Wnioskowanie statystyczne

Lista zadań nr 4, 16 kwietnia 2018.

1. Dane zawierają dzienną liczbę ogłoszeń w gazetach A, B oraz C. Testujemy hipotezę: **średnia liczba ogłoszeń w trzech gazetach jest taka sama**.

```
> dane1 <- read.csv("ww0401.csv", header=F)
> head(dane1, n=3)
   V1 V2
1 246  A   2 231  A   3 236  A
> class(dane1)   [1] "data.frame"
> colnames(dane1)   [1] "V1" "V2"
> colnames(dane1) <- c("inseraty", "gazeta")
> (res1 <- aov(inseraty ~ gazeta, data=dane1))
Call:      aov(formula = inseraty ~ gazeta, data = dane1)
Terms:              gazeta Residuals
Sum of Squares  2870.933  1514.800
Deg. of Freedom        2        12
Residual standard error: 11.23536
Estimated effects may be unbalanced
> summary(res1)
            Df Sum Sq Mean Sq F value Pr(>F)
gazeta       2   2871  1435.5   11.37 0.0017 **
Residuals   12   1515   126.2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2. Dane: dzienna liczba zgłoszeń w czterech działach (A, B, C, D) firmy. Testowana hipoteza: **liczba zgłoszeń jest średnio taka sama we wszystkich oddziałach**.

```
> dane2 <- read.csv("ww0402.csv", header=T)
> dane2$A
 [1] 41 43 45 44 42 48 49 48 47 45 47 45 46 45 48 40 44 39
> (n <- nrow(dane2))      [1] 18
> A <- cbind(dane2$A, rep('A', n))
> B <- cbind(dane2$B, rep('B', n))
> C <- cbind(dane2$C, rep('C', n))
> D <- cbind(dane2$D, rep('D', n))
```

```
> dane2 <- rbind(A, B, C, D)
> (dane2 <- data.frame(na.omit(dane2)))
   X1 X2
1  41  A      2  43  A
...
61 42  D      62 36  D
> dane2$X1 <- as.numeric(dane2$X1)
> (res2 <- aov(X1 ~ X2, data=dane2))
Call:      aov(formula = X1 ~ X2, data = dane2)
Terms:              X2 Residuals
Sum of Squares  374.8532  400.6468
Deg. of Freedom        3        58
Residual standard error: 2.628251
Estimated effects may be unbalanced
> summary(res2)
            Df Sum Sq Mean Sq F value   Pr(>F)
X2           3  374.9  124.95   18.09 2.09e-08 ***
Residuals   58  400.6    6.91
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3. W czterech punktach obserwacyjnych notujemy liczbę wstrząsów. Testujemy hipotezę: **punkty położone są w obrębie tej samej struktury geologicznej**.

```
> dane3 <- read.csv("ww0403.csv", header=T)
> dane3$A
 [1] 64 28 33 83 37 31 60 40 56 45 10 26 49 25 52 62 31 25
> (n <- nrow(dane3))   [1] 18
> A <- cbind(dane3$A, rep('A', n))
...
> (res3 <- aov(X1 ~ X2, data=dane3))
Call:      aov(formula = X1 ~ X2, data = dane3)
Terms:               X2 Residuals
Sum of Squares    160.768 14654.668
Deg. of Freedom         3        58
Residual standard error: 15.89549
Estimated effects may be unbalanced
> summary(res3)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
X2           3    161   53.59   0.212  0.888
Residuals   58  14655  252.67
```

