LABORATORIA 1

**zadanie 1 - Obliczyć wartości następujących wyrażeń w środowisku R:**

------------------------------------------------------------------------------------------------------------------------------------------

**# a)**

4 \* 5^2 + log(30, 3)

**# b)**

7^(1/5)

**# c)**

(6^(1/7))^(1/3)

**zadanie 2 - Dane są macierze A i B. Zdefiniować macierze i obliczyć tam gdzie to możliwe:**

------------------------------------------------------------------------------------------------------------------------------------------

**# MACIERZ A**

# 1 sposob tworzenia macierzy

W1 = c(3, 2, -1) # wiersz pierwszy

W2 = c(4, 1, 0) # wiersz drugi

A1 = rbind(W1, W2) # łączymy wiersze w całośc (w macierz A)

# 2 sposob tworzenia macierzy

k1 = c(3, 4)

k2 = c(2, 1)

k3 = c(-1, 0)

A2 = rbind(k1, k2, k3)

# 3 sposob

A = matrix(c(3, 4, 2, 1, -1, 0), 2, 3) # --> MACIERZ A

**# MACIERZ B**

# 1 sposob tworzenia macierzy

W1 = c(7, 2, 12) # wiersz pierwszy

W2 = c(-11, -6, 1) # wiersz drugi

W3 = c(3, -2, 3) # wiersz trzeci

B = rbind(W1, W2, W3) # łączymy wiersze w całośc (w macierz B)

# macierz A (wyznacznik)

det(A) # --> nie wyjdzie bo musi byc kwadrtatowa maciewrz

# macierz B (wyznacznik)

det(B) # --> wyjdzie bo musi byc kwadrtatowa maciewrz

# macierz A (odwrotne macierze)

solve(A)

# macierz B (odwrotne macierze)

solve(B)

# macierz A transponowana i ich wyznacznik

t(A)

det(t(A))

# macierz B transponowana i ich wyznacznik

t(B)

det(t(B))

# iloczyn macierzy A i B

A %\*% B

B %\*% A # NIE WOLNO

A %\*% A # NIE WOLNO

B %\*% B

# iloczyn skalarny

# wiersz pierwszy macierzy A

# W1 = c(3, 2, -1)

A[1,]

# druga kolumna macierzy B

# V2 = c(2, -6, -2)

B[,2]

A[1,] %\*% B[,2]

**zadanie 3 - Wykorzystując zapis macierzowy rozwiązać układ równań (użyć /solve/):**

------------------------------------------------------------------------------------------------------------------------------------------

# 2 4 -3 1 # 71

# 3 -2 8 -11 # -20

# 1 3 2 5 # 26

# 4 -3 -5 -3 # 49

A = matrix(c(2, 3, 1, 4, 4, -2, 3, -3, -3, 8, 2, -5, 1, -11, 5, -3), 4, 4) # --> niewiadome

B = matrix(c(71, -20, 26, 49), 4, 1) # --> wyrazy wolne

solve(A, B)

**zadanie 4 - Utworzyć wektor kwadratów liczb od 1 do 80, a następnie ustalić, które cyfry oraz jak często występują na pozycji jedności w wyznaczonych kwadratach (użyć operatora modulo oraz funkcji /summary/ i /factor/).**

------------------------------------------------------------------------------------------------------------------------------------------

# 1. Wektor kwadratów liczb od 1 do 80

wektor\_kwadratow <- (1:80)^2

# 2. Wyciągnięcie ostatnich cyfr (modulo 10)

ostatnie\_cyfry <- kwadraty %% 10

# 3. Przekształcenie na factor i podsumowanie

w <- factor(ostatnie\_cyfry, levels = 0:9)

podsumowanie <- summary(factor(ostatnie\_cyfry)

