

Zestaw 3 (przedziały ufności)

rozkłady prawdopodobieństwa:

binom - dwumianowy
geom - geometryczny
pois - poissona
norm - normalny
t - t studenta
chisq - chi-kwadrat
f - Fshedecora
exp - wykładniczy
unif - jednostajny

przedrostki przed nazwą rozkładu
d - wartość $f(x)$ lub $P(X=x)$
p - wartość $F(x)$
q - wartość kwantyla
r - generowanie liczb losowych

Zadanie 1

Średnia cena 50 losowo wybranych podręczników akademickich wyniosła 28.40 zł. Wiadomo, że odchylenie standardowe cen podręczników wynosi 4.75 zł. Wyznaczyć 95% przedział ufności dla średniej ceny podręcznika akademickiego zakładając, że rozkład cen jest rozkładem normalnym.

$n=50$
 $\alpha=0.05$
 $\text{sig}=4.75$
 $sr = 28.4$

Model 1

X_1, X_2, \dots, X_n i.i.d. $N(\mu, \sigma)$, σ -znane:

$$\mu \in \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right),$$

gdzie $z_{1-\frac{\alpha}{2}}$ oznacza kwantyl rozkładu normalnego standardowego rzędu $1 - \frac{\alpha}{2}$.

$(prz=sr+c(-1,1)*\text{sig}*qnorm(1-\alpha/2)/\text{sqrt}(n))$

Zadanie 2

Przeprowadzono 18 niezależnych pomiarów temperatury topnienia ołowiu i otrzymano następujące wyniki (w C):

330.0 322.0 345.0 328.6 331.0 342.0
342.4 340.4 329.7 334.0 326.5 325.8
337.5 327.3 322.6 341.0 340.0 333.0

Zakładamy, że temperatura topnienia ołowiu ma rozkład normalny. Wyznaczyć dwustronny przedział ufności dla wartości oczekiwanej i odchylenia standardowego temperatury topnienia ołowiu na poziomie ufności 0.95.

$x = c(330, 322, 345, 328.6, 331, 342, 342.4, 340.4, 329.7, 334, 326.5, 325.8, 337.5, 327.3, 322.6, 341, 340, 333)$
 $\alpha = 0.05$

Wartość Oczekiwana:

Model 2

X_1, X_2, \dots, X_n i.i.d. $N(\mu, \sigma)$, σ -nieznane:

$$\mu \in \left(\bar{X} - t_{1-\frac{\alpha}{2}}^{[n-1]} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}}^{[n-1]} \frac{S}{\sqrt{n}} \right),$$

gdzie $t_{1-\frac{\alpha}{2}}^{[n-1]}$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu t-Studenta o $n - 1$ stopniach swobody.

$\text{mean}(x) + c(-1,1) * qt(1-\alpha/2, \text{length}(x)-1) * \text{sd}(x) / \text{sqrt}(\text{length}(x))$

lub

```

prz.ufn.mi=function(x,alfa){
n=length(x)
sr=mean(x)
s=sd(x)
q=qt(1-alfa/2,n-1)
prz=sr+c(-1,1)*s*q/sqrt(n)
list(przedzial=prz)
}
prz.ufn.mi(x,alfa)

```

lub

`t.test(x, conf.level = (1 - alfa))` // `x` – wektor danych; `conf.level` – poziom ufności

Odchylenie Standardowe

gdzie $\chi^2_{1-\frac{\alpha}{2}, n-1}$ i $\chi^2_{\frac{\alpha}{2}, n-1}$ są kwantylami rzędu, odpowiednio, $1 - \frac{\alpha}{2}$ i $\frac{\alpha}{2}$ rozkładu chi-kwadrat o $n - 1$ stopniach swobody;

$$\sigma \in \left(\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}} \right). \quad (7)$$

lewy kraniec:

`sqrt((length(x)-1)*var(x)/qchisq(1-alpha/2, length(x)-1))`

prawy kraniec:

`sqrt((length(x)-1)*var(x)/qchisq(alpha/2, length(x)-1))`

lub

```

prz.ufn.odch=function(x,alfa){
n=length(x)
s2=var(x)

```

`lk= (n-1)*s2/ qchisq(1-alfa/2,n-1)`

`pk= (n-1)*s2/ qchisq(alfa/2,n-1)`

`prz=c(sqrt(lk),sqrt(pk))`

`list(przedzial=prz)`

`}`

`prz.ufn.odch(x,0.05)`

Zadanie 3

Pojemność 10 losowo wybranych baterii wyniosła (w ampero-godzinach):

140, 136, 150, 144, 148, 152, 138, 141, 143, 151.

Zakładając, że pojemność baterii ma rozkład normalny, wyznaczyć 99% przedział ufności dla wariancji pojemności.

