

RAMKA DANYCH (= data_frame)

Zad.1

Używając funkcji *data.frame(...)* - której argumentami są nazwy kolejnych kolumn przyrównane do wektora wartości jakie odpowiednia kolumna ma przyjmować - utwórz ramkę *dane_* posiadającą kolumny:

'Lp' o zawartości 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

'Imię' o zawartości 'Jan', 'Robert', 'Anna', 'Magda', 'Kamil', 'Franek', 'Michał', 'Dorota', 'Agnieszka', 'Marek'

'Wiek' o zawartości 18, 27, 41, 19, 68, 33, 29, 20, 42, 40

'Odp' o zawartości 'T', 'N', 'N', 'T', 'T', 'N', 'N', 'N', 'T', 'T'.

Wywołaj ramkę *dane_*. Utwórz na dysku lokalnym nowy dokument tekstowy (plik z rozszerzeniem .txt) i wygeneruj w nim 'na piechotę' (tj. używając spacji i enter'ów) tabelkę taką jak powyżej. Zimportuj do R'a utworzony plik używając funkcji *read.table('..nazwa pliku z rozszerzeniem..', header=TRUE)*.

// np. *imported_ = read.table(...)*

Wywołaj tą zmienną. Zwróć kolumnę *Lp*, potem kolumnę *Wiek*. W końcu zwróć trzy pierwsze wiersze razem z ostatnim z tej tabelki.

Zad.2

Spróbuj wywołać bezpośrednio kolumny *Lp*, *Imię*. Używając komendy *attach(..data frame name..)* dołącz do bieżącego środowiska tabelkę przechowywaną przez zmienną *imported_*. Spróbuj ponownie wywołać bezpośrednio kolumnę *Lp*, *Imię*. Używając polecenia *detach(..data frame name..)* odłącz ze środowiska w/w tabelkę.

ROZKŁADY ZMIENNYCH LOSOWYCH

Nazwy rozkładów: beta, binom, cauchy, chisq, exp, gamma, geom, norm, pois, t, wilcox,

// > ..distribution name(..argument.., ..parameters..) //

Poprzedzając nazwę rozkładu literką:

d – uzyskamy funkcję gęstości o zadanym rozkładzie

p – uzyskamy dystrybuantę

q – uzyskamy funkcję kwantylową(*)

r – uzyskamy generator liczb losowych z zadanego rozkładu

*Przypomnijmy, że liczbę x_α nazywamy kwantylem rzędu α zmiennej losowej X jeżeli spełnia warunki:

i) $P(X \leq x_\alpha) \geq \alpha$

ii) $P(X \geq x_\alpha) \geq 1 - \alpha$

Np. $>qnorm(0.5)$ zwróci kwantyl rzędu 0.5 rozkładu normalnego standardowego, tj. taką wartość $x_{1/2}$, dla której $P(N(0,1) \leq x_{1/2}) = 0.5$.

Zad. 3

a) znajdź pr-stwo, że snajper skuteczny w 7 na 10 oddanych strzałów, strzelając 20 razy do jabłoni, strąci do koszyka dokładnie 11 jabłek.

b) j.w. ale tym razem, że strąci co najmniej 11 jabłek

c) znajdź pr-stwo, że zmienna losowa o rozkładzie Poissona z parametrem $\lambda=0.1$ przyjmie wartość 2

d) j.w. ale tym razem, że przyjmie wartość mniejszą niż 2

e) policz kwantyl rzędu $\alpha=0.995$ dla rozkładu t-Studenta z 5-ma stopniami swobody

// ..q+distribution name(..order value.., df = ...) //

- f) policz wartość funkcji gęstości oraz wartość dystrybuanty rozkładu normalnego standardowego w argumentcie 0
- g) policz kwantyl rzędu $\alpha=0.95$ dla rozkładu Chi kwadrat z 15-ma stopniami swobody
- h) wygeneruj losowy zestaw 30 liczb z populacji w której cecha ma rozkład normalny ze średnią 50 i odchyleniem standardowym 5a
- i) wygeneruj pochodzący z rozkładu wykładniczego (z parametrem $\lambda = 2$) zestaw 100 liczb

Zad.4

Używając funkcji *sample()*, która generuje wektor danych losowych z innego wektora, zasymuluj:

