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:**

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

tworzenie macierzy

1 sposob

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

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

A1 = rbind(W1, W2) # łączymy wiersze/wektory w całość

2 sposob

k1 = c(3, 4)

k2 = c(2, 1)

k3 = c(-1, 0) # kolumny jako wektory

A2 = rbind(k1, k2, k3) # łączenie kolumn/wektrów w całość

3 sposób (funkcja wbudowana)

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

B = matrix(c(7,-11,3,2,-6,12,1,3),3,3)

**a) wyznaczniki macierzy**

det(B) // det(A) --> nie wyjdzie bo musi byc kwadrtatowa macierz

**b) macierze odwrotne**

solve(A) // solve(B) --> nie wyjdzie bo musi byc kwadrtatowa macierz

**c) macierze transponowane i ich wyznaczniki**

macierz transponowana

t(A)

t(B)

wyznacznik macierzy transponowanej

det(t(A))

det(t(B))

**d) iloczyn macierzy A i B**

A %\*% B

B %\*% B

B %\*% A --> NIE WOLNO, bo jest 3x3 i 2x3 więc środkowe nie są takie same

A %\*% A --> NIE WOLNO, bo jest 2x3 i 2x3 więc środkowe nie są takie same

**e) iloczyn skalarny między pierwszym wierszem macierzy A, a drugą kolumną macierzy B**

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

# 1. niewiadome

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

# 2. wyrazy wolne

B = matrix(c(71, -20, 26, 49), 4, 1)

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 = wektor\_kwadratow %% 10

# 3. Przekształcenie na factor i podsumowanie

x = factor(ostatnie\_cyfry, levels = 0:9)

wynik = summary(x)

**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,

sinus = sin(rad(stopnie)),

cosinus = cos(rad(stopnie)),

tangens = tan(rad(stopnie)),

ctangens = 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**

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

litery = rep(c("X", "Y", "Z"), length.out=40)

numery = 1:40

lancuch = paste(litery, numery, sep=".")

print(noquote(lancuch))

LABORATORIA 2

**ZMIENNE LOSOWE TYPU SKOKOWEGO**

TEORIA – lab2

**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)

**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)

dbinom(3,300,0.002)

dpois(3, 300\*0.002)

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

pbinom(3, 300, 0.002)

ppois(3, 300 \* 0,002)

**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)

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))

pgeo(6,1/9)-pgeom(0,1/9)

**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

**ZMIENNE LOSOWE TYPU CIĄGŁEGO**

TEORIA – lab3

**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 Fishera**

**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)**

**kwantyl rozkładu normalnego** qnorm(0.98)

**b) t(0.95, 18)**

**kwantyl rozkładu t Studenta** qt(0.95,18)

**c) X^2(0.975, 23)**

**kwantyl rozkładu chi kwadrat** qchisq(0.975,23)

**d) F(0.99, 5, 24)**

**kwantyl rozkładu Fishera** qf(0.99,5,24)

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

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

# rozkład normalny N(3,6): (norm)

# μ srednia --> 3

# σ odchylenie standardowe --> 6

**a) P(X < 5)** ( F(5) = 0.6305587 ) P(X ≤ a) -> pnorm(a, μ, σ)

pnorm(5,3,6)

**b) P(X > 4)** (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)** ( 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(|X - 4| <= 0.5)** ( P(4 - 0.5 <= X <= 4 + 0.5) = F(4.5) - F(3.5) = 0.06549957)

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

**e) P(|3X - 8| < 1)** ( P(8 – 1 < 3X < 8 + 1) = 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) + P(X > = -1 + 7) = F(-8) + 1 - F(6) = 0.341914 )

pnorm(6,3,6,0)+pnorm(-8,3,6) lub 1-(pnorm(6,3,6)-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) lub 1-(pnorm(7/2,3,6)-pnorm(-1/2,3,6))

# 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.**

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

N(200,10)

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:**

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

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

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

# 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** X >= 2.3 (najmniej tyle)

(1-pnorm(2.3,2.25,0.2))\*40

round((1-pnorm(2.3,2.25,0.2))\*40) odp.: 16 zawodników

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

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) lub pnorm(25,13,18,0)

**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

**ZARZADZANIE DANYMI**

TEORIA – lab4

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")

LABORATORIA 5

**ELEMENTY STATYSTYKI OPISOWEJ**

RAMKA DANYCH: wywołanie zmiennej w ramce np. Ankieta.M$Waga

ZADANIA – lab5

**Zadanie 1 - Napisać funkcję parametry.opisowe, która dla dowolnej zmiennej wyznaczy parametry opisowe do łączenia wartości użyć rbind): srednia, kwartyl.1, mediana, kwartyl.3, min, max, rozstep.empiryczny = max(x) – min(x), rozstep.miedzykwartylowy, wariancja, odchylenie.standardowe, wspolczynnik.zmiennosci = sd(x)/mean(x),wspolczynnik.asymetrii, wspolczynnik.skupienia**

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

parametry.opisowe = function(x)

{ rbind(

srednia = mean(x),

kwartyl.1 = quantile(x, 0.25), # wartosc liczbowa ponizej kotra jest 25 procent populacji

mediana = median(x), # wartosc srodkowa

kwartyl.3 = quantile(x, 0.75), # wartosc liczbowa ponizej kotra jest 75 procent populacji

min = min(x),

max = max(x),

rozstep.empiryczny = max(x) - min(x), # jaki przedzial danych

rozstep.miedzykwartylowy = IQR(x), # kwartyl.3 - kwartyl.1

wariancja = var(x), # srednie kwadratowe odchylenie od sredniej

odchylenie.standardowe = sd(x), # pierwiastek wariancji

wspolczynnik.zmiennosci = sd(x)/mean(x),

wspolczynnik.skupienia=kurtosis(x))