Model 2

X_1, X_2, \dots, X_n i.i.d. $N(\mu, \sigma)$, μ -nieznane:

$$\sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \right), \quad (6)$$

gdzie $\chi^2_{1-\frac{\alpha}{2}, n-1}$ i $\chi^2_{\frac{\alpha}{2}, n-1}$ są kwantylami rzędu, odpowiednio, $1 - \frac{\alpha}{2}$ i $\frac{\alpha}{2}$ rozkładu chi-kwadrat o $n - 1$ stopniach swobody;

```

prz.ufn.var=function(x,alfa){

```

```

n=length(x)
s2=var(x)

lk= (n-1)*s2/ qchisq(1-alfa/2,n-1)
pk= (n-1)*s2/ qchisq(alfa/2,n-1)

prz=c(lk,pk)
list(przedzial=prz)
}

y=c(140,136,150, 144, 148, 152, 138, 141, 143, 151)
prz.ufn.var(y, 0.01)

```

Zadanie 4

W sondażu przeprowadzonym przez magazyn „Time” („Time”, 22 czerwca 1987) 578 spośród 1014 dorosłych respondentów stwierdziło, że dla dobra dzieci lepiej jest, gdy matka nie pracuje poza domem. Wyznaczyć 95% przedział ufności dla odsetka dorosłych podzielających ten pogląd.

X_1, X_2, \dots, X_n i.i.d. $Bern(p)$, n – duże:

$$p \in \left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right).$$

```

prz.ufn.p=function(k,n,alfa){
pp=k/n
prz=pp+c(-1,1)*qnorm(1-alfa/2)*sqrt(pp*(1-pp)/n)
list(przedzial=prz)
}

```

```
prz.ufn.p(578,1014,0.05)
```

Zadanie 5

Jak dużą próbę należy pobrać, aby z maksymalnym błędem 2.5% oszacować, na poziomie ufności 0.95, odsetek dorosłych Polaków czytających codziennie przynajmniej jedną gazetę? Uwzględnić rezultaty wstępnych badań, z których wynika, że interesująca nas wielkość jest rzędu 20%. Porównać otrzymaną liczbę próby z liczbą, jaka byłaby wymagana, gdyby pominąć rezultaty badań wstępnych.

Model III. Cecha X ma rozkład dwupunktowy $P(X = 1) = p$, p - nieznany
 jeżeli znany jest szacunkowy procent p_0 , to $n \geq u_{1-\alpha/2}^2 \frac{p_0(1-p_0)}{d^2}$
 jeżeli nie jest znany szacunkowy procent p_0 , to $n \geq u_{1-\alpha/2}^2 \frac{1}{4d^2}$

```

licz.pr=function(d,alfa,pp){
n=ceiling((qnorm(1-alfa/2)/d)^2*pp*(1-pp))
list(n=n)
}

```

//p0 (pp) nieznane – pp przyjmujemy na 0.5

```

licz.pr(0.025,0.05,0.5) // pp nie znane
licz.pr(0.025,0.05,0.2) // pp znane

```

Zadanie 6

Wygenerować 10000 próbek 10-elementowych z rozkładu normalnego. Następnie zakładając, iż o próbkach wiemy tylko tyle, że pochodzą one z rozkładu normalnego o nieznanym parametrach, wyznaczyć dla każdej próbki przedział ufności dla wartości oczekiwanej na poziomie ufności 0.95. Porównać frakcję pokryć przez przedział ufności faktycznej wartości oczekiwanej z założonym poziomem ufności.

```
n=10
mi=4
sig=3
alfa=0.05
q=qt(1-alfa/2,n-1)
i=0
for(k in 1:10000){
  x= rnorm(n,mi,sig)
  m=mean(x)
  s=sd(x)
  if(m-s*q/sqrt(n)<mi & m+s*q/sqrt(n)>mi)
    i=i+1
}
i/10000
```

Zestaw 4 (Weryfikacja Hipotez)

Zadanie 1

Wytrzymałość na ciśnienie wewnętrzne szkła butelek jest ich ważną charakterystyką jakościową. Pewna rozlewnia zainteresowana jest butelkami, których średnia wytrzymałość przewyższa $1,20 \text{ N/mm}^2$. Na podstawie dotychczasowych doświadczeń wiadomo, że rozkład ciśnienia jest normalny z odchyleniem standardowym $0,07 \text{ N/mm}^2$. Pobrano próbę losową 20 butelek, które następnie umieszczono w maszynie hydrostatycznej, zwiększając ciśnienie aż do zniszczenia butelki. Otrzymano następujące wyniki (w N/mm^2):

1.36, 1.14, 1.27, 1.15, 1.20, 1.29, 1.27, 1.18, 1.23, 1.36,
1.38, 1.37, 1.30, 1.21, 1.33, 1.28, 1.32, 1.29, 1.33, 1.25.

Na poziomie istotności 0.04 stwierdzić, czy dana partia butelek spełnia postawione wymagania jakościowe.

