

## Statistička metode za informatičare

### Zadatak:

<https://www.kaggle.com/kumarajarshi/life-expectancy-who>

- Opišite varijable statističkog skupa i grafički ih prikažite. Izračunajte korelacije između kvantitativnih varijabli i grafički ih prikažite.
- Temeljem kvantitativnih varijabli *Number of years schooling* i *Average Body mass index of entire population (BMI)* definirajte nove varijable na sljedeći način

*Number of years schooling* kvalitativna =

$$\begin{cases} 0 & 0 \leq \text{Number of years schooling} \leq 8 \\ 1 & 8 < \text{Number of years schooling} \leq 12 \\ 2 & 12 < \text{Number of years schooling} \end{cases}$$

$$\text{BMI kvalitativna} = \begin{cases} \text{vrlo jaka pothranjenost} & 0 \leq BMI \leq 15 \\ \text{jaka pothranjenost} & 15 < BMI \leq 16 \\ \text{umjeren pothranjenost} & 16 < BMI \leq 18,5 \\ \text{normalna težina} & 18,5 < BMI \leq 25 \\ \text{umjerena pretilost} & 25 < BMI \leq 30 \\ \text{jaka pretilost} & 30 < BMI \leq 40 \\ \text{vrlo jaka pretilost} & BMI > 40 \end{cases}$$

- Opišite novodobivene varijable statističkog skupa i grafički ih prikažite.
- Korištenjem hi-kvadrat testa ispitajte postoji li povezanost kvalitativnih statističkih varijabli *Number of years schooling* kvalitativna i *BMI* kvalitativna.
- Korištenjem hi-kvadrat testa ispitajte postoji li povezanost kvalitativnih statističkih varijabli *Status* (Developed/Developing) i *BMI* kvalitativna.
- Korištenjem odgovarajućeg parametarskog i neparametarskog testa ispitajte postoje li razlike varijable *Life expectancy in years* po razinama (modalitetima) varijable *Number of years schooling* kvalitativna. Za provedbu parametarskog testa ispitajte pretpostavke za provedbu. Ukoliko se korištenjem parametarskog testa utvrdi da postoje signifikantne razlike provedite i post hoc test.
- Korištenjem odgovarajućeg parametarskog i neparametarskog testa ispitajte postoje li razlike varijable *Life expectancy in years* po razinama (modalitetima) varijable *BMI* kvalitativna. Za provedbu parametarskog testa ispitajte pretpostavke za provedbu. Ukoliko se korištenjem parametarskog testa utvrdi da postoje signifikantne razlike provedite i post hoc test.
- Korištenjem dvofaktorske analize varijace ispitajte postoje li signifikantne razlike varijable *Life expectancy in years* po tretmanima varijabli *Status* i *Number of years schooling* kvalitativna.
- Korištenjem odgovarajućeg parametarskog i neparametarskog testa ispitajte postoje li razlike varijable *Life expectancy in years* po razinama (modalitetima) varijable *Status*. Za provedbu parametarskog testa ispitajte pretpostavke za provedbu.
- Provedite regresijsku analizu kod koje će zavisna varijabla biti *Life expectancy in years*, a nezavisne sve ostale kvantitativne varijable. Komentirajte parametre regresije: koeficijent determinacije i korigirani koeficijent determinacije. Interpretirajte skupni i

pojedinačne testove signifikantnosti regresije za svaku od promatranih nezavisnih varijabli. Komentirajte rezultate regresije. Provedite izbor varijabli koristeći neku od metoda za izbor varijabli (Forward Selection Procedure, Backward Selection Procedure, Backward/Forward). Nacrtajte normalni prikaz rezidualnih vrijednosti.