}

# pierwiastek wariancji podzielic na srednia

# (ktora probka jest bardziej zroznicowana)

#wspolczynnik.asymetrii = skewness(x),

# czy wartosc jest dodatnia czy ujemna

# (dodatnia - prawostronna asymetria, ujemna - lewostronna asymetria)

#wspolczynnik.skupienia = kurtosis(x) ,

# kurtoza, wartosc rowna 3 dla ukladu normalnego

# mniej skupiony - ponizej 3

# bardziej skupiony - ponizej 3

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

# Zadanie 2 - Dla zmiennych Waga i Wzrost w grupie mężczyzn:

# ------------------------------------------------------------------------- WAGA

# a) wyznaczyć i zinterpretować parametry opisowe;

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

parametry.opisowe(Zaliczenie$Średnica)

# srednia - średnia waga zbadanych mężczyzn wynosi 74.01 kg

# kwartyl.1 - waga 25% mężczyzn nie przekracza 64.75 kg

# mediana - waga 50% mężczyzn nie przekracza 73.5 kg

# kwartyl.3 - waga 75% mężczyzn nie przekracza 81 kg

# min, max - waga minimalna wynosi 45 kg, a maksymalna 104 kg

# rozstep.empiryczny - waga mężczyzn zmienia się w zakresie 59 kg

# rozstep.miedzykwartylowy - 50% środkowych wartości wagi zmienia się w zakresie 16.25 kg

# wariancja - przeciętne kwadratowe odchylenie wagi mężczyzn od średniej wynosi 151.55 kg^2

# odchylenie.standardowe - przeciętne odchylenie wagi mężczyzn od średniej wynosi 12.31 kg

# wspolczynnik.zmiennosci - odchylenie standardowe wagi mężczyzn stanowi 16.63 % średniej

# wspolczynnik.asymetrii - rozkład wagi mężczyzn jest prawostronnie asymetryczny, gdyż wsp. as. jest dodatni

# wspolczynnik.skupienia - rozkład wagi mężczyzn jest słabiej skupiony wokół średniej niż w rozkładzie normalnym,

# gdyż kurtoza jest mniejsza niż 3

# b) narysować histogram (przyjąć szerokość klasy 10);

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

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

ggplot (Ankieta.M, aes(`Waga [kg]`)) + geom\_histogram (fill = "red", col = "black", binwidth = 10) + ylab ("Liczność")

# c) wyznaczyć i zinterpretować diagram łodyga i liście;

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

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

# d) wyznaczyć i zinterpretować wykresy ramka-wąsy z podziałem ze względu na płeć

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

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

ggplot (Ankieta, aes (x = Płeć, y = `Waga [kg]` )) + geom\_boxplot (fill = c("lightpink","lightblue"), col = "black")

# ----------------------------------------------------------------------- WZROST

# a) wyznaczyć i zinterpretować parametry opisowe;

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

parametry.opisowe(Ankieta.M$`Wzrost [cm]`)

# srednia - średnia waga zbadanych mężczyzn wynosi 180.54 cm

# kwartyl.1 - waga 25% mężczyzn nie przekracza 177.00 cm

# mediana - waga 50% mężczyzn nie przekracza 180.00 cm

# kwartyl.3 - waga 75% mężczyzn nie przekracza 185.00 cm

# min, max - waga minimalna wynosi 163.00 cm a maksymalna 195.00 cm

# rozstep.empiryczny - waga mężczyzn zmienia się w zakresie 32.00 cm

# rozstep.miedzykwartylowy - 50% środkowych wartości wagi zmienia się w zakresie 8.00 cm

# wariancja - przeciętne kwadratowe odchylenie wagi mężczyzn od średniej wynosi 37.47 cm^2

# odchylenie.standardowe - przeciętne odchylenie wagi mężczyzn od średniej wynosi 6.12 cm

# wspolczynnik.zmiennosci - odchylenie standardowe wagi mężczyzn stanowi 0.034 średniej

# wspolczynnik.asymetrii - rozkład wagi mężczyzn jest lewostronnie asymetryczny, gdyż wsp. as. jest ujemna

# wspolczynnik.skupienia - rozkład wagi mężczyzn jest bardziej skupiony wokół średniej niż w rozkładzie normalnym,

# gdyż kurtoza jest większa niż 3

# b) narysować histogram (przyjąć szerokość klasy 10);

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

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

ggplot (Ankieta.M, aes(`Wzrost [cm]`)) + geom\_histogram (fill = "red", col = "black", binwidth = 10) + ylab ("Liczność")

# c) wyznaczyć i zinterpretować diagram łodyga i liście;

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

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

# d) wyznaczyć i zinterpretować wykresy ramka-wąsy z podziałem ze względu na płeć

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

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

ggplot (Ankieta, aes (x = Płeć, y = `Wzrost [cm]` )) + geom\_boxplot (fill = c("lightpink","lightblue"), col = "black")

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

# Zadanie 3 - Dla zmiennej Wzrost w grupie mężczyzn wyznaczyć szereg rozdzielczy

# przedziałowy (krok – 10 cm, od minimum obciętego w dół z dokładnością do 10 cm,

# użyć table i cut) i utworzyć pomocnicze zmienne Wzrost.środki i Wzrost.wagi.

# Dla tak zgrupowanych danych obliczyć średnią i odchylenie standardowe stosując

# pomocniczą zmienną Wzrost.szereg (użyć rep). Czy otrzymane średnia i odchylenie

# standardowe są takie same jak parametry dla zmiennej Wzrost w grupie mężczyzn bez grupowania?

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

table(cut(Ankieta.M$Wzrost,c(160,170,180,190,200)))