4. Dane: wytrzymałość trzech rodzajów opon pochodzących od pięciu producentów. Testowane hipotezy: **średnia wytrzymałość opony jest niezależna od producenta** oraz **różne rodzaje opon mają różną wytrzymałość**.

```
> (dane4 <- read.csv("ww0404.csv", header=T))
  Czynnik Zimowe Letnie Uniwersalne
1       A    250    255         262
2       B    255    250         262
3       C    245    257         265
4       D    250    255         262
5       E    250    255         262
> library(tidyr)
> d4_long <- gather(dane4, rodzaj, pomiar,
      Zimowe:Uniwersalne, factor_key=T)
> d4_long
   Czynnik      rodzaj pomiar
1        A      Zimowe    250
...
15       E Uniwersalne    262
> (res4 <- aov(pomiar ~ Czynnik + rodzaj, data=d4_long))
Call:      aov(formula = pomiar ~ Czynnik + rodzaj, data = d4_long)
Terms:
                 Czynnik   rodzaj Residuals
Sum of Squares    0.0000 408.9333   84.4000
Deg. of Freedom        4        2         8
Residual standard error: 3.248076
Estimated effects may be unbalanced
> summary(res4)
            Df Sum Sq Mean Sq F value   Pr(>F)
Czynnik      4    0.0    0.00    0.00 1.000000
rodzaj       2  408.9  204.47   19.38 0.000857 ***
Residuals    8   84.4   10.55
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5. Dane: czynnik A – normalny (1) lub podwyższony poziom cukru (2); czynnik B – waga normalna (1) lub powyżej normy (2). Mierzymy ciśnienie. **czy istnieje związek między poziomem cukru a ciśnieniem**, **czy istnieje związek między wagą a ciśnieniem** oraz **czy istnieje interakcja wagi z poziomem cukru**?

```
> (res5 <- aov(Odpowiedź ~ Czynnik.A + Czynnik.B, data=dane5))
Call:      aov(formula = Odpowiedź ~ Czynnik.A + Czynnik.B, data = dane5)
Terms:
                Czynnik.A Czynnik.B Residuals
Sum of Squares     781.25    594.05    739.25
Deg. of Freedom         1         1        17
Residual standard error: 6.594338
Estimated effects may be unbalanced
> dane5 <- read.csv("ww0405.csv", header=T)
> (res5i <- aov(Odpowiedź ~ Czynnik.A * Czynnik.B, data=dane5))
Call:      aov(formula = Odpowiedź ~ Czynnik.A * Czynnik.B, data = dane5)
Terms:
                Czynnik.A Czynnik.B Czynnik.A:Czynnik.B Residuals
Sum of Squares     781.25    594.05                0.05    739.20
Deg. of Freedom         1         1                   1        16
Residual standard error: 6.797058
Estimated effects may be unbalanced
> summary(res5)
            Df Sum Sq Mean Sq F value   Pr(>F)
Czynnik.A    1  781.2   781.2   17.97 0.000554 ***
Czynnik.B    1  594.0   594.0   13.66 0.001793 **
Residuals   17  739.2    43.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(res5i)
                    Df Sum Sq Mean Sq F value   Pr(>F)
Czynnik.A            1  781.2   781.2  16.910 0.000815 ***
Czynnik.B            1  594.0   594.0  12.858 0.002472 **
Czynnik.A:Czynnik.B  1    0.0     0.0   0.001 0.974163
Residuals           16  739.2    46.2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