- a) 10 rzutów sześcienną kostką do gry (dane mogą – a w tym przypadku nawet muszą – się powtarzać).
Jako pierwszy argument wstaw wektor możliwych rezultatów pojedynczego rzutu, jako drugi – liczbę rzutów, natomiast jako trzeci: atrybut *replace=TRUE*.
- b) 8 dwukrotnych (jeden po drugim) rzutów monetą, gdzie orła interpretujemy jako sukces ('1') natomiast reszkę jako porażkę ('0') // np. '1_0' niech oznacza że w konkretnym dwurzucie otrzymaliśmy najpierw orła a tuż po tym reszkę //

Zad.5

Wygeneruj przykładową 14-wyrazową wariacjębar

- a) bez powtórzeń
 - b) z powtórzeniami
- zbioru 19-elementowego {23, 29, 35, 41, ..., 125, 131}.

WYKRESY SŁUPKOWE, KOŁOWE

Zad.6

Używając funkcji *round()* oraz *rnorm()* ze średnią 4 i odchyleniem standardowym 1.5 wygeneruj wektor *tv_* osiemdziesięciu liczb będących czasem spędzonym przez 80 respondentów przed telewizorem. Analogicznie wygeneruj wektor *pc_* 80-ciu danych z rozkładu normalnego ze średnią 6 i odchyleniem 2, będących liczbami godzin spędzanych przed komputerem.

- a) użyj funkcji *barplot()* na każdym z wygenerowanych wektorów
- b) użyj funkcji *barplot()* złożonej z funkcją *sort()* na powyższych wektorach
- c) użyj funkcji *barplot()* na argumentcie *table(tv_, tv_)* a potem na argumentcie *table(pc_, pc_)*, zinterpretuj wykresy.
- d) użyj w końcu *barplot()* z trzema argumentami: *table(pc_, tv_)*, *legend.text=TRUE* oraz *col=rainbow(12)*;

Następnie, przy użyciu atrybutu *ylim=c(..., ...)* ustaw żądane granice wartości dla osi y.

- e) wywołaj polecenie *axis(2, at = seq(...))* aby zagęścić odpowiednio miarkę na osi pionowej
- f) poleceniem *abline(h = seq(...))* nanieś na wykres poziome linie ułatwiające odczyt danych, zinterpretuj otrzymany wykres

Zad.7

- a) Używając funkcji *pie()* z argumentem *table(tv_)* a następnie na argumentcie *table(pc_)*, utwórz wykresy kołowe ilości godzin spędzanych przed telewizorem i komputerem, odpowiednio.
- b) Chcemy ulepszyć wykres kołowy z informacją dot. godzin spędzonych przed tv.
 - i) utwórz faktor *f_tv*,
 - ii) pod zmienną *labels_* zapisz wartości unikatowe faktora, natomiast pod zmienną *weights_* wagi odpowiadające poszczególnym wartościom unikatowym (wykorzystaj w tym celu funkcję *table(...)* oraz *as.vector(...)*)
 - iii) utwórz zmienną *percentage_*, która przechowuje procentowy udział wag poszczególnych unikatów
 - iv) nadpisz zmienną *labels_ = paste(labels_ , '___', percentage_)*

v) nadpisz ponownie: `labels_ = paste(labels_ , '%', sep='')`

Teraz ponownie utwórz wykres kołowy dla danych trzymanych w zmiennej `weights_` i etykiet (`labels`) przechowywanych w wektorze `labels_`. Atrybutem `col=rainbow(...)` ustaw wybrany wachlarz kolorów natomiast atrybutem `main='...'` nadaj wykresowi tytuł.

HISTOGRAMY

Zad.8

Używając komendy `data(rivers)` załaduj dane dotyczące długości rzek w USA. Wywołaj zmienną `rivers` i histogram zestawu danych przy użyciu funkcji `hist(...)`.

Wygeneruj histogram ponownie wymuszając

a) żadaną liczbę przedziałów (klas) = 40 (wartość 40 jako drugi argument)

b) zadane z góry klasy, wstawiając tym razem w miejsce drugiego argumentu `breaks=seq(0, 400, by=200)`

Popraw pionową oś tak, by jej wartości sięgały do 100 oraz zageść miarkę aby zaznaczała co 10 jednostek.

Nanieś też poziome linie (np. co 5 jednostek).

Zad.9

Sporządź wykres gęstości (uciągłony histogram) dla zestawu danych, składając funkcję `plot()` z funkcją `density()` wywołaną na argumencie `rivers`.