Materiały do zajęć 4 z Programowania narzędzi analitycznych

1. Działania macierzowe

`t(A)` - zwraca macierz transponowaną do macierzy `A`
`solve(A)` - zwraca odwrotną macierz do macierzy `A`
`solve(A,B)` - rozwiązuje układ równań `A%*%x=B`

2. Wczytywanie danych

`read.table` - wczytanie danych z pliku tekstowego, dane w pliku mają postać tabeli
`read.csv` - wczytywanie danych z plików `csv`, domyślnie dane rozdzielane są przecinkami
`read.csv2` - wczytywanie danych z plików `csv`, domyślnie dane rozdzielane są średnikami
`read.delim` - wczytywanie plików z wartościami rozdzielanymi (np. tabulatorem `''\ t''`
`read.xls` - wczytywanie plików `.xls`, procedura z biblioteki `xlsReadWrite`
`read.xlsx` - wczytywanie plików `.xlsx`, z biblioteki `xlsx`

3. Typy zmiennych

`list()` - tworzy listę. Podobnie jak wektor, lista to również uporządkowany zbiór elementów. W przeciwieństwie do wektora, elementy listy mogą mieć różne typy. Podobnie jak w przypadku wektora poszczególne elementy mogą mieć nazwy.
`list(imie=c("Jan","Tomasz"), nazwisko="Kowalski", wiek=25, czyWZwiazku=TRUE)`
`data.frame()` - tworzy ramkę danych. Szczególnym typem jest ramka danych, nazywana również tabelą danych. Ramka danych jest zazwyczaj kojarzona z macierzową/tabelaryczną strukturą, której elementy w każdej kolumnie są tego samego typu, ale mogą różnić się typami pomiędzy kolumnami. Z tego powodu ramkę danych można traktować jak listę wektorów o tej samej długości, każdy wektor odpowiada jednej kolumnie.
`factor()` tworzy zmienną typu czynnikowego. Typ czynnikowy nazywany jest też typem wyliczeniowym lub kategorycznym.
`factor(x=character(),levels=sort(unique.default(x),na.last=TRUE),`
`labels=levels, exclude=NA, ordered=is.ordered(x))`

Przemysław Biecek, Przewodnik po pakiecie R, 2014, (link)

4. Testy zgodności dla rozkładów ciągłych

Na podstawie *Przemysław Biecek, Wybrane testy normalnosci, 2013. Materiały Fundacji Smarter-Poland.pl.* (link)

`cvm.test(x)` - test Cramera von Misesa na normalność rozkładu dostępny w bibliotece `nortest`
`ad.test(x)` - test Andersona-Darlinga na normalność rozkładu dostępny w bibliotece `nortest`
`ks.test(scale(x), ''pnorm'')` - test Kołmogorowa-Smirnowa na normalność rozkładu
`ks.test(scale(x), ''p...'')` - test Kołmogorowa-Smirnowa na zgodność z określonym rozkładem
`ks.test(x,y)` - test Kołmogorowa-Smirnowa na zgodność dwóch rozkładów
`lillie.test(x)` - test Lillieforsa na normalność rozkładu dostępny w bibliotece `nortest`
`lillieTest(x)` - test Lillieforsa na normalność rozkładu dostępny w bibliotece `fBasics`
`dagoTest(x)` - test D'Agostino-Pearsona na normalność rozkładu oparty o skośność i kurtozę dostępny w bibliotece `fBasics`
`jarque.bera.test(x)` - test Jarque-Bera na normalność rozkładu oparty o skośność i kurtozę dostępny w bibliotece `tseries`
`shapiro.test(x)` - test Shapiro-Wilka na normalność rozkładu
`sf.test(x)` - test Shapiro-Francia na normalność rozkładu dostępny w bibliotece `nortest`

4. Testy zgodności dla rozkładów dyskretnych
Na podstawie *Przemysław Biecek, Wybrane testy normalnosci, 2013. Materiały Fundacji Smarter-Poland.pl.* (link)

`chisq.test(x=observed, p=expected)` - test zgodności χ^2 Pearsona

`g.test(x=observed, p=expected)` - test zgodności Kullbacka-Leiblera
https://en.wikipedia.org/wiki/G-test (link)

5. Tabela liczności
`table(x)` - tablica częstości dla zmiennej `x`

6. Polecenie `lm`
`lm()` - polecenie do szacowania modeli regresji liniowej

Materiały do zajęć 3 z Programowania narzędzi analitycznych

1. Wykresy

Wybrane funkcje do tworzenia wykresów:
`plot` - tworzy wykresy: liniowe lub rozproszenia
`hist` - histogram
`barplot` - wykres słupkowy
`boxplot` - wykres pudełkowy
`qqnorm` - wykres kwantylowy
`curve(x^2, from=0, to=2)` - tworzy wykres krzywej określonej wzorem, argumentem jest x