6. Czynnik A to środek reklamy, czynnik B to główny temat reklamy, dane to procent pozytywnych reakcji na reklamę. **Czy istnieje związek między nośnikiem reklamy a skutecznością reklamy**, **który typ reklamy jest najskuteczniejszy** oraz **czy istnieje związek między rodzajem reklamy a jej nośnikiem**?

```
> dane6 <- read.csv("ww0406.csv", header=T)
> colnames(dane6) <- c("nośnik", "typ", "Sprzedaż")
> summary(aov(Sprzedaż ~ nośnik, data=dane6))
            Df Sum Sq Mean Sq F value   Pr(>F)
nośnik       2  680.1   340.1   12.43 0.000655 ***
Residuals   15  410.3    27.4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> aggregate(Sprzedaż~nośnik, FUN=mean, data=dane6)
    nośnik Sprzedaż
1 Internet 31.16667
2    Prasa 28.50000
3    Radio 17.00000
>
> summary(aov(Sprzedaż ~ typ, data=dane6))
            Df Sum Sq Mean Sq F value Pr(>F)
typ          1      8    8.00   0.118  0.735
Residuals   16   1082   67.65
> aggregate(Sprzedaż~typ, FUN=mean, data=dane6)
     typ Sprzedaż
1   Cena 24.88889
2 Jakość 26.22222
>
> summary(aov(Sprzedaż ~ typ * nośnik, data=dane6))
            Df Sum Sq Mean Sq F value   Pr(>F)
typ          1    8.0     8.0   1.029 0.330507
nośnik       2  680.1   340.1  43.721 3.09e-06 ***
typ:nośnik   2  309.0   154.5  19.864 0.000156 ***
Residuals   12   93.3     7.8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Witold Karczewski

A	B	C
246	243	265
231	246	260
236	243	265
217	235	253
246	235	291
średnie		
235,2	240,4	266,8
średnia		
247,4667		
150,47	49,94	373,78
2 870,93		pomiędzy
116,64	6,76	3,24
17,64	31,36	46,24
0,64	6,76	3,24
331,24	29,16	190,44
116,64	29,16	585,64
1 514,80		wewnątrz

ANOVA - Single Factor

Alfa 0,05

Grupy	Zlicz	Suma	Średnia	Wariancja
Column 1	5	1176	235,2	145,7
Column 2	5	1202	240,4	25,8
Column 3	5	1334	266,8	207,2

Source of Var	SS	df	MS	F	P-value	F critical
Between Grou	2 870,93	2	1 435,47	11,37	0,0017	3,89
Within Groups	1 514,80	12	126,23			
Razem	4 385,73	14				

A	B	C	D
41	44	43	34
43	40	40	37
45	37	42	37
44	43	41	41
42	41	41	37
48	43	39	39
49	42	45	36
48	40	38	41
47	43	40	42
45	42	45	37
47	44	34	37
45	42	43	33
46	40	43	42
45	37	42	36
48		42	
40		39	
44			
39			

ANOVA - Single Factor

Alfa	0,05

Grupy	Zlicz	Suma	Średnia	Wariancja
Column 1	18,0	806,0	44,8	8,4
Column 2	14,0	578,0	41,3	5,1
Column 3	16,0	657,0	41,1	7,7
Column 4	14,0	529,0	37,8	8,0

Source of Vari	SS	df	MS	F	P-value	F critical
Between Grou	390,09	3	130,03	17,57	0,0000	2,76
Within Groups	429,26	58	7,40			
Razem	819,35	61				

A	B	C	D
64	64	55	39
28	28	43	38
33	33	53	45
83	83	18	44
37	37	59	29
31	31	32	30
60	60	27	50
40	40	30	63
56	56	38	29
45	45	25	31
10	10	42	14
26	26	45	48
49	49	20	52
25	25	46	38
52		46	
62		36	
31			
25			

ANOVA - Single Factor

Alfa 0,05

Grupy	Zlicz	Suma	Średnia	Wariancja
Column 1	18,0	757,0	42,1	332,3
Column 2	14,0	587,0	41,9	364,5
Column 3	16,0	615,0	38,4	152,5
Column 4	14,0	550,0	39,3	152,2

Source of Vari	SS	df	MS	F	P-value	F critical
Between Grou	160,77	3	53,59	0,21	0,8876	2,76
Within Groups	14 654,67	58	252,67			
Razem	14 815,44	61				

Czynn	Zimowe	Letnie	Uniwersalne
A	250	255	262
B	255	250	262
C	245	257	265
D	250	255	262
E	250	255	262

ANOVA - Two Factor

Alpha 0,05

Grupy	Zlicz	Suma	Średnia	Wariancja
Column 1	5,0	1250,0	250,0	12,5
Column 2	5,0	1272,0	254,4	6,8
Column 3	5,0	1313,0	262,6	1,8
Row 1	3,0	767,0	255,7	36,3
Row 2	3,0	767,0	255,7	36,3
Row 3	3,0	767,0	255,7	101,3
Row 4	3,0	767,0	255,7	36,3
Row 5	3,0	767,0	255,7	36,3

Source of Vari	SS	df	MS	F	P-value	F critical
Rows	0,00	4	0,00	0,00	1,0000	3,84
Columns	408,93	2	204,47	19,38	0,0009	4,46
Error	84,40	8	10,55			
Total	493,33	14				

Wnioskowanie statystyczne

Lista zadań nr 5, 16 kwietnia 2018.

1. Niezależne zmienne losowe mają rozkłady $Y_1 \sim N(1,3)$ oraz $Y_2 \sim N(2,5)$. Znaleźć rozkład łączny zmiennych $W_1 = Y_1 + 2Y_2$, $W_2 = 4Y_1 - Y_2$.
2. Niezależne zmienne losowe mają rozkłady $Y_1 \sim N(2,3)$ oraz $Y_2 \sim N(1,5)$. Znaleźć rozkład łączny zmiennych $W_1 = 2Y_1 + Y_2$, $W_2 = Y_1 - 2Y_2$.

Do zadań 3–4. Niezależne zmienne losowe mają rozkłady $Y_1 \sim N(0,1)$ oraz $Y_2 \sim N(3,4)$.

3. Znaleźć rozkład zmiennej $Z_1 = Y_1^2$.
4. Niech $Y = \begin{bmatrix} Y_1 & \dfrac{Y_2 - 3}{2} \end{bmatrix}^T$. Znaleźć rozkład zmiennej $Z_2 = Y^T Y$.

Do zadań 5–9. Zmienna $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ ma rozkład $N(\mu, \Sigma)$, gdzie $\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ oraz $\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}$. Niech $U = \dfrac{1}{2\sqrt{15}}(-3Y_1 + 2Y_2)$ oraz $V = \dfrac{1}{2\sqrt{21}}(3Y_1 + 2Y_2 - 12)$.

5. (2 p.) Wykazać, że $Z_1 = U^2 + V^2$. Z_1 jak w zadaniu 8.
6. Sprawdzić, że U, V są niezależne.
7. Sprawdzić, że $U, V \sim N(0,1)$.
8. Jaki jest rozkład zmiennej $Z_1 = (Y - \mu)^T V^{-1} (Y - \mu)$?
9. Jaki jest rozkład zmiennej $Z_2 = Y^T V^{-1} Y$?

   Wsk.: Sprawdzić, że $Z_2 = U^2 + \left(V + \dfrac{2\sqrt{3}}{\sqrt{7}}\right)^2$

Witold Karczewski

rozwiązania – 28 kwietnia 2018

Wnioskowanie statystyczne

Lista zadań nr 6, 16 kwietnia 2018.

1. Dane: *czas*, *wynik* – czas przeznaczony na naukę oraz wynik testu. Zakładamy liniową zależność wyniku testu od czasu przygotowań, $y = \beta_0 + \beta_1 x + \epsilon$. Znaleźć, metodą najmniejszych kwadratów, oszacowania parametrów β_0, β_1.
2. Dane: *rok*, *wartość sprzedaży* produktu przez firmę. Zakładamy liniową zależność sprzedaży od czasu. Oszacować wartość sprzedaży w 2018 roku.
3. Dane: *ciśnienie*, *waga*. Testujemy hipotezę H_0: **ciśnienie nie zależy od wagi**. UWAGA: zmienna niezależna znajduje się w drugiej kolumnie danych.
4. Dane jak w poprzednim zadaniu. Znaleźć współczynnik korelacji i współczynnik korelacji Spearmana pomiędzy *wagą* a *ciśnieniem*.
5. Dane: *czas przeznaczony na TV*, *wynik testu*. Obliczyć współczynniki korelacji pomiędzy powyższymi zmiennymi.

Regresja liniowa

Postać równania regresji: $y = \beta_0 + \beta_1 \cdot x + \varepsilon$,
Oznaczenia: y – zmienna zależna, x – zmienna niezależna
Założenia: błędy regresji podlegają rozkładowi normalnemu, z wartością oczekiwaną 0, wielkość błędu nie zależy od wielkości x, błędy są niezależne
Wzory:

$$SS_{xx} = \sum (x - \bar{x})^2, \quad SS_{xy} = \sum (x - \bar{x})\,(y - \bar{y}),$$
$$b_1 = \frac{SS_{xy}}{SS_{xx}}, \quad b_0 = \bar{y} - b_1\,\bar{x}. \tag{1}$$

Wnioskowanie o nachyleniu prostej regresji

Hipoteza: $H_0 \ : \ \beta_1 = c$.
Statystyka testowa:

$$SE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2} \cdot \frac{1}{SS_{xx}}}$$

$$t = \frac{b_1 - c}{SE} \sim t_{n-2} \tag{2}$$

Współczynnik korelacji:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}} \tag{3}$$

Współczynnik określoności

$$r^2 = \frac{SSR}{SS(total)}. \tag{4}$$

Oznaczenia: $SSR = \sum(\hat{y}_i - \bar{y}_i)^2$ – wariancja modelu, $SSE = \sum(y_i - \hat{y}_i)^2$ – wariancja residualna.

Źródło	df	Skwad	Średnia	F
Model	1	SSR	$MSR = SSR/1$	$F = MSR/MSE$
Residua	$n-2$	SSE	$MSE = SSE/(n-2)$	
Razem	$n-1$	$SS(total)$		

Przedziały ufności i prognozy

Przedział prognozy dla wartości y przy $x = x_0$:

$$\hat{y} \pm t_{n-2}(\alpha/2)\sqrt{\frac{SSE}{n-2}}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}. \tag{5}$$

Przedział ufności dla średniej wartości y przy $x = x_0$:

$$\hat{y} \pm t_{n-2}(\alpha/2)\sqrt{\frac{SSE}{n-2}}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}. \tag{6}$$

Witold Karczewski

rozwiązania – 12 maja 2018

Wnioskowanie statystyczne

Lista zadań nr 3, 9 kwietnia 2018.

Od zadania 2 – obliczamy *wartość_p*.

1. Zmierzono wzrost 10 mężczyzn i 10 kobiet. Testujemy hipotezę $H_0 : \mu_m = \mu_k$, wobec hipotezy $H_1 : \mu_m > \mu_k$. Obliczyć (poziom istotności $\alpha = .05$):

   (a) wartość statystyki testowej i przedział akceptacji,

   (b) przedział ufności dla różnicy wartości średnich.