**zadanie 5 - Utworzyć tablice trygonometryczne, w których zebrane będą informacje o wartościach funkcji sinus, cosinus, tangens i cotangens dla kątów od 30 stopni do 60 stopni co 5stopni (funkcje trygonometryczne w R przyjmują argumenty w radianach). W tym celu napisać funkcję rad (użyć function) do zamiany stopni na radiany (stała π w R ma nazwę pi), utworzyć wektor argumentów w radianach oraz ramkę danych Tablice (użyć data.frame).**

------------------------------------------------------------------------------------------------------------------------------------------

rad = function(stopnie)

{

stopnie\*pi/180

}

stopnie = seq(30, 60, by = 5)

tablica = data.frame(

stopnie,

sin = sin(rad(stopnie)),

cos = cos(rad(stopnie)),

tan = tan(rad(stopnie)),

ctan = 1/(tan(rad(stopnie)))

)

**zadanie 6 - Utworzyć wektor 40 łańcuchów znaków następującej postaci: litera.liczba, gdzie litera to trzy duże litery X, Y, Z występujące cyklicznie, a liczba to kolejne liczby od 1 do 40 czyli X.1 Y.2 Z.3 X.4 itd. Wykorzystać funkcję paste, która łączy napisy**

------------------------------------------------------------------------------------------------------------------------------------------

lit <- rep(c("X", "Y", "Z"), length.out=40)

nums <- 1:40

lan <- paste(lit, nums, sep=".")

print(noquote(lan))

