import pandas as pd`

Главни објект је **Dataframe** који симболише **ред** на стабеларним подацима, може се делијати и као матрица редова и колоне и имају суштински додатни функционалности

- **дипломатичка (функционална) зависност** - општији број зависних променљивих  
дом. X (предмет) - Тачко је описан однос између ПРОМЕНЉИВЕ

- **линейчка зависност** - и да разматрамо, веза изм. где променљиве имају линеарну везу која ће нумеричким начином описати гајије  $\hat{y} = \alpha x + b$

$x_i$  - бројтака независне променљиве за свако измерење i

$y_i$  - бројтака зависне променљиве за свако измерење i

$\epsilon_i$  - предвиђене бројтаке зависне променљиве за свако измерење i

У њој се описује у неким измерењима, а чим је ово увршијено сматрају  
линейчку са десета променљивија  $e_i = y_i - \hat{y}_i$

- један од најчешћих метода **OLS** - минимизира суму квадратних  
погрешака **SSE** (збори који се добијају описуји везе)

- **линейност** - представљава да веза између X и Y је линеарна. то смисло да  
сага шематично ВИЗУЕЛНО општо график, а други начин је испитивање коришћење  
линейних тестова (тестови Т-статистике)

- **T-статистика** је тест који се користи да провери однос између X и Y - често представљава да  
НЕ ПОСТОЈИ да  $\alpha=0$  и  $y=\alpha x + b$

• T-брејтова туђи честа (правило 25% тип) - објаснији хипотезу и постепено да је  
однос између X и Y **линейаран**

• T-брејтова честа (око тип) - неможно да објасни хипотезу  
ако је бр честа и знамо да је веза **линейарна** или **нелинейарна**.

- уместо T-брејтова, линеарност можемо проверити и користећи **P-брејтова** - и да је  
вероватноћа да немоје т-дистрибуције извучи конкретну T-брејтова или неку другу  
наме вероватној

- ако је P-брејтова **мала**, T-брејтова туђи честа и веза је **линейарна**

P-брејтова је мала  $< 0,05$   $\rightarrow$  95%. што сигурно да је веза линеарна

- веће не можемо утврдити да ли је веза линеарна, и то проверили линеарност бит кориснији т-статистика

- Укин бр. са којим ради по скривен изглед на 0280 кутији у Фабрици. Користи се линеарни модел  $y = \alpha + \beta x$  како јесто процењених вр. кутија (тешко је објаснити оцењивање свих кутија, бет сачувати о агр. др. кутија)
  - држи објашње да смо имали - доб. јесто модел са пр. вр. за најбољи и прве
  - зато ће се променити парни модел тада јесто проширили скривен изглед који је увр. модел добар - да јесто да смо објавили - јер ће најбољи изглед који је увр. објашњен да је увр. вр. кутија те се најбољи модел вр. кутија је параметар (узвеси бр. дајући најбољи јесто имали све објашње)
  - интервал објашње креира се ја неки проречани изглед који је увр. јесто ког р-вредност је узимаши да је 95%. - јако и обје јесу једнаке да са 95% је јасно да међу којима вр. са најбољим јесто
- model.conf.int() @ model.summary()
- да сад смо интервал објашње параметарима да је параметар најбољи и објечан, али можемо да је дефинисан и за САМЕ ПРЕДИКЦИЈЕ
  - јаснују којима бр. се најбољи  $y_i$  за дају  $x_i$ ? (који је речик чини је увр. са што)
- model.get\_prediction(ex-with-const).summary\_frame("роман")

L.1.N.E - Честотабавне (морал) да баше јесто јако јесто било смештити да је редовити модел објашњен за дају проблем)

- 1) Линеарност - вако када је однос између  $X$  и  $Y$  линеаран
 

може да графику претидат да су је дужу праве  $x = y$  тада ће проречијући било јасно да је линеарна ојачавања - модел је јако ајекланат  
МАДА ојачавања - модел је јако ајекланат
- 2) независноста грешака - вако када су грешаки јаснији независно
 

постоји се да ће резидуалима не да ће осећати да је резидуали - проречијући грешака из узорка изгледају. Разлик између праве и проречије вр.  $y = \hat{y}$

постоји јасно што ИПР ТАКО ГРАФИК РЕЗИДУАЛА

Резидуал има јаснији ШАБЛОН - проречији најчешће има јаснији ШАБЛОН - проречији таји најчешће

Двогодишњи Записи стапајући чин

$1,5 \leq d \leq 2$  тада ауторегресије  $\rightarrow$  ПРЕД. ВАИН  
 $d < 1,5$  или  $d > 2$  нити  $\rightarrow$  не ВАИН

- 3) нормалноста грешака

- средња вр. 0 за грешаке
- грешаке су нормално расподељене Ауторегресији - Двогодишњи стапајући чини - бројнији  
 • вр.  $d > 2$   $\rightarrow$  расподељене НОРМАЛНА  
 • вр.  $d < 1,5$   $\rightarrow$  расподељене НИФ. НОРМАЛНА

- оп бр 0 али тијесу нормално дистрибуирати  $\rightarrow$  ПРЕД. ИС ВАРІЈ
- али ако је узорак подједнако велики, застварајући тијес нормално дистрибуирати  
 $\rightarrow$  дрети и. баланси

4) Једнака варјација - варјатија која јављајуше промене око референтне вредбе између једнаких варјација (а чија вредност са фиксни резултати) - која неће имати варјатију

Sjordeng - Кватитет искл - вредна је бројност

• оп бр  $\geq$  дрети - ВАРИЈАЦИЈА ЈЕ ЈЕДНАЧИНА

• оп бр  $<$  дрети - ВАРИЈАЦИЈА НЕЋЕ ЈЕДНАЧИНА

## ШИШЕСТРУКА

+ још једна прашт.

5) не јаснији савремени количарстви између 2 или више независних променљивих савремених количарстава који су променљиве  $x_1$  и  $x_2$  обезбједије количарством формулом:

$$x_2 = ax_1 + b$$

количарстви уочавамо преко МАТРИЦЕ КОРЕЛАЦИЈЕ

-  $S_{xy} =$  има већи приказателј  $r$

- ИНТЕРПОЛАЦИЈА - ПРЕМЕШАЊА ПОДАТАКА