```
# Zadanie 1.
dane1 <- read.csv("w-03-01.csv", sep=";", header=FALSE)
dane1
mean1 <- mean(dane1[,1])   # 170.9
mean2 <- mean(dane1[,2])   # 165.7
s1sq <- sum((dane1[,1]-mean1)^2)
s2sq <- sum((dane1[,2]-mean2)^2)
n <- nrow(dane1)
ssq <- (s1sq+s2sq)/(n+n-2)   # 11.83
# Wartość statystyki testowej
tstat <- (mean1 - mean2)/(sqrt(ssq*2/n))   # 3.3801
# Wartość dystrybuanty
pt(tstat, df=2*n-2)   # 1.734064
# Prawy kraniec obszaru akceptacji
t1 <- qt(0.95, df=2*n-2)   # 1.7341
# Przedział ufności
t1 <- qt(0.975, df=2*n-2)   #  2.1009
t2 <- qt(0.025, df=2*n-2)   # -2.1009
roznica <- mean1 - mean2
polowa <- sqrt(ssq*2/n)
przedzial <- c(roznica - polowa, roznica + polowa)
# [3.6616; 6.7384]
```

2. Dane zawierają wiek pierwszego małżeństwa dla 50 mężczyzn i 50 kobiet (odpowiednio w 1. i 2. kolumnie). Testujemy hipotezę $H_0 : \mu_m - \mu_k = 2$, wobec hipotezy $H_1 : \mu_m - \mu_k > 2$. Obliczyć *wartość_p*.