LABORATORIA 2

TEORIA – lab2

# ZMIENNE LOSOWE TYPU SKOKOWEGO #

# ------------------------------------------------------------------------------

# WYBRANE ROZKŁADY TYPU SKOKOWEGO

# ------------------------------------------------------------------------------

# 1) binom(x, n, p) --> rozkład dwumianowy (binomialny)

# - NIEZALEZNE PROBY, WYNIKI: SUKCES/PORAZKA

# /x/ - liczba sukcesow (dokladna) | p - mniejsza, rowna

# /n/ - liczba niezaleznych prób z wynikami:sukces/porazka

# /p/ - pp sukcesu

# kiedy: policzyc liczbe sukcesow (x) w skonczonej liczbie niezaleznych prob (n) (sukces/porazka)

# np.: rzucamy monetą 10 razy: ile razy wypadnie orzeł?

# ------------------------------------------------------------------------------

# 2) geom(n, p) --> rozkład geometryczny

# - PIERWSZY SUKCES PO OKRESLONEJ LICZBIE PORAZEK

# /n/ - liczba prób/porazek (przed pierwszym sukcesem)

# /p/ - pp sukcesu

# kiedy: ile porazek niepowodzen /k/ przed pierwszym sukcesem

# np.: ile razy trzeba rzucić kostką, by pierwszy raz wypadła 6?

# ------------------------------------------------------------------------------

# 3) pois(x, λ(lambda)) --> Rozkład Poissona

# x(d) - dokladnia liczba zdarzen/sukcesow = x | (p) mniejsza lub rowna x

# λ (lambda) - średnia liczba zdarzeń na jednostkę czasu/przestrzeni

# kiedy: policzyć liczbę zdarzeń w dnaym okresie gdy coś dzieje sie w

# czasie/przestrzeni z pewną średnią częstością λ

# np.: Ile telefonów przychodzi na infolinię w ciągu godziny?

# PREFIKSY

# ------------------------------------------------------------------------------

# d - funkcja rozkladu (DOKLADNA WARTOSC prawdopodobieństwo dla konkretnej wartości)

# ------------------------------------------------------------------------------

# p - wartosc dystrybuanty (NIE DOKLADNA, pp że zmienna (co najwyzej) ≤ x)

# P(X<= 8) ppois(8, lambda = 6)

# P(x > 8) = 1 - P(X<= 8) 1 - ppois(8, 6)

# ------------------------------------------------------------------------------

# q – wartość kwantyla (jaki wynik x (sukcesow) daje określone prawdopodobieństwo)

# ------------------------------------------------------------------------------

# r – generator liczb losowych (Gdy chcesz wygenerować dane losowe z danego

# rozkładu)

# ------------------------------------------------------------------------------

ZADANIA – lab2

**zadanie 1 - Prawdopodobieństwo, że przeciętny student nie zrobi pewnego zadania na kolokwium wynosi 3/7. Nauczyciel wybiera przypadkowo 5 prac różnych studentów.Wyznaczyć rozkład prawdopodobieństwa liczby osób spośród wybranych, które nie zrobiły tego zadania.Sprawdzić, czy rozkład jest dobrze określony.**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład dwumianowy - NIEZALEZNE PROBY, WYNIKI: SUKCES/PORAZKA

# /x/ - liczba sukcesow --> 0:5 (od 0 (wszyscy zrobili zadanie) do 5 (nikt

# nie zrobił zadania)) (jakie pp ze 2 nie zrobi: x = 2)

# /n/ - liczba prób --> 5 studentów (5 prac)

# /p/ - pp sukcesu --> 3/7 (sukces: NIE zrobienie zadania)

# Wyznaczyć rozkład prawdopodobieństwa liczby osób spośród wybranych, które nie zrobiły tego zadania.

dbinom(0:5,5,3/7)

# Sprawdzić, czy rozkład jest dobrze określony (wszystkie pp sumują się do 1)

sum(dbinom(0:5,5,3/7))

#teraz przeprowadzimy analizę dla wszystkich możliwych sukcesów

dbinom(0:5,5,3/7)

# Przedstawimy to w postaci macierzowej

rozklad=rbind(x\_i=0:5, p\_i=dbinom(0:5,5,3/7))

**zadanie 2 - W pewnej rodzinie (dokladnie) dwoje spośród trojga dzieci urodziło się w środę. Jakie jest prawdopodobieństwo takiego zdarzenia?**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład dwumianowy - NIEZALEZNE PROBY, WYNIKI: SUKCES/PORAZKA

# /x/ - liczba sukcesow --> 2 (2 z 3) (dokladnie)

# /n/ - liczba prób --> 3 (3 dzieci)

# /p/ - pp sukcesu --> 1/7 (sukces: urodzenie się w środę)

dbinom(2,3,1/7)

# KOLOKWIUM #

**zadanie 3 - Prawdopodobieństwo awarii pewnego urządzenia podczas uruchomiania wynosi 0.003. Obliczyć prawdopodobieństwo, że pierwsza awaria zdarzy się przy dokladnie szóstym włączeniu.**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład geometryczny - PIERWSZY SUKCES PO OKRESLONEJ LICZBIE PORAZEK

# /n/ - liczba prób/porażek (przed pierwszym sukcesem) --> 5 (bo szósty to awaria (sukces)

# /p/ - pp sukcesu --> 0.003 (sukces: awaria)

# kiedy: ile porazek/niepowodzen przed pierwszym sukcesem

dgeom(5,0.003) # (odp.: 0.002955269)

**zadanie 4 - W skład pewnej wtryskarki wchodzi 300 elementów określonego rodzaju. Prawdopodobieństwo uszkodzenia w ciągu miesiąca każdego z tych elementów wynosi 0.002 i nie zależy od stanu pozostałych. Obliczyć prawdopodobieństwo uszkodzenia w ciągu miesiąca:**

**a) dokładnie trzech elementów,**

**b) nie więcej niż trzech elementów.**

**W obu podpunktach obliczyć przybliżenie rozkładem Poissona.**

------------------------------------------------------------------------------------------------------------------------------------------

# Rozkład Poissona (x, lambda)

# x - liczba zdarzen/sukcesow (dokladnie 3 czyli d)

# λ - średnia liczba zdarzeń 300 \* 0,002

# a) dokładnie trzech elementów, (dokladna wartosc - d)

300 \* 0.002

dpois(3, 0.6)