$z=c(1.36, 1.14, 1.27, 1.15, 1.20, 1.29, 1.27, 1.18, 1.23, 1.36, 1.38, 1.37, 1.30, 1.21, 1.33, 1.28, 1.32, 1.29, 1.33, 1.25)$

$H: \mu = 1.2$
 $K'': \mu > 1.2$

$$H : \mu = \mu_0, \quad (1)$$

$$K : \mu \neq \mu_0 \quad (2)$$

$$K' : \mu < \mu_0$$

$$K'' : \mu > \mu_0.$$

Model 1

X_1, X_2, \dots, X_n i.i.d. $N(\mu, \sigma)$, σ -znane

Statystyka testowa:

$$T = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}. \quad (3)$$

Obszar krytyczny:

$$\begin{aligned} W_\alpha &= (-\infty, -z_{1-\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, +\infty), \\ W'_\alpha &= (-\infty, -z_{1-\alpha}], \\ W''_\alpha &= [z_{1-\alpha}, +\infty), \end{aligned} \quad (4)$$

gdzie $z_{1-\frac{\alpha}{2}}$ i $z_{1-\alpha}$ są, odpowiednio, kwantylami rozkładu normalnego $N(0, 1)$ rzędów $1 - \frac{\alpha}{2}$ i $1 - \alpha$.

```
stat.z=function(x,m0,sig){
n=length(x)
sr=mean(x)
z=(sr-m0)*sqrt(n)/sig
list(z=z)
}
```

```
stat.z(z,1.2,0.07)
qnorm(1-0.04) /// obszar krytyczny
```

lub

```
(pv=1-pnorm(stat.z(z,1.2,0.07)$z))
```

jeżeli $pv < \alpha$ - odrzucamy hipotezę H
jeżeli $pv > \alpha$ - nie odrzucamy hipotezy H

$pval$ jest mniejsze od α - odrzucamy H

Zadanie 2

Nominalna waga netto kawy sprzedawanej w opakowaniu szklanym winna wynosić 150 g. Występuje jednakże duża zmienność wagi. Istotnie, próba losowa siedmiu słoiczków kawy konkretnej marki wykazała następujące wagi netto (w gramach):

142, 151, 148, 151, 145, 150, 141.

a) Zakładając normalność rozkładu wagi, przetestować hipotezę głoszącą, że waga netto tej marki kawy wynosi faktycznie 150 g. Przyjąć poziom istotności $\alpha = 0.05$.

$$H: \mu = 150$$

$$K: \mu \neq 150$$

Model 2

X_1, X_2, \dots, X_n i.i.d. $N(\mu, \sigma)$, σ -nieznane

Statystyka testowa:

$$T = \frac{\bar{X} - \mu_0}{S} \sqrt{n}, \quad (5)$$

Obszar krytyczny:

$$W_\alpha = (-\infty, -t_{1-\frac{\alpha}{2}}^{[n-1]}) \cup [t_{1-\frac{\alpha}{2}}^{[n-1]}, +\infty), \quad (6)$$

$$W'_\alpha = (-\infty, -t_{1-\alpha}^{[n-1]}],$$

$$W''_\alpha = [t_{1-\alpha}^{[n-1]}, +\infty),$$

gdzie $t_{1-\frac{\alpha}{2}}^{[n-1]}$ i $t_{1-\alpha}^{[n-1]}$ są, odpowiednio, kwantylami rozkładu t-Studenta o $n-1$ stopniach swobody rzędów $1-\frac{\alpha}{2}$ i $1-\alpha$.

$x=c(142, 151, 148, 151, 145, 150, 141)$

```
stat.t=function(x,m0){  
  n=length(x)  
  sr=mean(x)  
  S=sd(x)  
  t=(sr-m0)*sqrt(n)/S  
  list(t=t)  
}
```

```
n = length(x)  
2*pt(stat.t(x,150)$t,n-1) ## p-value
```

lub

```
t.test(x,mu=150,conf.level=0.95,alt="two.sided") ## alt="two.sided - alternatywa różny od(=/=)
```

$\alpha < pval$ - nie odrzucamy H

b) Sprawdzić, czy odchylenie standardowe wagi netto słoika kawy różni się istotnie od 4g.

$$H: \sigma = 4$$

$$K: \sigma \neq 4$$

Model 2

X_1, X_2, \dots, X_n i.i.d. $N(\mu, \sigma)$, μ -nieznane

Statystyka testowa:

$$T = \frac{(n-1)S^2}{\sigma_0^2},$$

Obszary krytyczny

$$W_\alpha = (0, \chi_{\frac{\alpha}{2}, n-1}^2] \cup [\chi_{1-\frac{\alpha}{2}, n-1}^2, +\infty)$$

```
stat.chi=function(x,sig0){
  n=length(x)
  S2=var(x)
  chi=(n-1)*S2/sig0^2
  list(chi=chi)
}
```

```
stat.chi(x,4)
```

```
[1] 6.678571
```

```
qchisq(0.05/2,n-1)
qchisq(1-0.05/2,n-1)
```

Wartość statystyki nie wpada w wyznaczony obszar, więc pozostajemy przy H_0

Zadanie 3

Wylosowana niezależnie z partii żarówek 12 elementowa próba dała następujące wyniki pomiarów czasu świecenia (w godzinach):

2852, 3060, 2631, 2819, 2805, 2835, 2955, 2595, 2690, 2723, 2815, 2914.

a) Zakładając normalność rozkładu czasu świecenia żarówek wyznaczyć 97% przedział ufności dla średniego czasu świecenia żarówek oraz dla odchylenia standardowego czasu świecenia żarówek.

```
x=c(2852, 3060, 2631, 2819, 2805, 2835, 2955, 2595, 2690, 2723, 2815, 2914)
```

```
t.test(x, conf.level = 0.97)
```

odchylenie

```
prz.ufn.odch=function(x,alfa){
  n=length(x)
  s2=var(x)
```

```
lk= (n-1)*s2/ qchisq(1-alfa/2,n-1)
pk= (n-1)*s2/ qchisq(alfa/2,n-1)
```

```
prz=c(sqrt(lk),sqrt(pk))
list(przedzial=prz)
}
```

```
prz.ufn.odch(x,0.03)
```

b) Czy średni czas świecenia żarówek jest istotnie krótszy od 2900 godzin? Przyjąć poziom istotności 0.05.

