Wrocław, 24.10.2022 r.

**PODSTAWY ANALIZY DANYCH – PRAKTYCZNE WARSZTATY**

**ZADANIA DO SAMODZIELNEGO PRZEĆWICZENIA, LISTA NR 2**

1. Korzystając np. z dokumentacji R proszę obejrzeć funkcję *round()*, pozwalającą na zaokrąglanie liczb w różny sposób i przećwiczyć jej działanie na dowolnych przykładach. W praktyce często używa się wariantów *floor()* i *ceiling()*, warto je poznać.
2. Proszę przejrzeć podrozdział 2.4 z książki p. Przemysława Biecka <http://www.biecek.pl/R> o podstawowych sposobach wizualizacji danych.
3. Aby wygenerować *x* (zob. dalej😊) liczb losowych z rozkładu normalnego o zadanych; wartości oczekiwanej *mean* i odchyleniu standardowym *sd* możemy użyć funkcji *rnorm*(*x, mean, sd*)*.* Jeżeli pominiemy *mean* i *sd*, liczby zostaną wygenerowane ze standardowego rozkładu normalnego, dla którego *mean* = 0 i *sd* = 1. (Proszę porównać wynik działania funkcji *rnorm*(x) i *rnorm*(x, 0, 1). Powtarzalność wyników zapewnimy przez użycie przed komendą *rnorm*(x) funkcji *set.seed*(wartość\_ziarna), gdzie wartość\_ziarna to liczba całkowita. Proszę pobawić się generowaniem wektorów liczb losowych o różnej długości, łączeniem ich w jeden wektor, tworzeniem sum wektorów itp. Zbiory danych uzyskane w ten sposób mogą mieć rozkłady wielomodalne. Proszę postarać się zwizualizować strukturę danych przy pomocy histogramów tak, aby jak najlepiej pokazać strukturę zbiorowości, także wielomodalność. Powinni Państwo zauważyć, ze większa liczba danych sprzyja zauważeniu, że dane były generowane z rozkładów normalnych.
4. Proszę wygenerować 500 liczb z rozkładu normalnego o dowolnie zaproponowanych parametrach, a następnie wygenerować domyślny histogram o nierównych rozpiętościach przedziałów klasowych. Rozpiętości przedziałów klasowych proszę zaproponować dowolnie. Co pokazuje taki histogram? Dla porównania, dla tych samych danych proszę wygenerować domyślny histogram o równych przedziałach klasowych, a następnie zmodyfikować go przez wprowadzenie parametru *probability = T*.
5. Proszę zaokrąglić dane wykorzystywane w ćwiczeniu nr 3 do wartości całkowitych, a następnie pokazać dane na wykresach słupkowych. Podczas zabawy powinni Państwo zauważyć, że wykresy lepiej pokazują strukturę zbiorowości, jeżeli wariantów cechy jest stosunkowo niewiele.
6. Proszę poszukać w dokumentacji R-a, jak umieścić na wykresie legendę. Proszę się pobawić kolorami, długościami osi, podpisami osi, legendą itp. wykresów sporządzonych w ćwiczeniach nr 3 i 4.
7. Proszę sprawdzić, czy w zbiorowościach wygenerowanych w poprzednich ćwiczeniach są obserwacje odstające. Można to zrobić graficznie, korzystając np. z wykresu ramka-wąsy lub sprawdzając, czy w zbiorowości występują wartości X spełniające warunek:

X > Q3+1.5\*IQR lub X < Q1-1.5\*IQR. Podczas oceny Q1 i Q3 proszę uważać, bo Państwa dane mogą być dyskretne lub ciągłe😊.

1. W naszym zespole na platformie MS Teams znajdą Państwo trzy pliki Excela z danymi: „Proszki\_do\_prania”, „Produkcja\_koszt” i „Przestoje”. Dane pochodzą z podręcznika S. Ostasiewicz, Z. Rusnak i U. Siedleckiej „Statystyka. Elementy teorii i zadania”, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 1989, str. 330, odpowiednio zad. nr 12.34, 12.35 i 12.37 (opis danych został dostosowany do niniejszych zajęć).

„Proszki\_do\_prania” zawierają dane na temat zużycia proszków do prania w pewnym kraju, w różnych momentach czasowych. Zatem, dane przedstawiają szereg czasowy.

Dane z pliku „Produkcja\_koszt” pokazują, jak zmieniała się produkcja i koszty jednostkowe produkcji w pewnym zakładzie. Dane z pliku „Przestoje” pokazują wyniki działań podejmowanych przez firmę w celu zmniejszenia czasu trwania przestojów w pracy pewnej firmy.

Proszę obejrzeć dane na wykresach rozrzutu (scatterplots). Wykresy można sporządzić korzystając z funkcji plot(x,y) z pakietu „base” lub wykorzystując funkcję scatterplot(x,y) z pakietu „car”. Korzystając z pakietu car warto wyrzucić prostą regresji, która się automatycznie pokaże (parametr regLine=F) i efekt działania parametru smooth (wpisujemy smooth=F).