# b) nie więcej niż trzech elementów (nie wiecej niz x <= 3)

ppois(3, 300 \* 0,002)

# lub

1 - pbinom(3, 600, 0.004)

**zadanie 5 - Wadliwość produkowanych w pewnej firmie kości pamięci wynosi 0.4%. Pobrano losowo do kontroli partię 600 kości pamięci. Obliczyć prawdopodobieństwo, że liczba uszkodzonych kości pamięci jest większa niż 3.**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład dwumianowy (binomialny) - NIEZALEZNE PROBY, WYNIKI: SUKCES/PORAZKA

# p - pytaja o wieksza niz 3

# /x/ - liczba sukcesow --> liczba uszkodzonego sprzetu (od 4 do 600)

# próba - sprawdzenie jednego sprzetu

# /n/ - liczba prób --> 600 (liczba sprzetu do kontroli)

# sukces - sprzezt jest wadliwy

# porazka - sprzezt nie jest wadliwy

# /p/ - pp sukcesu --> 0.4% = 0.004 (sukces: uszkodzony sprzęt)

# P(X>3) = 0.2210252

sum(dbinom(4:600,600,0.004))

# 1-F(3) = 1-P(X<=3)

1-pbinom(3,600,0.004)

#lub

pbinom(3,600,0.004, lower.tail=FALSE)

#lub

pbinom(3,600,0.004,0)

**zadanie 6 - Rzucamy jednocześnie trzema monetami aż wypadną trzy orły. Jakie jest prawdopodobieństwo, że będziemy musieli rzucać więcej niż 5 razy?**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład geometryczny - PIERWSZY SUKCES PO OKRESLONEJ LICZBIE PORAZEK

# p - PYTAJA O WIECEJ NIZ 5 RZUTY

# /n/ - liczba prób/porazek (przed pierwszym sukcesem) --> 4 bo rzut 1,2,3,4

# /p/ - pp sukcesu --> 1/8 (sukces: wypadną trzy orły)

1-pgeom(4,1/8)

#lub

pgeom(4,1/8,0)

**zadanie 7 - Pewne urządzenie zawiera 650 lamp. Prawdopodobieństwo przepalenia dowolnej lampy w ciągu jednej doby pracy urządzenia jest jednakowe i wynosi 0.003. Obliczyć prawdopodobieństwo, że w ciągu jednej doby pracy urządzenia przepalą się co najmniej 2 lampy.**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład dwumianowy (binomialny)

# /x/ - liczba sukcesow --> 1 (co najmniej 2 lampy)

# /n/ - liczba prób/mozliwosci --> 650 (liczba lamp)

# /p/ - pp sukcesu --> 0.003 (sukces: przypalenie lampy)

# P(X>=2) = 0.5806874

sum(dbinom(2:650,650,0.003))

# 1-F(1) = 1-P(X<=1)

1-pbinom(1,650,0.003)

# P(X>1)

pbinom(1,650,0.003,0)

# lub

# n = 650 p= 0.003 lambda = n\*p=1.95

# ppois(conajmniej 2 , lambda)

1 - ppois(1,1.95)

**zadanie 8 - Rzucamy jednoczenie dwiema kostkami aż na obu wypadnie co najmniej 5 oczek. Obliczyć prawdopodobieństwo, że zdarzy się to:**

**a) w trzecim rzucie,**

**b) w co najmniej drugim, ale nie później niż w siódmym rzucie.**

------------------------------------------------------------------------------------------------------------------------------------------

# a) w trzecim rzucie,

# /n/ - liczba prób/porazek (przed pierwszym sukcesem) --> 2 (1 rzut (porazka) 2 rzut (porazka) 3 rzut (sukces))

# 6\*6=36, kombinacje: 55,56,65,55, czyli 4/36=1/9

# /p/ - pp sukcesu --> 1/9 (4/36) (sukces: wypada 5 oczek)

dgeom(2,1/9)

