

---

## ZARZĄDZANIE DANYMI I ELEMENTY STATYSTYKI OPISOWEJ

---

**Biblioteki:** dplyr, ggplot2, moments, stats

**ZAD. 1.** Otworzyć plik z danymi z ankiety w Excelu

- a) w kolumnie Szkoła średnia zamienić łańcuchy *klasa z rozszerzoną matematyką* na *RM* oraz *klasa z podstawową matematyką* na *PM*, *Technikum Informatyczne* na *TI*,
- b) zmienić nagłówki kolumn na krótsze wprowadzając nazwy: Waga, Wzrost, M.zamieszk, Sz.średnia, ECTS, Algebra, MSzS1, Narz.inż, Prog1, WdI, L.godz, L.sys.op, System, Wiek, F.stud,
- c) plik po zmianach zapisać pod nazwą Ankieta.xlsx.

**ZAD. 2.** Wczytać w RStudio plik Ankieta.xls ustalając numeryczny typ zmiennych: Waga, Wzrost, ECTS, Algebra, MSzS1, Narz.inż, Prog1, WdI, L.sys.op, Wiek (File/Import Data-set/From Excel).

- a) zmienić typ znakowy (*char*) na typ czynnikowy (*factor*) zmiennych: Płeć, M.zamieszk, Sz.średnia, System.
- b) sprawdzić działanie funkcji *summary* dla zmiennych: Waga, Wzrost, M.zamieszk, System,
- c) w ramce Ankieta utworzyć nową zmienną *Śr.kursy*, zawierającą średnią ocen z kursów,
- d) przenieść kolumny z ocenami z kursów do podzbioru Ankieta.kursy (użyć *subset*),
- e) usunąć kolumny Algebra, MSzS1, Narz.inż, Prog1, WdI z ramki danych Ankieta,
- f) utworzyć funkcję **zakres3sigm**, która zwróci dla dowolnej zmiennej ramkę danych z nagłówkami: *lewy.kres*, *prawy.kres* oznaczającymi średnią minus/plus trzy odchylenia standardowe (użyć *function*, *data.frame*, *mean* - wartość średnia, *sd* – odchylenie standardowe),
- g) dla zmiennej *Śr.kursy* wyznaczyć ewentualne dane odstające i zastąpić je symbolem braku danych, a później średnią zaokrągloną do części dziesiętnych,
- h) wyodrębnić podzbiory mężczyzn i kobiet o nazwach Ankieta.M oraz Ankieta.K odpowiednio (wykorzystać funkcję *filter*, biblioteka *dplyr*)
- i) wyznaczyć ewentualne wartości odstające dla *Wagi* i *Wzrostu* w grupie kobiet oraz w grupie mężczyzn symbolem braku danych a następnie zastąpić je wartościami średnimi w tych grupach, wyznaczonymi z dokładnością do części dziesiętych,
- j) utworzyć nową zmienną *L.g.kody*, w której zostaną umieszczone 3 przedziały liczbowe rozdzielone wartościami: 0, 4, 8, 24, odpowiadające kategoriom średniego czasu spędzanego przed komputerem w ciągu doby „krótko”, „średnio”, „długo” (użyć *cut*) oraz zliczyć te kategorie.

**ZAD. 3.** Sporządzić histogramy dla zmiennych jakościowych *M.zamieszk*, *Sz.średnia*, *System*.