$H_0: \mu = 2900$

$K: \mu < 2900$

```
t.test(x,mu=2900,conf.level=0.95,alt="less")
```

p-value < alfa odrzucamy H_0

Zadanie 4

W czasie poprawnej pracy maszyny frakcja wytwarzanych przez nią elementów wadliwych nie powinna przekraczać 4%. Jeżeli frakcja elementów wadliwych przekroczy 4%, wówczas należy podjąć czynności mające na celu wyregulowanie procesu produkcji. Pracownik zajmujący się kontrolą jakości pobrał próbkę losową 200 elementów i znalazł w niej 14 elementów wadliwych. Czy zaistniała sytuacja wymaga wyregulowania procesu produkcji?

Zweryfikować odpowiednią hipotezę na poziomie istotności 0.05.

H: $p_0 = 0.04$

K: $p_0 > 0.04$

```
prop.test(14,200,p=0.04, alt="greater",correct=F)
```

p-value < alfa - odrzucamy H -> trzeba regulować

Zadanie 5

Przeprowadzono ankietę wśród pracowników naukowych pewnej uczelni dotyczącą stażu pracy. Stwierdzono, że wśród 140 respondentów znalazło się 47 osób o stażu krótszym niż 10 lat, 53 osoby pracujące co najmniej 10, ale nie dłużej niż 15 lat oraz 40 osób o stażu pracy dłuższym niż 15 lat. Zweryfikować hipotezę, że 30% pracowników tej uczelni legitymuje się stażem pracy przekraczającym 15 lat. Przyjąć poziom istotności 0.05.

H: $p_0 = 0.3$

K: $p_0 \neq 0.3$

```
prop.test(40,140,p=0.3, alt="two.sided",correct=F)
```

p-value > alfa – przyjmujemy H

Zadanie 6

Badania dotyczące stażu pracy pracowników naukowych pewnego wydziału dały następujące rezultaty:

Staż pracy (lata) Liczba pracowników

1 – 5	3
5 – 9	5
9 – 13	9
13 – 17	4
17 – 21	1

a) Zakładając, że rozkład stażu pracy jest normalny, zweryfikować hipotezę, że średnia stażu pracy na tym wydziale przekracza 8 lat. Przyjąć poziom istotności 0.05.

H: $\mu_i = 8$

K: $\mu_i > 8$

```
xl = c(1,5,9,13, 17) #lewe granice  
xp= c(5,9, 13,17, 21) #prawe granice
```

```
xio=(xl+xp)/2 # srodki klas
```

```
ni=c(3,5,9,4,1)  
n=sum(ni)
```

```
sr=sum(xio*ni)/n  
s2=sum(ni*(xio-sr)^2/(n-1)) # #var
```

```
stat.t=(sr-8)*sqrt(n)/sqrt(s2)
```

```
qt(1-.05,n-1)
```

wpada, odrzucamy H

b) Wyznaczyć 95% przedział ufności dla średniego stażu pracy na tym wydziale.

```
alfa = 0.05  
sr + c(-1,1) *qt(1-alfa/2, n -1)* sqrt(s2)/sqrt(n)
```


c) Wyznaczyć 95% przedział ufności dla odchylenia standardowego stażu pracy na tym wydziale.

lewy kraniec:

$$\sqrt{(n-1) \cdot s^2 / q_{\text{chisq}}(1-\alpha/2, n-1)}$$

prawy kraniec:

$$\sqrt{(n-1) \cdot s^2 / q_{\text{chisq}}(\alpha/2, n-1)}$$

Zadanie 8

W stopie metalicznym pewnego typu zastosowano dwa różne pierwiastki utwardzające. Wyniki pomiarów twardości przeprowadzonych później na próbkach tego stopu utwardzanych obiema metodami wyglądają następująco:

Metoda I: 145 150 153 148 141 152 146 154 139 148

Metoda II: 152 150 147 155 140 146 158 152 151 143 153

Przyjmuje się, że twardość ma rozkład normalny oraz że odchylenia standardowe σ_1 i σ_2 dla obu metod są równe. Czy na podstawie przeprowadzonych pomiarów można stwierdzić, że średnia twardość stopu utwardzanego drugą metodą przewyższa średnią twardość stopu utwardzanego pierwszą metodą?

$$H: \mu_1 = \mu_2$$

$$K: \mu_1 < \mu_2$$

$$m1 = c(145, 150, 153, 148, 141, 152, 146, 154, 139, 148)$$

$$m2 = c(152, 150, 147, 155, 140, 146, 158, 152, 151, 143, 153)$$

$$t.test(m1, m2, alt="less", var.equal=T)$$

p-val > alfa przyjmujemy H (niema podstaw do odrzucania)