# b) w co najmniej drugim, ale nie później niż w siódmym rzucie.

# od 2 do 7 czyli od 1 do 6 porażek

sum(dgeom(1:6,1/9))

**zadanie 9 - W centrali telefonicznej jest 1000 linii, które działają niezależnie od siebie. Prawdopodobieństwo tego, że linia nie jest zajęta wynosi 0.88. Obliczyć prawdopodobieństwo, że liczba zajętych linii różni się od 100 o mniej niż 15.**

------------------------------------------------------------------------------------------------------------------------------------------

#prawdopodobieństwo sukceku (linia zajęta) = 0.12

#ilość sukcesów między 86 a 114

# ilość prób 1000

sum(dbinom(86:114,1000,0.12))

LABORATORIA 3

TEORIA – lab3

# ZMIENNE LOSOWE TYPU CIĄGŁEGO #

# WYBRANE ROZKŁADY TYPU CIĄGŁEGO: (nie ma znaczenia >, < czy >=, <= bo punktowe nie ma sensu)

# ------------------------------------------------------------------------------

# norm --> standardowy rozkład Gaussa

# rozkład normalny np.: N(3,6):

# /𝜇/ średnia = 3

# /𝜎/ odchylenie standardowe = 6

# t --> rozkład t-Studenta

# liczba stopni swobody (df)

# chisq --> rozkład chi-kwadrat

# f --> rozkład F

# PREFIKSY:

# ------------------------------------------------------------------------------

# d – funkcja gęstości

# p – wartość dystrybuanty (prawdopodobienstwo)

# μ,σ -> srednia, standardowa

# P(X ≤ a) -> pnorm(a, μ, σ)

# P(X > a) -> pnorm(a, μ, σ, lower.tail=FALSE)

# P(a < X ≤ b) -> pnorm(b,μ,σ)−pnorm(a,μ,σ)

# q – wartość kwantyla

# r – generator liczb losowych

# ------------------------------------------------------------------------------

ZADANIA – lab3

# KOLOKWIUM

**# zadanie 1 - Obliczyć kwantyle:**

------------------------------------------------------------------------------------------------------------------------------------------

# a) u(0.98) (odp. 2.053749)

qnorm(0.98)

qnorm(0.07)

# b) t(0.95, 18) (odp. 1.734064)

qt(0.95,18)

# c) X^2(0.975, 23) (odp. 38.07563)

qchisq(0.975,23)

# d) F(0.99, 5, 24) (odp. 3.89507)

qf(0.99,5,24)

**# zadanie 2 - Zmienna losowa X ma rozkład normalny N(3,6). Obliczyć prawdopodobieństwo:**

------------------------------------------------------------------------------------------------------------------------------------------

# rozkład normalny N(3,6): (norm)

# srednia --> 3

# odchylenie standardowe --> 6

# a) P(X<5) (odp.: F(5) = 0.6305587) # P(X ≤ a) -> pnorm(a, μ, σ)

pnorm(5,3,6)

# b) P(X>4) (odp.: 1-F(4) = 0.4338162) # P(X > a) -> 1 - pnorm(a, μ, σ)

1-pnorm(4,3,6)

# lub

pnorm(4,3,6,0)

# c) P(-1<X<=1) (odp.: F(1)-F(-1)= 0.1169488) # P(a < X ≤ b) -> pnorm(b,μ,σ)−pnorm(a,μ,σ)

pnorm(1,3,6)-pnorm(-1,3,6)

# d) P(4-0.5<=X<=4+0.5) (odp.: F(4.5)-F(3.5) = 0.06549957)

pnorm(4.5,3,6)-pnorm(3.5,3,6)

# e) P(8-1<3X<8+1) (odp.: P(7/3<X<3) = F(3)-F(7/3) = 0.04423588)

pnorm(3,3,6)-pnorm(7/3,3,6)