Wzrost.środki=c(165,175,185,195) #srodki przedziualow

Wzrost.wagi=c(8,42,42,4) #// ile elementow w przedziale

Wzrost.szereg = #rep(Wzrost.środki,Wzrost.wagi) // powtarza konkretny element zadana liczbe razy

parametry.opisowe(Wzrost.szereg) # powstal sztuczny tymczasowy wektor i je sobie ogladamy

# zamiast danych pozbieranych, chcemy podzial zaproponowac, punkty wpadaja do przedzialow,

# liczymy kropeczki, w stuczny sposob tworzymy wektor

# czy takie stuczne wprowadzone dane (przyblizone) do srodka pewnego przedzialu cos zmieniaja

# dokladnie nie to samo ale nie istotnie odmienne wyniki ;)

# ale bledy sa (np odchylenie standardowe)

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

# Zadanie 4 - Dla zmiennej Średnia:

# a) wyznaczyć i zinterpretować parametry opisowe;

# b) wyznaczyć i zinterpretować wykresy ramka-wąsy z podziałem ze względu na płeć, miejsce zamieszkania i szkołę średnią.

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

LABORATORIA 6

# TEST ZGODNOŚCI I ESTYMACJA DLA JEDNEJ POPULACJI #

# w każdym teście:

# sformułować hipotezy, podać statystyke oraz poziom p oraz wniosek

# rozklad normalny ;)

# rozkald nie normalny ;|

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

# Zadanie 1 - Wyświetlić w jednym układzie współrzędnych gęstości rozkładów

# stosowanych w statystyce matematycznej:

# a) t Studenta (dt) dla stopni swobody: 1, 2, 10, 50 (przyjąć zakres od −4 do 4),

# b) X^2 (dchisq) dla stopni swobody: 2, 10, 20, 50 (przyjąć zakres od −1 do 100).

# Wyciągnąć wnioski odnośnie rozkładów granicznych.

# a) im wiecej stopni swobody tym bardziej jest zbiezny do ukladu normalnego

# blad popelniany nie az tak duzy

ggplot (data.frame (x = c(-4, 4)), aes (x)) +

stat\_function (fun = dt, args = 1, col = "blue", size = 1.25) +

stat\_function (fun = dt, args = 2, col = "red", size = 1.25) +

stat\_function (fun = dt, args = 10, col = "green", size = 1.25) +

stat\_function (fun = dt, args = 50, col = "brown", size = 1.25) +

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

ylab ("Gęstość rozkładu t Studenta")

# Rozkład t Studenta jest zbieżny do rozkładu normalnego standaryzowanego,

# więc rozkład t Studenta można zastąpić rozkładem normalnym dla dużej próby

# b)

# od najmniejszej do najwiekszej

ggplot (data.frame (x = c(-1, 100)), aes (x)) +

stat\_function (fun = dchisq, args = 2, col = "blue", size = 1.25) +

stat\_function (fun = dchisq, args = 10, col = "red", size = 1.25) +

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

stat\_function (fun = dchisq, args = 50, col = "brown", size = 1.25) +

ylab ("Gęstość rozkładu chi^2")

# od najmniejszej do najwiekszej (brazowa zaczyna prypominac rozklad normalny,

# mozna podmiany robic miedzy tymi rozkladami przy duzych danych)

# Rozkład chi kwadrat jest zbieżny do rozkładu normalnego,

# więc rozkład chi kwadrat można zastąpić rozkładem normalnym dla dużej próby

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

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

# Zadanie 2 - Dla zmiennej Wzrost w grupie mężczyzn:

# a) dokonać wstępnej oceny zgodności z rozkładem normalnym w populacji

# generalnej na podstawie histogramu z gęstością teoretyczną (dnorm);

ggplot (Ankieta.M, aes(`Wzrost [cm]`)) +

geom\_histogram (aes (y = ..density..), fill = "blue", col = "black", binwidth = 5) +

stat\_function (fun = dnorm, args = list (mean (Ankieta.M$`Wzrost [cm]`), sd (Ankieta.M$`Wzrost [cm]`)), col = "red", size = 1.25)+

ylab ("Częstość")

# ocena zgodnosci z rozkladem normalnym: NIE

# Przy uogólnianiu wyników badania próbnego na pop.gen,

# estymacja i testowanie parametrów wymaga zgodności z rozkładem normalnym

# Struktura testu w skrypcie:

# H0:

# H1:

# Statystyka testowa: =

# p-value =

# Wniosek:

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

# w idealnym swiecie: srodkowy wykres slupkowy a czerwona linia przez jej srodek

# narzedzie do weryfikacji czy jest poprawne i jest zgodne

# populacja generalna: WE WSZYSTKICH ZADANIACH ZAKŁADAMY, ŻE POPULACJĄ GENERALNĄ

# SĄ WSZYSCY STUDENCI I-GO ROKU NA WI

# na podstawie pewnej proby uogooniamy wnioski dla danej populacji generalnej - bo duzo danych duze koszty

# chcemy zeby winiki byly zgodne z rozkladem normalnym

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

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

# (1) SHAPIRO-WILK (CZY ROZKLAD JEST NORMALNY)

# b) na poziomie istotności 0.05 sprawdzić założenie o normalności rozkładu

# testem Shapiro Wilka (użyć shapiro.test)

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

# ----------------------------------------------------------- FORMULOWANIE TESTU

# Struktura/schemat testu w skrypcie:

# H0: ROZKLAD JEST NORMALNY (alfa < p, nie ma podstaw)

# testujemy ja, dla niej budujemy tekst

# zawsze parametr ktory testujemy jest zgodny z rozkladem normalnym

# np.: Hipoteza zerowa: rozkład wzrostu mężczyzn w pop.gen jest normalny

# H1: NIE JEST (alfa >= p, odrzucamy H0)

# Hipoteza alternatywna H1: ~H0

# skoro zakladamy ze spelnione, to potem zakladamy ze nie spolenione

# np.: Hipoteza alternatywa: nie jest normalny

# ---------------------------------------------------------- SPRAWDZENIE WYNIKOW

shapiro.test(Ankieta.M$`Wzrost [cm]`)