2. Dodatkowe elementy wykresów

`abline(3, 5)` - dodaje do wykresu linię określoną wzorem $y = 3 + 5x$
`abline(v = 2)` - dodaje do wykresu linię pionową, $x = 2$
`abline(h = 0)` - dodaje do wykresu linię poziomą, $y = 0$
`lines` - dodaje do wykresu dodatkową linię/wykres
`points` - dodaje punkty
`text` - dodaje tekst
`title` - dodaje tytuł

3. Przykłady

```
plot(table(rpois(100, 5)), type = "h", col = "red", lwd = 10,
     main = "rpois(100, lambda = 5)")

plot(-1:1, -1:1, type = "n", xlab = "Re", ylab = "Im")
K <- 16; text(exp(1i * 2 * pi * (1:K) / K), col = 2)

plot(States03$Unemp, States03$Poverty, xlab = "Unemployment", ylab = "Poverty")
plot(Poverty ~ Unemp, data = States03, xlab = "Unemployment", ylab = "Poverty")
hist(States03$Poverty, main = "Poverty", xlab = "percent",
xlim = c(0, 24), ylim = c(0, 20))
plot(1:19, 1:19, pch = 1:19, xlab = "x", ylab = "y")
pie(rep(1, 8), col = 1:8)
curve(x^2, from = 0, to = 2)
curve(cos(x), from = 0, to = pi)
curve(cos(x), from = 0, to = pi, lty = 4, col = "red")
```

3. Opcje wykresów

`pch` - styl znaczników
`lty` - typ linii, Line types can either be specified as an integer (0=blank, 1=solid (default), 2=dashed, 3=dotted, 4=dotdash, 5=longdash, 6=twodash) or as one of the character strings ”blank”, ”solid”, ”dashed”, ”dotted”, ”dotdash”, ”longdash”, or ”twodash”, where ”blank” uses 'invisible lines' (i.e., does not draw them).
`lwd` - grubość linii
`col` - kolor, `(col=''red'', ''blue'',...)`
`xlim` - zakres osi x: `xlim=c(min, max)`
`ylim` - zakres osi y
`xlab` - etykieta osi x: `xlab=''moja etykieta''`
`ylab` - etykieta osi y
`main` - główny tytuł
`sub` - podtytuł

Na podstawie Mathematical Statistics with Resampling and R (link).

Materiały do zajęć 2 z Programowania narzędzi analitycznych

1. Statystyka opisowa

`sum(wektor)` - suma elementów wektora lub macierzy
`mean(wektor)` - wylicza średnią z wektora
`median(wektor)` - wyznacza medianę z wektora
`sd(wektor)` - wylicza odchylenie standardowe wartości wektora
`var(wektor)` - wylicza wariancję wartości wektora
`abs(liczba)` - wylicza moduł/wartość absolutną liczby
`quantile(x, probs=0.25)` - wylicza pierwszy kwartyl

2. Moda/dominanta

```
Mode <- function(x, na.rm = FALSE) {
  if(na.rm){
    x = x[!is.na(x)]
  }

  ux <- unique(x)
  return(ux[which.max(tabulate(match(x, ux)))])
}
```

Źródło: link: https://stackoverflow.com/

3. Kwantyle

`qnorm(x, mean=0, sd=1)` - kwantyle rozkładu standardowego normalnego
`pnorm(x, mean=0, sd=1)` - wartość dystrybuanty rozkładu w punkcie `x`, tj. $\Phi(x)$

Rozkład	Kwantyl	Gęstość	Dystrybuanta	Liczby losowe
Normalny	**q**norm(p, mean, sd)	**d**norm(x,mean, sd)	**p**norm(q, mean, sd)	**r**norm(n, mean, sd)
Beta	**q**beta(p, s1, s2)	**d**beta(x, s1, s2)	**p**beta(q, s1, s2)	**r**beta(n, s1, s2)
χ^2_n	**q**chisq(p, df)	**d**chisq(x, df)	**p**chisq(q, df)	**r**chisq(n, df)
Wykładniczy	**q**exp(p, rate)	**d**exp(x, rate)	**p**exp(q, rate)	**r**exp(n, rate)
t-Studenta	**q**t(p, df)	**d**t(x, df)	**p**t(q, df)	**r**t(n, df)
Jednostajny(0,1)	**q**unif(p, min, max)	**d**unif(x, min, max)	**p**unif(q, min, max)	**r**unif(n, min, max)
Gamma	**q**gamma(p, s, r)	**d**gamma(x, s, r)	**p**gamma(q, s, r)	**r**gamma(n, s, r)
F-Snedecora	**q**f(p, df1, df2)	**d**f(x, df1, df2)	**p**f(q, df1, df2)	**r**f(n, df1, df2)
Dwumianowy	**q**binom(p, s, p)	**d**binom(x, s, p)	**p**binom(q, s, p)	**r**binom(n, s, p)
Poissona	**q**pois(p, lambda)	**d**pois(x, lambda)	**p**pois(q, lambda)	**r**pois(n, lambda)