Zadanie 9

Dokonano po 5 niezależnych pomiarów ciśnienia w komorze spalania silnika dla dwóch gatunków paliwa. Otrzymano następujące rezultaty (w kG/cm²):

pierwszy gatunek: 40.32 39.85 41.17 40.62 40.04

drugi gatunek: 51.07 49.60 50.45 50.59 50.29

Zakładamy, że ciśnienie w komorze spalania ma rozkład normalny. Na poziomie istotności 0.05 zweryfikować hipotezę o jednakowej wariancji ciśnienia dla obu gatunków badanego paliwa.

$$H: \sigma_1^2 = \sigma_2^2$$

$$K: \sigma_1^2 \neq \sigma_2^2$$

$$g1 = c(40.32, 39.85, 41.17, 40.62, 40.04)$$

$$g2 = c(51.07, 49.60, 50.45, 50.59, 50.29)$$

$$var.test(g1, g2, alt="two.sided")$$

p-val > alfa przyjmujemy H (niema podstaw do odrzucania)

Zadanie 10

Spośród pracowników pewnego przedsiębiorstwa wylosowano niezależnie 15 pracowników fizycznych i 9 pracowników umysłowych. Otrzymano następujące dane dotyczące stażu pracy (w latach):

pracownicy umysłowi: 14 17 7 33 2 24 26 22 12

pracownicy fizyczni: 13 15 3 2 25 4 1 18 6 9 20 11 5 1 7

Wiadomo, że rozkład stażu pracy w przedsiębiorstwie jest normalny. Zweryfikować hipotezę, że średni staż pracy pracowników fizycznych jest istotnie krótszy niż staż pracy pracowników umysłowych

$$H: \mu(u) = \mu(f)$$

$$K: \mu(u) > \mu(f)$$

sprawdzamy czy wariancje są równe:
H1 wariacja1= wariancja 2 ($\sigma^2 = \sigma^2$)
K1 wariancje różne.

$pu = c(14, 17, 7, 33, 2, 24, 26, 22, 12)$

$pf = c(13, 15, 3, 2, 25, 4, 1, 18, 6, 9, 20, 11, 5, 1, 7)$

$var.test(pu,pf)$

$p\text{-}v > \alpha \rightarrow$ niema podstaw do odrzucenia H1 \rightarrow wariacje rowne

$t.test(pu,pf,alt="greater",var.equal=T)$ ## niema podstaw do odrzucenia H1 wiec var equal

p-value < alfy \rightarrow odrzucamy H na rzecz K
wiec fizycznych jest krótszy

Zadanie 11

Grupę 10 dzieci poddano testowi pamięci. Po pewnym czasie, w którym dzieci wykonywały w domu ćwiczenia usprawniające pamięć, poddano je ponownie testowi. Na podstawie wyników zamieszczonych w tabeli stwierdzić, czy zaproponowane ćwiczenia w istotny sposób usprawniają pamięć. Założyć, że liczba zapamiętywanych przedmiotów ma rozkład normalny.

Dziecko	1	2	3	4	5	6	7	8	9	10
Liczba przedmiotów zapamiętanych przed serią ćwiczeń	27	21	34	24	30	27	33	31	22	27
Liczba przedmiotów zapamiętanych po serii ćwiczeń	29	32	29	27	31	26	35	30	29	28

H $\mu_i(\text{przd}) = \mu_i(\text{po})$

K $\mu_i(\text{przed}) < \mu_i(\text{po})$

$przed = c(27, 21, 34, 24, 30, 27, 33, 31, 22, 27)$

$po = c(29, 32, 29, 27, 31, 26, 35, 30, 29, 28)$

$t.test(przed,po,alt="less",paired=T)$ ##przed i po więc model 3 -, paired (bo pary)

p-val większe od alfa czyli nie odrzucamy, czyli cw nie pomagają

Zadanie 13

455 spośród 700 wylosowanych studentów Politechniki, będących absolwentami techników i 517 spośród 1320 wylosowanych studentów będących absolwentami liceów, nie zdało egzaminu z matematyki na zakończenie pierwszego semestru. Czy na podstawie powyższych wyników można stwierdzić, że absolwenci techników są słabiej przygotowani z matematyki niż absolwenci liceów?

$n = c(700, 1320)$

$k = c(455, 517)$

$prop.test(k,n,alt="greater")$

p val mniejsze od alfa

Zadanie 14

W pewnym przedsiębiorstwie wylosowano niezależnie 100 kobiet i 200 mężczyzn, dla których uzyskano następujące dane dotyczące stażu pracy: Liczba pracowników

Staż pracy	Liczba pracowników	
	Kobiety	Mężczyźni
0 – 5	8	20
5 – 10	14	48
10 – 15	25	80
15 – 20	30	32
20 – 25	23	20

Zweryfikować hipotezę, że odsetek kobiet pracujących w tym przedsiębiorstwie krócej niż 10 lat jest mniejszy niż odsetek mężczyzn o tym samym stażu pracy. Przyjąć poziom istotności 0.05.