# W = 0.98665, p-value = 0.4442

# Poziom istotności, alfa = 0.05

# Statystyka testowa: W = 0.98665

# p-value = 0.4442

# wartosc na podstawie ktorej podejmujemy decyzje, nie prawdziwa ale ze brak

# podstaw zeby ja odrzucic hipoteze zerowa

# p-value to najmniejszy poziom istotności pozwalający odrzucić hipotezę H0,

# wyznaczany na podstawie statystyki testowej

# ---------------------------------------------------------------------- DECYZJA

# na podstawie probki danych - decydujemy o populajci ogolnej

# Decyzje w pakietach statystycznych:

# jeśli alfa >= p, to odrzucamy hipotezę H0 na korzyść hipotezy alternatywnej,

# jeśli alfa < p, to nie ma podstaw do odrzucenia hipotezy H0

# ---------------------------------------------------------------------- WNIOSEK

# Wniosek - jaka jest decyzja, uwazamy ze material dowody wspiera ho czy odrzucamy i przyjmujemy h1

# twierdzenia padajace nie sa na 100 procent, pamietac o tym trzeba

# Wniosek: alfa = 0.05 < p, więc nie ma podstaw do odrzucenia hipotezy H0,

# tzn. na poziomie istotności 0.05 różnice między rozkładem normalnym

# a rozkładem empirycznym nie są statystycznie istotne

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

# KOLOKWIUM ---> tabelka - wpisac statystyke testowa

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

# c) wyznaczyć przedziały ufności dla średniej wzrostu w populacji generalnej

# (poziomy ufności 0.95 oraz 0.98, użyć t.test) – jak poziom ufności wpływa na

# szerokość przedziału ufności?

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

# 180,5 - pewne dane i przenosimy na cala populacje to prawd.=0, nie jest sie

# w stanie powiedziec przedzial ufnosci

t.test(Ankieta.M$`Wzrost [cm]`, conf.level=0.95)

# Przedział liczbowy (179.3013, 181.7820) z prawdopodobieństwem 0.95

# obejmuje prawdziwą nieznaną średnią wzrostu mężczyzn w pop.gen

t.test(Ankieta.M$Wzrost, conf.level=0.98)

# Przedział liczbowy (179.0633, 182.0200) z prawdopodobieństwem 0.98

# obejmuje prawdziwą nieznaną średnią wzrostu mężczyzn w pop.gen

t.test(Ankieta.M$`Wzrost [cm]`, conf.level=1)

# Im większy poziom ufności tym szerszy przedział ufności

# powiekszyl sie, im wiekszy poziom ufnosci, tym wiekszy przedzial

# d) napisać funkcję przedzial.odchylenie, która zwróci wartości ocena.dolna

# i ocena.gorna zgodne z modelem 1 estymacji odchylenia standardowego

# (funkcja var oblicza wariancję)

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

przedzial.odchylenie = function(x, ufnosc)

{

n = length(x)

alpha = 1 - ufnosc

kwartyl.1 = qchisq(1-alpha/2, n-1)

kwartyl.2 = qchisq(alpha/2, n-1)

licznik = (n-1) \* var(x)

data.frame(

ocena.dolna = sqrt(licznik/kwartyl.1),

ocena.gorna = sqrt(licznik/kwartyl.2)

)

}

# e) wyznaczyć przedział ufności dla odchylenia standardowego wzrostu w populacji

# generalnej (poziom ufności 0.97);

przedzial.odchylenie(Ankieta.M$Wzrost, 0.97)

# Przedział liczbowy (5.287252, 7.256975) na poziomie ufności 0.97

# obejmuje prawdziwe nieznane odchylenie standardowe wzrostu mężczyzn w pop.gen

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

# ------------------> KOLOKWIUM

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

#Zadanie 3 - Na poziomie istotności 0.01 sprawdzić, czy można szacować średnią i odchylenie

# standardowe dla liczby godzin spędzanych przy komputerze w ciągu doby w populacji generalnej

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

# SHAPIRO-WILK -> zawsze ten schemat

# Estymację stosujemy, gdy w zadaniu występuje poziom ufności

# Testy stosujemy, gdy w zadaniu występuje poziom istotności

#H0: rozkład liczby godzin spedzonych przy komputerze w ciagu doby w pop.gen jest normalny

#H1: ~H0 (nie jest normalny)

shapiro.test(Ankieta$L.godzin)

#Statystyka testowa: W = 0.94768

#p-value = 0.0001467

#Wniosek: odrzucamy hipotezę H0 na rzecz hipotezy alternatywnej

# tzn. na poziomie istotności 0.01 różnice między rozkładem normalnym

# a rozkładem empirycznym są statystycznie istotne

# Nie można szacować średniej i odchylenia standardowego

#dodatek

ggplot (Ankieta, aes(L.godzin)) +

geom\_histogram (aes (y = ..density..), fill = "blue", col = "black", binwidth = 3) +

stat\_function (fun = dnorm, args = list (mean (Ankieta$L.godzin), sd (Ankieta$L.godzin)), col = "red", size = 1.25)+

ylab ("Częstość")