Materiały do zajęć 1 z Programowania narzędzi analitycznych

1. Pomoc

`help(''nazwa_polecenia'')` - wyświetla informacje o poleceniu
`?nazwa_polecenia` - wyświetla informacje o poleceniu
`help.search(''slowoKluczowe'')` - Przemysław Biecek, Przewodnik po pakiecie R, 2008, str. 18, (link): Przegląda opisy funkcji znajdujących się w zainstalowanych pakietach i wyświetla te pozycje, w których znaleziono wskazane slowoKluczowe. W tym przypadku slowoKluczowe może oznaczać również kilka słów lub zwrot. W liście wyników znajduje się również informacja, w którym pakiecie znajdują się znalezione funkcje.
`demo` - pliki instruktażowe. Przemysław Biecek, Przewodnik po pakiecie R, 2008, str. 10, (link): Dla wielu pakietów oraz funkcji dostępnych w R zostały przygotowane prezentacje, pokazujące możliwości danego pakietu lub funkcji. Takie prezentacje uruchamia się funkcją demo(utils).

2. Polecenia ogólne

`rm(a)` - usuwa zmienną `a`
`rm(list=ls())` - usuwa z pamięci komputera wszystkie zmienne i funkcje
`ctrl+l` - czyści okno poleceń
`getwd` - podaje bieżący katalog roboczy
`setwd` - zmiana katalogu domyślnego
`Sys.Time()` - wyświetla datę i godzinę
`Sys.Date()` - wyświetla datę
`q()` - wyłączenie programu R
`#` - tworzenie komentarza

3. Działania matematyczne

= lub `<-` przypisanie wartości np.: `a=4`
`+`,`-`,`*`,`/` - podstawowe działania matematyczne
^ - podniesienie do potęgi

4. Operatory i funkcje logiczne

== równy, porównanie
`!=` nierówny
`!` nierówny, negacja
`<`,`<=`,`>`,`>=` relacje
`&` - koniunkcja, "i", "AND"
`|` - alternatywa, "lub", "OR"
`xor` - alternatywa wykluczająca
`is.infinite(x)` - czy liczba `x` jest równa nieskończoności
`is.finite(x)` - czy liczba `x` jest skończona
`is.nan(x)` - czy `x` jest brakiem danych

5. Podstawowe funkcje

`sqrt(x)` - pierwiastek z liczby x ($\sqrt{x}$)
`exp(x)` - eksponenta z x (e^x)
`log(x)` - logarytm naturalny ($\ln(x)$)

`log10(x)` - logarytm o podstawie 10 z liczby `x`
`log2(x)` - logarytm o podstawie 2 z liczby `x`
`log(x, base=b)` - logarytm o podstawie `b` z liczby `x`
`sin()`,`cos()`,`tan()` - funkcje trygonometryczne
`asin()`,`acos()`,`atan()` - odwrotne funkcje trygonometryczne
`sinh()`,`cosh()`,`tanh()` - funkcje hiperboliczne
`abs()` - wartość bezwzględna
`factorial(n)` - silnia z liczby `n` ($n!$)
`floor()` - zaokrąglenie liczby w dół
`ceiling()` - zaokrąglenie liczby w górę
`c %% d` - reszta z dzielenia liczby `c` przez liczbę `d`, modulo
`round()` - zaokrąglenie do najbliższej liczby całkowitej
`sign()` - zwraca (-1) dla liczb ujemnych i 1 dla dodatnich i 0 dla zera

6. Stałe

`pi` $= \pi = 3.14159$
`Inf` - nieskończoność
`NaN` - brak danych

7. Generowanie macierzy i wektorów

`matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE)` - tworzy macierz
`diag(wielkość)` - tworzy macierz jednostkową
`diag(macierz)` - wybiera diagonalę z macierzy

8. Funkcje dla macierzy i wektorów

`det(macierz)` - wyznacznik macierzy
`t(A)` - zwraca macierz transponowaną do macierzy `A`
`length(wektor)` - zwraca długość wektora
`dim(macierz)` - zwraca wektor z wymiarami macierzy
`nrow(macierz)` - zwraca liczbę wierszy macierzy
`ncol(macierz)` - zwraca liczbę kolumn macierzy
`eigen(macierz)` - zwraca wartości i/lub wektory własne macierzy
`sum(wektor)` - suma elementów wektora lub macierzy
`min(wektor)` - wyznacza minimalną wartość z wektora
`max(wektor)` - wyznacza maksymalną wartość z wektora
`prod(wektor)` - wyznacza iloczyn wszystkich elementów wektora
`cumsum(wektor)` - wylicza sumę narastającą
`mean(wektor)` - wylicza średnią z wektora
`median(wektor)` - wyznacza medianę z wektora
`sd(wektor)` - wylicza odchylenie standardowe z wartości wektora
`which` - funkcja znajduje elementy macierzy spełniające warunek

10. Łączenie macierzy i wektorów

`c(v1, v2)` - łączenie wektorów `v1` i `v2`
`rbind(v1, v2)` - łączenie wektorów `v1` i `v2` wiersz pod wierszem
`cbind(v1, v2)` - łączenie wektorów `v1` i `v2` kolumna obok kolumny