H: $p_k = p_m$

K: $p_k < p_m$

$n = c(100, 200)$ # wektor licznosci

$k = c(14+8, 20+48)$

`prop.test(k,n,alt="less")`

Zestaw 5

Zadanie 2

Badania grupy krwi 200 osób dały następujące wyniki: grupę 0 miały 73 osoby, grupę A – 74 osoby, grupę B – 34 osoby, natomiast grupę AB miało 19 osób.

a) Czy na podstawie tych wyników można przyjąć hipotezę o równomiernym rozkładzie wszystkich grup krwi?

$$H: p_1 = p_2 = p_3 = p_4 = 1/4$$

$$K: \sim H$$

```
krew = c(73,74,34,19)
```

```
chisq.test(krew)
```

małe p-value więc odrzucamy

b) Zweryfikować hipotezę, że grupa krwi O występuje średnio u 36,7% ludzi, grupa A – u 37,1%, B – u 18,6%, natomiast grupa AB występuje u 7,6% ogółu ludzi

$$H: p_1 = 0.367 \quad p_2 = 0.371 \quad p_3 = 0.186 \quad p_4 = 0.076$$

$$K: \sim H$$

```
pr=c(0.367,.371,.186,.076)
```

```
chisq.test(krew, p=pr)
```

p-value > alfa => nie odrzucamy H

Zadanie 3

W losowo wziętym tygodniu wydarzyło się w Warszawie 414 wypadków i kolizji drogowych, przy czym ich rozkład w poszczególnych dniach tygodnia wyglądał następująco:

Dzień	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota	Niedziela
Liczba wypadków	78	56	52	58	83	42	45

Stwierdzić, czy rozkład liczby wypadków w poszczególne dni tygodnia jest równomierny.

```
wypadki = c(78,56,52, 58, 83, 42, 45)
```

```
chisq.test(wypadki)
```

małe p-value odrzucamy H

Zadanie 4

Na podstawie danych dotyczących preferowanego miejsca spędzania wakacji (plik **wakacje.csv**) stwierdzić, czy rozkład preferencji jest równomierny.

$$H: p_1 = p_2 = p_3 = 1/3$$

$$K: \text{nie prawda ze } H$$

```
wakacje= read.csv2("P:\smwd\wakacje.csv")
```

```
tab=table(wakacje$miejsce)
```

```
chisq.test(tab)
```

duże p-value, większe od alfa nie odrzucamy H

Zadanie 5

Zmierzono czas trwania siedmiu rozmów telefonicznych i otrzymano (w minutach):

2.5 1.8 6.0 0.5 8.75 1.2 3.75.

Na poziomie istotności 0,01 zweryfikować hipotezę, że czas trwania rozmowy ma rozkład wykładniczy o wartości średniej 4 minuty.

H: czas ma rozkład wykładniczy z parametrem $\lambda = 1/4$ (1/średnia)

nie idzie testu chi kwadrat bo żeby go dać do ciągłego to musi być duża próba

```
ks.test(czas, "pexp", 0.25) ## pexp dystrybuanta , 0.25 lambda
```

duże p-value większe od alfa nie odrzucamy H

```
ks.test(x, "pnorm", mi, sigma) // zgodność z r normalnym
```

```
ks.test(x, "pt", n) // zgodność z r t-studenta , n liczba st swobody :)
```

Zadanie 7

W celu zbadania, czy istnieje związek pomiędzy dochodem i posiadanym wykształceniem przeprowadzono badanie na 450 osobowej próbie losowej i otrzymano następujące wyniki:

	Roczny dochód (w tys. PLN)		
	poniżej 120	120-250	powyżej 250
Wykształcenie wyższe	80	115	55
Brak ukończonych studiów	95	70	35

Zweryfikować odpowiednią hipotezę na poziomie istotności 1%.

H: dochód nie zależy od wykształcenia

K: nie prawda ze H

```
dane = matrix(c(80,115,55,95,70,35), nrow = 2, byrow=T)
```

```
## nrow liczba wierszy byrow wpisuj wierszami
```

```
chisq.test(dane)
```

p-value < alfa mniejsze od alfa odrzucamy H

Zadanie 8

Na podstawie wyników badania 200 losowo wybranych osób (plik **wakacje.csv**) stwierdzić, czy istnieje zależność między płcią a preferowanym miejscem spędzania wakacji.

```
wakacje = read.csv2("P:\\smwd\\wakacje.csv")
```

```
tab = table(wakacje)
```

```
chisq.test(tab)
```

pvalue > alfa nie odrzucamy H -> niema zależności

Zadanie 9

Psycholog pracujący w poradni rodzinnej zebrał dane dotyczące powodów kryzysów małżeńskich, które wymieniane były przez przychodzące do poradni pary. Dane te, zamieszczone w poniższej tabeli, pokazują źródła kryzysu postrzegane przez każde z małżonków.