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

LABORATORIA 7

# ESTYMACJA I TESTY DLA JEDNEJ POPULACJI #

# w każdym teście sformułować hipotezy, podać statystykę, poziom p oraz wniosek

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

# Zadanie 1 - Wiedząc, że zmienna Wzrost w grupie mężczyzn ma rozkład normalny,

# na poziomie istotności 0.05 zweryfikować hipotezę, że średni wzrost mężczyzn

# w populacji generalnej jest większy niż 179 cm (użyć t.test).

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

# wzrost

# uklad normalny (wiemy ze ma wiec lecimy dalej, jak nie to musimy sprawdzic)

# nie rozklad normalny to wtedy weryfikujemy wtedy to!!!!

# Prawdziwą nieznaną średnią w populacji generalnej oznaczamy literą m

# ----------------------------------------------------------- FORMULOWANIE TESTU

# (2) TEST DLA SREDNIEJ (t.test)

# H0: m = 179 (zawsze równość !!!!) # sredni wzrost mezczyzn rowny 179

# (zawsze rownosc tu musi byc)

# H1: m > 179 (jest wiekszy niz 179 cm)

# ---------------------------------------------------------- SPRAWDZENIE WYNIKOW

shapiro.test(Zaliczenie$Średnica)

t.test(Zaliczenie$Średnica, mu = 77.4, alternative ="less")

# Statystyka testowa: t = 2.4676

# p-value = 0.007696

# poziom istotności = 0.05

# ---------------------------------------------------------------------- DECYZJA

# Decyzje w pakietach statystycznych:

# jeśli alfa >= p, to odrzucamy hipotezę H0 na korzyść hipotezy alternatywnej,

# jeśli alfa < p, to nie ma podstaw do odrzucenia hipotezy H0

# ---------------------------------------------------------------------- WNIOSEK

# Wniosek: alfa = 0.05 >= p, więc odrzucamy hipotezę H0 na korzyść hipotezy

# alternatywnej,tzn. na poziomie istotności 0.05 prawdziwa średnia wzrostu

# mężczyzn w pop.gen jest istotnie większa od 179 cm

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

# Zadanie 2 - Dla zmiennej Waga w grupie mężczyzn:

# a) na poziomie istotności 0.01 sprawdzić założenie niezbędne do wykonania

# punktów b-d

# b) wyznaczyć przedział ufności dla średniej wagi w populacji generalnej

# (poziom ufności 0.96)

# c) wyznaczyć przedział ufności dla odchylenia standardowego wagi w populacji

# generalnej (poziom ufności 0.99)

# d) na poziomie istotności 0.05 zweryfikować hipotezę, że średnia waga

# w populacji generalnej wynosi 77 kg

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

# a) na poziomie istotności 0.01 sprawdzić założenie niezbędne do wykonania

# punktów b-d

# TEST ZGODNOSCI SHAPIRO (czy rozklad normalny)

# ----------------------------------------------------------- FORMULOWANIE TESTU

# H0 -> rozkład wagi mężczyzn w pop.gen jest normalny

# H1 -> rozkład wagi mężczyzn w pop.gen nie jest normalny

# ---------------------------------------------------------- SPRAWDZENIE WYNIKOW

shapiro.test(Ankieta.M$`Waga [kg]`)

# W = 0.97607, p-value = 0.07602

# alpha = 0.01

# ---------------------------------------------------------------------- DECYZJA

# jeśli alfa >= p, to odrzucamy hipotezę H0 na korzyść hipotezy alternatywnej,

# jeśli alfa < p, to nie ma podstaw do odrzucenia hipotezy

# alpha = 0.01 vs. p-value = 0.07602

# ---------------------------------------------------------------------- WNIOSEK

# alpha < p-value --> nie ma podstaw do odrzucenia hipotezy zero!!!!

# Można wyznaczać przedziały ufności i weryfikować hipotezy parametryczne

# dla wagi mężczyzn w pop.gen

# zalozenie spelnione

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

ggplot (Ankieta.M, aes(`Waga [kg]`)) +

geom\_histogram (aes (y = ..density..), fill = "blue", col = "black", binwidth = 10) +

stat\_function (fun = dnorm, args = list (mean (Ankieta.M$Waga), sd (Ankieta.M$Waga)), col = "red", size = 1.25)+

ylab ("Częstość")

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

# b) wyznaczyć przedział ufności dla średniej wagi w populacji generalnej

# (poziom ufności 0.96)

# PRZEDZIALY UFNOSCI (t.test)

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

t.test(Ankieta.M$`Waga [kg]`, conf.level=0.96)

# Przedział liczbowy (71.39718 76.62991) z prawdopodobieństwem 0.96

# obejmuje prawdziwą nieznaną średnią wagę mężczyzn w pop.gen

# c) wyznaczyć przedział ufności dla odchylenia standardowego wagi w populacji

# generalnej (poziom ufności 0.99)

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

przedzial.odchylenie(Ankieta.M$`Waga [kg]`, 0.99)