# f) P(X<=-1-7)+P(X>=-1+7) = F(-8)+1-F(6) = 0.341914

pnorm(6,3,6,0)+pnorm(-8,3,6)

# g) P(2X < 3-4)+P(2X > 3+4) = F(-1/2) + 1-F(7/2) = 0.7466277

pnorm(-1/2,3,6)+pnorm(7/2,3,6,0)

# KOLOKWIUM

**Zadanie 3 - Czas świecenia żarówek pochodzących z masowej produkcji jest zmienną losową X o rozkładzie normalnym N(200 h, 10 h). Oblicz, ile przeciętnie żarówek spośród 10000 świeci krócej niż 175 h.**

------------------------------------------------------------------------------------------------------------------------------------------

# P(X<175) pp dla jednej zarowki (# P(X ≤ a) -> pnorm(a, μ, σ))

pnorm(175,200,10)

# pp dla 10 000

10000\*pnorm(175,200,10)

#62 żarówki

**# Zadanie 4 - Przy założeniu, że wyniki w skoku wzwyż mężczyzn mają rozkład normalny z parametrami 2.25 m oraz 0.2 m, obliczyć:**

------------------------------------------------------------------------------------------------------------------------------------------

# N(2.25, 0.2)

# μ,σ -> srednia, standardowa

# P(X ≤ a) -> pnorm(a, μ, σ)

# P(X > a) -> 1 - pnorm(a, μ, σ)

# P(a < X ≤ b) -> pnorm(b,μ,σ)−pnorm(a,μ,σ)

**a) ilu zawodników na 40 osiągnie w skoku wzwyż co najmniej 2.3 m,**

(odp.: 16 zawodników) X >= 2.3 (najmniej tyle)

1 - pnorm(2.3 , 2.25 , 0.2)

40\*pnorm(2.3 ,2.25 ,0.2, 0)

**b) jaki jest wynik uzyskany przez zawodników, poniżej którego jest 20% najsłabszych rezultatów?**

(P(X<x) = 0.2 x-kwantyl(q) rzędu 0.2)

qnorm(0.2, 2.25, 0.2)

#pułap 2.08

# P(X<x) = 0.2 x-kwantyl rzędu 0.2

**Zadanie 5 - Przyjmując, że opóźnienie pociągu do Poznania jest zmienną losową o rozkładzie normalnym N(13 min, 18 min), obliczyć prawdopodobieństwo, że pociąg, który miał przyjechać o 14.25 przyjedzie:**

------------------------------------------------------------------------------------------------------------------------------------------

# N(13 min, 18 min)

# μ,σ -> srednia, standardowa

# P(X ≤ a) -> pnorm(a, μ, σ)

# P(X > a) -> 1 - pnorm(a, μ, σ)

# P(a < X ≤ b) -> pnorm(b,μ,σ)−pnorm(a,μ,σ)

# X = 14:25

# 14:40(25+15) - 14:45(25+20) --> 15 < X < 20

# P(a < X ≤ b) -> pnorm(b,μ,σ) − pnorm(a,μ,σ)

**c) między 14.40 a 14.45, (opoznienie 15 - 20 min)**

pnorm(20, 13, 18) - pnorm(15, 13, 18)

**d) po 14.50**

P(X > 25)

1 - pnorm(25, 13, 18)

**Zadanie 6 - Zmienna losowa ma rozkład N(25, 8). Wyznaczyć nieznane wartości całkowite k1, k2, k3, k4, jeżeli wiadomo, że zmienna ta przyjmuje wartość:**

------------------------------------------------------------------------------------------------------------------------------------------

**a) mniejszą niż k1 z prawdopodobieństwem 0.5987, --> P(X< k) = 0.5987**

k1 <- qnorm(0.5987, mean = 25, sd = 8)

