VARIABILITY PARTITIONING

$y = mx + c$ correlation$(x, y) = R$

R^2 = strength of relationship between X and Y

$R^2 \longrightarrow$ how much variability in Y can be explained by X

$y = m_1 x_1 + m_2 x_2 + \ldots c \; [\underline{m_0^c} \, \underline{x_0'}]$

Partitioning variability of Y using $x_1, x_2, x_3 \ldots$ [ANOVA]

sample mean

Y = {real observed values}

Y' = predictions of your linear model

$\sum (Y - \bar{Y})^2$ = total variability in Y

$\sum (Y - Y')^2$ = unexplained variability in Y (error)

explained variability = $\underbrace{\sum (Y - \bar{Y})^2}_{\text{variability}} - \underbrace{\sum (Y - Y')^2}_{\hookrightarrow \text{errors}}$

$$F\text{-statistic} = \frac{\text{explained variability}}{\text{unexplained variability}}$$

ML model $\longrightarrow$ statistically significant prediction

ADJUSTED R^2

$$y = m_1 x_1 + m_2 x_2 + \cdots \quad c$$

$$R^2_{adj} = 1 - \left(\frac{\text{unexplained variability}}{\text{total variability}} \times \frac{n-1}{n-k-1} \right)$$

compensate strength of explainability between y and $\{x_1, x_2 \ldots\}$

n = number of observations
k = predictors count

every "variable" introduces error
every degree of freedom $(n-k-1)$ introduces error

model A $\longrightarrow R^2_{adj} = A'$
model B $\longrightarrow R^2_{adj} = B'$

$B' > A'$, choose model B over A

T-TEST: BRIEF

inferential statistic

$\bar{X}_1 \longleftrightarrow \bar{X}_2$

$y = m_1 x_1 + m_2 x_2 \ldots + c$

$H_0 = m_1, m_2, m_3 \ldots = 0$

H_A = atleast $m_i \neq 0$

$$t = \frac{m_{i=2} - 0}{SE = \frac{\sigma}{\sqrt{n}}}$$

p value for second variable $\longleftarrow$ pt $\longleftarrow$ degrees of model and $t_{i=2}$

scipy.stats