#ocena.dolna ocena.gorna

# 10.35606 15.08749

# Przedział liczbowy (10.35606, 15.08749) na poziomie ufności 0.99

# obejmuje prawdziwe nieznane odchylenie standardowe wagi mężczyzn w pop.gen

# d) na poziomie istotności 0.05 zweryfikować hipotezę, że średnia waga

# w populacji generalnej wynosi 77 kg

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

# Test dla średniej

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

# H0: m=77

# H1: m ≠ 77 lub H1: m != 77 lub H1: m =/= 77 lub H1: ~H0

t.test(Zaliczenie$Średnica, mu=77, alternative="two.sided")

t.test(Ankieta.M$Waga, mu=77) # alternatywa

# Statystyka testowa: t = -2.3769

# p-value = 0.01946

# Wniosek: alfa = 0.05 >= p, więc odrzucamy hipotezę H0 na korzyść hipotezy alternatywnej,

# tzn. na poziomie istotności 0.05 prawdziwa średnia waga mężczyzn w pop.gen

# różni się istotnie od 77 kg

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

# Zadanie 3 - Wylosowaną grupę 10 osób chorujących na nadciśnienie tętnicze

# poddano dwukrotnemu pomiarowi ciśnienia krwi przed podaniem i po podaniu

# pewnego leku, testowanego pod kątem skuteczności obniżania ciśnienia.

# Wartości ciśnienia skurczowego zawiera tabela

# Przed podaniem leku: 158 160 155 170 166 173 167 180 168 173

# Po podaniu leku 140 155 150 167 170 162 157 163 158 175

# Na poziomie istotności 0.05 sprawdzić, czy leczenie jest skuteczne

# (utworzyć pomocnicze zmienne Przed i Po, sprawdzić niezbędne założenie).

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

# 1) Próby zależne

# roznica -> proba zalezna - 10 zbadanych przed i 10 zbadanych po

przed=c(58, 160, 155, 170, 166, 173, 167, 180, 168, 173)

po = c(140, 155, 150, 167, 170, 162, 157, 163, 158, 175)

# 2) TEST ZGODNOSCI SHAPIRO (wersja skrócona)

shapiro.test(przed-po)

#W = 0.57851, p-value = 3.07e-05

# alpha = 0.05

# Rozkład różnic ciśnienia nie różni się istotnie od rozkładu normalnego

# 3) Test dla dwóch średnich w próbach zależnych

# H0: m.Przed-m.Po = 0

# H1: m.Przed-m.Po > 0 (jest wieksze od zera - skuteczne)

t.test(przed-po, alternative = "greater")

# Statystyka testowa: t = 3.1607

# p-value = 0.005769

# alpha = 0.05

# Wniosek: alfa = 0.05 >= p, więc odrzucamy hipotezę H0 na korzyść hipotezy

# alternatywnej, tzn. na poziomie istotności 0.05 w pop.gen średnia ciśnienia

# tętniczego przed podaniem leku jest istotnie większa niż po podaniu leku

# (leczenie jest skuteczne)

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

# Zadanie 4 - Używając zmiennej M.zamieszkania oraz funkcji table i prop.test

# a) wyznaczyć przedział ufności dla odsetka studentów mieszkających na stancji

# w populacji generalnej (poziom ufności 0.96);

# b) na poziomie istotności 0.05 zweryfikować hipotezę, że w akademiku mieszka

# 40 % studentów.

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

# a) wyznaczyć przedział ufności dla odsetka studentów mieszkających na stancji

# w populacji generalnej (poziom ufności 0.96); (wszytskich = 120)

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

# wskaznik struktury

table(Ankieta$M.zamieszkania)

# Akademik Mieszkanie z rodziną Stancja lub inne

# 43 56 21

prop.test(21, 120, conf.level = 0.96)

# Przedział liczbowy (0.1117629, 0.2615617) z prawdopodobieństwem 0.96

# obejmuje prawdziwy nieznany odsetek liczby osób w pop.gen. mieszkający na stancji

# b) na poziomie istotności 0.05 zweryfikować hipotezę, że w akademiku mieszka

# 40 % studentów.

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

# Prawdziwy nieznany odsetek w pop.gen oznaczamy literą p

# Test dla wskaźnika struktury

# H0: p = 0.4

# H1: ~H0

prop.test(43, 120, p = 0.4)

# wartosc statystyki tekstowej (hi-kwadrat)

# Statystyka testowa: chi^2 / X-squared = 0.70312

# p-value = 0.4017

# Wniosek: alfa = 0.05 < p, więc nie ma podstaw do odrzucenia hipotezy H0,

# tzn. na poziomie istotności 0.05 w pop.gen odsetek studentów mieszkających

# w akademiku nie różni się istotnie od 0.4

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

# Zadanie 5 - W populacji generalnej

# a) wyznaczyć przedział ufności dla odsetka absolwentów Liceum Ogólnokształcącego

# (RM) (poziom ufności 0.97);

# b) na poziomie istotności 0.01 zweryfikować hipotezę, że Technikum Informatyczne

@ ukończyło mniej niż 35% studentów.

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

# a)

table(Ankieta$Sz.średnia)

# LO RM --> 51

prop.test(51, 120, conf.level = 0.97)

# Przedział liczbowy 0.3278395 0.5280612

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

# b)

# Technikum Informatyczne --> 39

# p=0.35 (ukonczylo rowno 35 %)

# p>0.35 (ukonczylo mniej niz 35 procent)

prop.test(39, 120, p=0.35, alternative = "less")

