

# 1 Počítanie $R^2$

Vytvorme funkciu `r_squared(x, y, beta)` - kde `x` je matica vektorov nezávislých premenných, `y` je vektor závislej premennej, `beta` je vektor optimálnych  $\beta$  koeficientov získaných lineárnou regresiou - ktorá bude počítat  $R^2$  koeficient podľa definície:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

```
def r_squared(x: np.ndarray, y: np.ndarray, beta: np.ndarray) -> float:
    # calculate y-hat and mean of y vector
    y_hat = beta[0] + x @ beta[1:]
    y_mean = np.mean(y)

    res1 = 0      # partial result for the numerator in the formula
    res2 = 0      # partial result for the denominator in the formula

    # calculate the sums
    res1 = np.sum((y - y_hat)**2)
    res2 = np.sum((y - y_mean)**2)

    # calculate the R^2 coefficient
    result = 1 - (res1 / res2)
    return result
```

Implementujeme metódu na dátach `A04wine.csv`. Načítame dáta pomocou, rozdelíme ich do premenných, vyriešime potrebné LP problémy pre načítané dáta (rovnako ako v predošlých úlohách) a vypočítame  $R^2$  koeficient:

```
betas = solve.x[:k+1]
betas_inf = solve_inf.x[:k+1]

r_squared(x, y, betas)
r_squared(x, y, betas_inf)
```

Vypočítané príslušné R-kvadráty teda sú:

$$R_{(1)}^2 \approx 0.78813$$

$$R_{(\infty)}^2 \approx 0.80649$$

Z toho môžeme usúdiť, že náš model sa dá považovať za relatívne vhodný pre tieto dáta. Tiež vidíme, že lineárna regresia pomocou Chebyshevovej normy lepšie zachytáva rozptyl dát.