**b) większą od k2 z prawdopodobieństwem 0.734, --> P(X>k) = 0.734**

k2 <- qnorm(0.266, mean = 25, sd = 8)

**c) odchylającą się od średniej nie więcej niż o k3 z prawdopodobieństwem 0.468, --> P(|X-25|<= k) 0.468**

z <- qnorm((1 + 0.468) / 2) # bo to symetrycznie wokół średniej

k3 <- z \* 8

**d) odchylającą się od średniej nie mniej niż o k4 z prawdopodobieństwem 0.617.--> P(|X-25|>= k) 0.617**

z <- qnorm((1 + 0.383) / 2)

k4 <- z \* 8

------------------------------------------------------------------------------------------------------- a) rozkład normalny N(0,1)

ggplot(data.frame(x=c(-5,5)), aes(x))+

stat\_function(fun=dnorm,col="blue",size=1.25)+

ylab("Gęstość rozkładu normalnego N(0,1)")

#u~N(0,1)

# ROZKLAD POSTACI

# N(0,1)

# 0 - srednia,

# 1 - odchylenie stardandowe

-------------------------------------------------------------------------------------------------- b) ROZKLAD TYPU T STUDENTA

ggplot(data.frame(x=c(-5,5)), aes(x))+

stat\_function(fun=dt,args=list(df=18),col="green",size=1.25)+

ylab("Gęstość rozkładu t Studenta")

# df - > stopnie swobody

------------------------------------------------------------------------------------------------------- c) ROZKLAD CHI KWADRAT

ggplot(data.frame(x=c(-2,75)), aes(x))+

stat\_function(fun=dchisq,args=list(df=20),col="red",size=1.25)+

ylab("Gęstość rozkładu chi kwadrat")

# df - > stopnie swobody

-------------------------------------------------------------------------------------------------------------- d) ROZKLAD FISHERA

ggplot(data.frame(x=c(-2,7)), aes(x))+

stat\_function(fun=df,args=list(df1=5,df2=24),col="brown",size=1.25)+

ylab("Gęstość rozkładu Fishera")

# df - > stopnie swobody

LABORATORIA 4

TEORIA – lab4

# zarzadzanie danymi #

# wywołanie kolumny w ramce np. Ankieta$Płeć

ZADANIA – lab4

**Zadanie 1 - Opracowanie ankiety**

------------------------------------------------------------------------------------------------------------------------------------------

**Zadanie 2**

**a) wyświetlić podsumowanie danych przed i po faktoryzacji zmiennych niemierzalnych**

------------------------------------------------------------------------------------------------------------------------------------------

summary(Ankieta)

# faktoryzacja zmiennych niemierzalnych

Ankieta$Płeć=factor(Ankieta$Płeć)

Ankieta$M.zamieszkania=factor(Ankieta$M.zamieszkania)

Ankieta$Sz.średnia=factor(Ankieta$Sz.średnia)

Ankieta$System=factor(Ankieta$System)

summary(Ankieta)

# ------------------------------------------------------------------------------

#(usunac pierwszy # wykonac instrukcje i ponownie postawic

# Ankieta=Ankieta[-32,] #usuwanie wiersza nr 32

# ------------------------------------------------------------------------------

**b) w ramce Ankieta utworzyć nową zmienną Średnia, zawierającą średnią ocen z kursów;**

------------------------------------------------------------------------------------------------------------------------------------------

Ankieta$Średnia = (Ankieta$Algebra + Ankieta$MSzS1 + Ankieta$Narz.inż + Ankieta$Prog1 + Ankieta$WdI)/5

summary(Ankieta)

**c) przenieść kolumny z ocenami z kursów do podzbioru Ankieta.kursy (użyć subset)**

------------------------------------------------------------------------------------------------------------------------------------------

Ankieta.kursy=subset(Ankieta,select=Algebra:WdI)

summary(Ankieta)