# X-squared = 0.22894

# p-value = 0.3162

# alfa = 0.01

# Wniosek: alfa = 0.05 < p, więc nie ma podstaw do odrzucenia hipotezy H0,

# tzn. na poziomie istotności 0.01 w pop.gen odsetek studentów ktorzy ukonczyki

# technikum informatyczne nie różni się istotnie od 0.4

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

LABORATORIA 8

# TESTY DLA DWÓCH POPULACJI #

# w każdym teście sformułować hipotezy, podać statystykę, poziom p oraz wniosek

# rozklad normalny dla 1, 2 populacji

# Poprawne porównanie średnich w dwóch populacjach wymaga sprawdzenia założeń:

# zgodności z rozkładem normalnym w obu populacjach oraz jednorodności wariancji

# (założenia obejmują trzy testy istotności)

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

# TESTY:

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

# 1)

# TEST ZGODNOSCI SHAPIRO WILKA

# czy jest rozkład normalny

# by (zmienna mierzalna, zmienna grupująca, shapiro.test)

# by(Ankieta$`Waga [kg]`, Ankieta$Płeć, shapiro.test)

# lub

# shapiro.test(Ankieta.K$`Waga [kg]`)

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

# 2)

# PARAMETRYCZNE

# test (data = zbiór danych, zmienna mierzalna~grupująca)

# lub

# test (zmienna mierzalna~grupująca, zbiór danych)

# t.test(Ankieta$Waga~Ankieta$Płeć, alternative="less", var.equal=TRUE)

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

# 3)

# TEST FISHER

# jednorodnosc wariancji

# var.test(Ankieta$`Waga [kg]`~Ankieta$Płeć)

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

# Zadanie 1 - Używając zmiennej Waga porównać grupę kobiet i mężczyzn w

# populacji generalnej:

# a) na poziomie istotności 0.01 sprawdzić założenie o normalności rozkładu

# w obu grupach testem Shapiro-Wilka;

# b) na poziomie istotności 0.05 sprawdzić jednorodność wariancji w obu grupach

# testem Fishera (użyć var.test);

# c) na poziomie istotności 0.05 zweryfikować testem Studenta hipotezę, że

# średnia wagi w populacji generalnej

# jest niższa w grupie kobiet (w przypadku jednorodnych wariancji wyłączyć

# poprawkę Welcha: var.equal = TRUE).

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

# a) na poziomie istotności 0.01 sprawdzić założenie o normalności rozkładu

# w obu grupach testem Shapiro-Wilka;

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

# TEST ZGODNOSCI SHAPIRO-WILKA

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

# sprawdzamy czy rozklad jest NORMALNY

# zgodności (rozkład normalny) – by (zmienna mierzalna, zmienna grupująca,

# shapiro.test)

# zmienna mierzalna - waga

# grupujaca - plec

#H0: waga w grupie kobiet w pop. generalnej ma rozklad normalny

#H1: ~H0

# 1)

by(Ankieta$`Waga [kg]`, Ankieta$Płeć, shapiro.test)

# poziom istotnosci = 0,01 = alfa

# alfa < p-value w obu przypadkach tak musi byc

#Ankieta$Płeć: K --> W = 0.95917, p-value = 0.4221

#Ankieta$Płeć: M --> W = 0.97607, p-value = 0.07602

# ukladny normalne

# 2)

shapiro.test(Ankieta.K$`Waga [kg]`)

#Statystyka testowa: W=0.95917

#p-value = 0.4221

# Wniosek: alfa=0.01<p.value więc nie ma podstaw do odrzucenia hipotezy H0

# tzn. na poziomie istotności 0.01 różnice między rozkładem normalnym

# a rozkładem empirycznym nie są statystycznie istotne

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

# b) na poziomie istotności 0.05 sprawdzić jednorodność wariancji w obu grupach

# testem Fishera (użyć var.test);

# Prawdziwe odchylenie standardowe w pop.gen oznaczamy σ

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

# TEST FISHERA (czy jednorodna wariancja)

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

# poziom istotnosci = 0,05

# H0: σ^2.K = σ^2.M (wariancja w grupie kobiet = wariancji w grupie mezczyzn)

# H1: ~H0 (nie są sobie rowne (negacja))

var.test(Ankieta$`Waga [kg]`~Ankieta$Płeć)

# Statystyka testowa: F = 0.57434

# p-value = 0.1294

# Wniosek: alfa = 0.05 < p, więc nie ma podstaw do odrzucenia hipotezy H0,

# tzn. na poziomie istotności 0.05 w pop.gen wariancje wag dla obu płci

# nie różnią się istotnie (wariancje są jednorodne)

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

# c) na poziomie istotności 0.05 zweryfikować testem Studenta hipotezę,

# że średnia wagi w populacji generalnej jest niższa w grupie kobiet

# (w przypadku jednorodnych wariancji wyłączyć poprawkę Welcha: var.equal = TRUE).

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

# Test dla dwóch średnich

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

# poziom istotnosci = 0,05

# I) ROZKLAD NORMALNY: waga zgodna z rozkladem normalnym w grupie M i k -> OK!

# gdy rozklad normalny tylko w jednej grupie jest - wtedy test parametryczny)

# II) TEST FISHERA: jednorodnosc wariancji -> TRUE albo FALSE (TRUE)

# H0: m.K = m.M (waga kobiet jest rowna grupie mezczyzn)

# H1: m.K < m.M (srednia w grupie kobiet) < (nizsza m.M (srednia w grupie mezczyzn)

t.test(Ankieta$Waga~Ankieta$Płeć, alternative="less", var.equal=TRUE)

