

Korelace

1. Analyzujte korelaci výšky, hmotnosti a měsíce narození skupiny lidí.

COVAR (A1:A50, B1:B50) – odhad kovariance

$$\hat{c}ov(x, y) = \langle xy \rangle - \langle x \rangle \langle y \rangle$$

PEARSON (A1:A50, B1:B50) – odhad korelace

$$\hat{\rho}(x, y) = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\hat{\sigma}_x \hat{\sigma}_y}$$

$$\hat{\rho}(x, y) = \frac{N-1}{N} \hat{\rho}(x, y)_{\text{Excel}}$$

$$\hat{\sigma}_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2}$$

$$\hat{\sigma}_y = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \langle y \rangle)^2}$$

Centrální limitní věta

2. Ukažte, že součet N náhodných proměnných s rovnoměrným rozdělením $U(0, 1)$ konverguje k normálnímu rozdělení $N\left(\frac{N}{2}, \sqrt{\frac{N}{12}}\right)$

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

N=1                # počet sumací
Nsim=100000        # počet vygenerovaných náhodných čísel

x=np.empty(N)
y=np.empty(Nsim)

for i in range(Nsim):
    x=np.random.random_sample(N)
    y[i]=np.sum(x)

# Co očekáváme z CLV za parametry gaussovky?
mu=N/2
sigma=np.sqrt(N/12)

# Vykreslíme gaussovku a histogram N-krát sečtených náhodných čísel
xp=np.arange(0,N,0.01)
yp=norm.pdf(xp,mu,sigma)
plt.pause(0.1)
plt.hist(y,bins=100,density='True')
plt.plot(xp,yp,c='red')
```