# Ankieta=Ankieta[,-(7:11)] #uzywamy tego raz (usuwamy kolumny)

**d) napisać funkcję zakres3sigm, która zwróci dla dowolnej zmiennej ramkę danych z nagłówkami lewy.kres / prawy.kres jako średnią minus / plus trzy odchylenia standardowe (użyć function, mean, sd, data.frame)**

------------------------------------------------------------------------------------------------------------------------------------------

zakres3sigm = function(x)

{

lewy.kres = mean(x)-3\*sd(x)

prawy.kres = mean(x)+3\*sd(x)

data.frame(lewy.kres, prawy.kres)

}

**e) dla zmiennej Średnia wyznaczyć ewentualne dane odstające i zastąpić je symbolem braku danych, a później średnią, zaokrągloną do części dziesiętnych (przy dużej liczbie danych użyć which)**

------------------------------------------------------------------------------------------------------------------------------------------

zakres3sigm(Ankieta$Średnia) # przedzial

summary(Ankieta$Średnia) # zakres danych

# na pierwszy rzut oka dane sa okej, bo mieszcza sie w przedziale

**f) utworzyć podzbiory danych Ankieta.M i Ankieta.K dla mężczyzn i kobiet odpowiednio (użyć filter)**

------------------------------------------------------------------------------------------------------------------------------------------

Ankieta.M=subset(Ankieta,Płeć=="M")

summary(Ankieta.M)

Ankieta.K=subset(Ankieta,Płeć=="K")

summary(Ankieta.K)

**g) dla zmiennych Waga i Wzrost wyznaczyć ewentualne dane odstające dla obu płci, zastąpić je symbolem braku danych, a później średnią, zaokrągloną do części dziesiątych;**

------------------------------------------------------------------------------------------------------------------------------------------

zakres3sigm(Ankieta.K$`Waga [kg]`)

summary(Ankieta.K$`Waga [kg]`)

zakres3sigm(Ankieta.K$`Wzrost [cm]`)

summary(Ankieta.K$`Wzrost [cm]`)

zakres3sigm(Ankieta.M$`Waga [kg]`)

summary(Ankieta.M$`Waga [kg]`)

zakres3sigm(Ankieta.M$`Wzrost [cm]`)

summary(Ankieta.M$`Wzrost [cm]`)

# > zakres3sigm(Ankieta.M$`Waga [kg]`)

#lewy.kres prawy.kres

#1 35.43449 113.4468

#> summary(Ankieta.M$`Waga [kg]`)

#Min. 1st Qu. Median Mean 3rd Qu. Max.

# 45.00 64.75 73.50 74.44 81.50 115.00

fix(Ankieta.M)

zakres3sigm(Ankieta.M$`Waga [kg]`)

summary(Ankieta.M$`Waga [kg]`)

# calosc

fix(Ankieta)

zakres3sigm(Ankieta$`Waga [kg]`)

summary(Ankieta$`Waga [kg]`)

**h) utworzyć nową zmienną L.g.kody, w której zostaną umieszczone 3 przedziały liczbowe odpowiadające ustalonym kategoriom: krótko, średnio, długo (użyć cut) i wyświetlić liczności przedziałów**

------------------------------------------------------------------------------------------------------------------------------------------

Ankieta$L.g.kody=cut(Ankieta$L.godzin,c(0,5,10,24))

summary(Ankieta$L.g.kody)

**Zadanie 3 - Wyznaczyć histogramy dla zmiennych M.zamieszkania, Sz.średnia i System.**

------------------------------------------------------------------------------------------------------------------------------------------

ggplot (Ankieta, aes(M.zamieszkania)) +geom\_bar (fill = "grey", col = "black") + ylab ("Liczność")

ggplot (Ankieta, aes(Sz.średnia)) +geom\_bar (fill = "orange", col = "black") + ylab ("Liczność")

ggplot (Ankieta, aes(System)) +geom\_bar (fill = "pink", col = "black") + ylab ("Liczność")