# Statystyka testowa: t = -5.5803

# p-value = 7.77e-08

# Wniosek: alfa = 0.05 >= p, więc odrzucamy hipotezę H0 na korzyść hipotezy

# alternatywnej, tzn. na poziomie istotności 0.05 w pop.gen średnia wagi

# w grupie kobiet jest istotnie mniejsza niż średnia wagi w grupie mężczyzn

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

# Zadanie 2 - Na poziomie istotności 0.05 zweryfikować hipotezę, że średnia

# ocen z kursów w populacji generalnej zależy od płci (na tym samym poziomie

# istotności sprawdzić wszystkie niezbędne założenia).

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

# 1) Testy zgodności Shapiro-Wilka

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

# poziom istotnosci = 0,05

# poziom istotnosci = 0,05 = alpha

#H0: średnia ocen w grupie kobiet/mezczyzn w pop. generalnej ma rozklad normalny

#H1: ~H0

by(Zaliczenie$Średnica, Zaliczenie$Kolor, shapiro.test)

# Ankieta$Płeć: K --> Statystyka testowa: W=0.95258 , p-value = 0.3079

# Ankieta$Płeć: M --> Statystyka testowa: W=0.98637, p-value = 0.4261

# 0,05 < p-value --> JEST rozklad normalny, nie ma podstaw

# Wniosek (w obu): alfa=0.05<p.value więc nie ma podstaw do odrzucenia hipotezy H0

# tzn. na poziomie istotności 0.05 różnice między rozkładem normalnym

# a rozkładem empirycznym nie są statystycznie istotne

# założenie można uznać za spełnione

# 2) Test jednorodnosci wariancji Fishera

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

# H0: σ^2.K = σ^2.M (srednia K i srednia M sa rowne)

# H1: ~H0

var.test(Zaliczenie$Średnia~Zaliczenie$Kolor)

# Statystyka testowa: F = 1.3253

# p-value = p-value = 0.3454

# Wniosek: alfa = 0.05 < p, więc nie ma podstaw do odrzucenia hipotezy H0,

# tzn. na poziomie istotności 0.05 w pop.gen wariancje sredniej dla obu płci

# nie różnią się istotnie (wariancje są jednorodne)

# Założenia są spełnione, można zrobić test t bez poprawki Welcha (czyli TRUE)

#3) Test dla dwóch średnich

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

# H0: m.K = m.M (srednia K i srednia M sa rowne)

# H1: ~H0

t.test(Zaliczenie$Średnica~Zaliczenie$Kolor, var.equal=TRUE)

# Statystyka testowa: t = 0.37918

# p-value = 0.7052

# Wniosek: alfa = 0.05 < p, więc nie ma podstaw do odrzucenia hipotezy H0,

# tzn. na poziomie istotności 0.05 w pop.gen średnia sredniej ocen w grupie kobiet

# nie rozni sie od tej w grupie mężczyzn

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

# Zadanie 3 - Na poziomie istotności 0.03 zweryfikować hipotezę, że mężczyźni

# preferujący system inny niż Windows spędzają przed komputerem więcej godzin

# niż pozostali (na tym samym poziomie istotności sprawdzić wszystkie niezbędne

# założenia).

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

# 1) TEST ZGODNOSCI SHAPIRO-WILKA (CZY ROZKLAD NORMALNY)

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

# poziom istotnosci/alfa = 0,03

# Inny

# H0: Rozkład liczby godzin w grupie mężczyzn korzystających z systemu

# Inny w pop. generalnej ma rozklad normalny

# H1: ~H0

by(Ankieta.M$L.godzin, Ankieta.M$System, shapiro.test)

# Statystyka testowa dla Inny: W=0.9012

# p-value = 0.1644

# Wniosek: alfa=0.03<p.value więc nie ma podstaw do odrzucenia hipotezy H0

# tzn. na poziomie istotności 0.03 różnice między rozkładem normalnym

# a rozkładem empirycznym nie są statystycznie istotne

# założenie można uznać za spełnione (rozklad normalny)

# Windows

#Windows

#H0: Rozkład liczby godzin w grupie mężczyzn korzystających z systemu Windows

# w pop. generalnej ma rozklad normalny

#H1: ~H0

#Statystyka testowa: W=0.968

#p-value = 0.03462

# Wniosek: alfa=0.03<p.value więc nie ma podstaw do odrzucenia hipotezy H0

# tzn. na poziomie istotności 0.03 różnice między rozkładem normalnym

# a rozkładem empirycznym nie są statystycznie istotne

# założenie można uznać za spełnione

# 2) TEST JEDNORODNOSCI WARIANCJI FISHERA

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

# H0: σ^2.I = σ^2.W (rowne)

# H1: ~H0 (nie rowne)

var.test(Ankieta.M$L.godzin~Ankieta.M$System)

# Statystyka testowa: F = 2.355

# p-value = 0.02785

# Wniosek: alfa = 0.03 > p, więc odrzucamy hipotezę H0 na rzecz hipotezy

# alternatywnej, tzn. na poziomie istotności 0.03 w pop.gen wariancje

# liczby godzin dla obu grup różnią się istotnie (wariancje nie są jednorodne)

# Założenia zgodności z rozkładem normalnym są spełnione, można zrobić test

# t z poprawką Welcha (czyli FALSE)

#3) Test dla dwóch średnich

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

# H0: m.I = m.W

# H1: m.I > m.W

t.test(Ankieta.M$L.godzin~Ankieta.M$System, alternative="greater", var.equal=FALSE)

# Statystyka testowa: t = 1.5506

# p-value = 0.0731

# Wniosek: alfa = 0.03 < p, więc nie ma podstaw do odrzucenia hipotezy H0

# tzn. na poziomie istotności 0.03 w pop.gen liczba godzin spędzanych przy

# komputerze wśród mężczyzn preferujących inny system operacyjny niż Windows

# nie różni się istotnie od liczby godzin przy komputerze w grupie mężczyzn

# preferujących Windows

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

# Zadanie 4 (bez rozkladu normalnego bedzie)

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

# Zadanie 5

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