		Przyczyny kryzysu wymieniane przez mężów			
		Pieniądze	Dzieci	Zainteresowania	Inne
Przyczyny kryzysu wymieniane przez żony	Pieniądze	86	31	132	19
	Dzieci	17	64	43	13
	Zainteresowania	54	39	132	33
	Inne	30	17	37	54

Czy na podstawie zebranych danych można stwierdzić, że istnieje zależność poglądów mężów i żon, co do przyczyn kryzysu w ich małżeństwach?

```
dane = matrix(c(86, 31, 132, 19, 17, 64, 43, 13, 54, 39, 132, 33, 30, 17, 37, 54), nrow = 4, byrow = T )  
chisq.test(dane)
```

p-value < alfa mniejsze od alfa odrzucamy H

Kolokwium 2 (14.05.2010)

Zadanie 1

Zbiór **Orange** zawiera m. In. Dane dotyczące obwodu pnia (zmienna **circumference**) drzewek pomarańczowych (mierzonych w mm)

Zbiór ten można otworzyć poleceniem **Orange**

- a) Zbadać, czy można przyjąć założenie normalności rozkładu obwodu pnia drzewek pomarańczowych. Przyjąć poziom istotności test 0,02

```
shapiro.test(Orange$circumference) ## H: próba pochodzi z rozkładu normalnego  
p-value = 0.08483 > alfa -> nie odrzucamy H
```

- b) Zakładając normalność rozkładu obwodu pnia drzewek pomarańczowych, podać 98% przedział ufności dla średniego obwodu pnia, oraz 95% przedział ufności dla odchylenia standardowego obwodu pnia drzewek pomarańczowych

```
t.test(Orange$circumference, conf.level = (0.98))
```

98 percent confidence interval:
92.13582 139.57847

```
prz.ufn.odch=function(x,alfa){  
  n=length(x)  
  s2=var(x)
```

```
  lk= (n-1)*s2/ qchisq(1-alfa/2,n-1)  
  pk= (n-1)*s2/ qchisq(alfa/2,n-1)
```

```
  prz=c(sqrt(lk),sqrt(pk))  
  list(przedzial=prz)  
}
```

```
prz.ufn.odch(Orange$circumference,0.05)  
$przedzial  
[1] 46.50058 75.32113
```

Zadanie 2

Zbiór **samochody.csv** zawiera dane dotyczące parametrów kilku wybranych marek samochodów, m. in. Moc silników (zmienna Moc)

Zbiór ten można otworzyć poleceniem: **read.csv2(„P:/smwd/samochody.csv”)**

Czy można twierdzić, że odsetek samochodów o mocy mniejszej niż 100 różni się istotnie od 0,7 ?

Przyjąć poziom istotności test 0,01. Uwaga na braki w danych !

```
read.csv2(„P:/smwd/samochody.csv”)  
moc=na.omit(samochody$moc)  
mocmniejsza = subset(moc,moc<100)
```

H: p0 = 0.7

K: p0≠0.7

```
prop.test(length(mocmniejsza), length(moc),p=0.7, alt="two.sided", correct=F, conf.level =0.99 )
```

p-value = 0.5579

p-value > alfa – niema podstaw do odrzucenia H -> nie można przyjąć że odsetek się różni

Zadanie 3

Zbiór **ChickWeight** zawiera m.in. dane dotyczące wagi kurczaków(zmienna **weight**) w zależności od stosowanej diety (Zmienna Diet)

a) Czy na poziomie istotności 0,05 można twierdzić, że średnia waga kurczaków przekracza 110 ?

H: $\mu = 110$

K: $\mu > 110$

```
t.test(ChickWeight$weight, mu=110, alt="greater", correct = F)
```

p-value = 7.223e-05

p-val < alfa -> odrzucamy H -> można twierdzić, że waga przekracza 110

b) Czy na poziomie istotności 0,05 można twierdzić, że przy stosowaniu diety nr 1 kurczaki mają średnio mniejszą wagę, niż przy stosowaniu diety nr 2? Przyjąć, że waga kurczaków przy obu rodzajach diet ma rozkład normalny z tą samą wariancją.

```
dieta1 = subset(ChickWeight$weight, ChickWeight$Diet==1)
```

```
dieta2 = subset(ChickWeight$weight, ChickWeight$Diet==2)
```

H: $\mu(\text{dieta1}) = \mu(\text{dieta2})$

K: $\mu(\text{dieta1}) < \mu(\text{dieta2})$

```
t.test(dieta1, dieta2, alt="less", var.equal=T)
```

p-value = 0.002517

p-value < alfa -> odrzucamy H -> kurczaki 1 mają mniejszą wagę od 2

Zadanie 4

Badano, czy istnieje zależność między zawodem uprawianym przez rodziców i ich dorosłe dzieci, w szczególności czy synowie lekarzy prawników i polityków kontynuują rodzinną tradycję. W tym celu zbadano losowo wybraną grupę ojców oraz ich synów i otrzymano następujące wyniki:

Zawód ojca	Zawód syna		
	Lekarz	Prawnik	Polityk
Lekarza	43	27	38
Prawnik	21	38	12
polityk	17	18	68

Czy na podstawie powyższych danych można twierdzić że istnieje taka zależność ? Zweryfikować hipotezę przyjmując poziom istotności testu 0,01.

H: Zawód syna zależy od zawodu ojca

K: Nie prawda że H

```
dane = matrix(c(43,27,38, 21,38,12,17,18,68 ), nrow =3, byrow=T )
```

```
chisq.test(dane)
```

p-value = 2.202e-11

p-value < alfa -> odrzucamy H -> zawód syna nie zależy od zawodu ojca

Kolokwium 2 (21.05.2010)

Zadanie 1

Zbiór **samochody.csv** zawiera dane dotyczące parametrów kilku wybranych marek samochodów, m. in. Mocy silników (zmienna **Moc**)

Zbiór ten można otworzyć poleceniem: **read.csv2(„P:/smwd/samochody.csv”)**

Czy można twierdzić, że odsetek samochodów których moc przekracza 80 jest mniejszy od 0,8 ?

Przyjąć poziom istotności test 0,05. Uwaga na braki w danych !

```
read.csv2(„P:/smwd/samochody.csv”)
moc=na.omit(samochody$moc)
mocwieksza = subset(moc,moc>80)
```

H: $p_0 = 0.8$

K: $p_0 < 0.8$

```
prop.test(length(mocwieksza), length(moc),p=0.8, alt="less", correct=F, conf.level =0.95 )
```

```
p-value = 3.067e-12
```

p-value < alfa -> odrzucamy H -> odsetek jest mniejszy niż 0.8

Zadanie 2

Badano czy zamknięcie jednego z warszawskich mostów istotnie wpłynęło na wydłużenie czasu stania w korkach . W tym celu wylosowano do badania 12 kierowców i zmierzono ile czasy stali w korku w pewnym punkcie miasta przed i po zamknięciu mostu

Przed	32	35	33	23	10	36	39	29	4	9	22	23
Po	38	67	52	28	16	52	61	50	12	19	58	43

Na poziomie istotności 0,05 zweryfikować hipotezę, że średni czas stania w korku po zamknięciu mostu jest istotnie większy od tego przed zamknięciem mostu. Założyć, że badana cecha ma rozkład normalny.

H $\mu_i(\text{przed}) = \mu_i(\text{po})$

K $\mu_i(\text{przed}) < \mu_i(\text{po})$

```
przed = c(32,35,33,23,10,36,39,29,4,9,22,23)
```

```
po = c(38,67,52,28,16,52,61,50,12,19,58,43)
```

```
t.test(przed,po,alt="less", paired=T) ##przed i po więc model 3 -, paired (bo pary)
```

```
p-value = 7.337e-05
```

p-value < alfa -> odrzucamy H -> czas po zamknięciu mostu jest istotnie większy

Zadanie 3

Zbiór **Pima.tr** – znajdujący się w bibliotece **MASS** Zawiera m.in. dane na temat wieku (zmienna **age**) oraz wyniku testu glukozowego (zmienna **glu**) przeprowadzonego wśród Indianek z plemienia Pima. Zbiór ten można otworzyć następująco najpierw otwieramy Bibliotekę poleceniem: **library(MASS)** a następnie zbiór poleceniem **Pima.tr**

- a) Zbadać, czy można przyjąć założenie o normalności rozkładu wyniku testu glukozowego wśród Indianek w wieku powyżej 30 lat.

Przyjąć poziom istotności testu 0,02

```
library(MASS)
```

```
indianki = subset (Pima.tr$glu, Pima.tr$age >30)
```

```
shapiro.test(indianki) ## H: próba pochodzi z rozkładu normalnego
```

p-value = 0.3366

p-value > alfa -> nie odrzucamy H

- b) Zakładając normalność rozkładu wyniku testu glukozowego wśród Indianek w wieku powyżej 30 lat, podać 98% przedział ufności dla średniego wyniku tego testu oraz 98% przedział ufności dla wariancji wyniku tego testu.

```
t.test(indianki, conf.level = (0.98))
```

98 percent confidence interval:

126.8105

144.0228

```
prz.ufn.var=function(x,alfa){  
  n=length(x)  
  s2=var(x)
```

```
  lk= (n-1)*s2/ qchisq(1-alfa/2,n-1)  
  pk= (n-1)*s2/ qchisq(alfa/2,n-1)
```

```
  prz=c(lk,pk)  
  list(przedzial=prz)  
}
```

```
prz.ufn.var(indianki, 0.02)
```

przedzial

791.969

1638.959

Zadanie 4

Poniższa tabela zawiera wyniki sondażu dotyczące preferowanego napoju, spożywanego na śniadanie, przeprowadzone badanie na losowej próbie mieszkańców pewnego miasta(każdy miał wskazać jeden napój z listy):

Napój	kawa	herbata	kakao	sok	mleko	Woda
Liczba osób	47	44	39	50	48	44

Czy na podstawie tych danych można twierdzić, że rozkład preferencji jest równomierny? Zweryfikować odpowiednią hipotezę przyjmując poziom istotności 0,06.

H: $p_1=p_2=p_3=p_4=p_5=p_6=1/6$ (rozkład jest równomierny)

K: nie prawda ze H

```
preferencje = c(47,44,39,50,48,44)  
chisq.test(preferencje)
```

p-value = 0.8937

p-value > alfa -> nie odrzucamy H -> rozkład jest równomierny