```
# Zadanie 2.
dane2 <- read.csv("w-03-02.csv", header=FALSE, sep=";")
dane <- dane2[,1]-dane2[,2]
mean1 <- mean(dane)   # 5.94
n <- length(dane)
ssq <- sum( (dane-mean1)^2 )
```

```
tstat <- mean1*sqrt(n-1)/sqrt(ssq)   # 1.3317
pt(tstat, df=n-1)   # 0.9054
# Akceptujemy hipotezę zerową
```

3. Dane (w kolumnach) przedstawiają pomiar wagi przed i po okresie stosowania określonej diety dla 16 osób. Testujemy hipotezę: **dieta ma wpływ na wagę**.

```
# Zadanie 3.
dane3 <- read.csv("w-03-03.csv", header=FALSE, sep=";")
dane <- dane3[,1]-dane3[,2]
mean1 <- mean(dane)
n <- length(dane)
ssq <- sum( (dane-mean1)^2 )
tstat <- mean1*sqrt(n-1)/sqrt(ssq)   # 0.9051
pt(tstat, df=n-1)   # 0.8101
# Akceptujemy hipotezę zerową
```

4. Wybrano próbkę 200 kobiet i 200 mężczyzn korzystających z sieci. W próbce tej 159 kobiet i 138 mężczyzn używało komunikatorów. Testujemy hipotezę: **kobiety częściej używają komunikatorów**.

```
# Zadanie 4.
n1 <-200
n2 <-200
n <- n1*n2/(n1+n2)   # 100
m1 <- 159
m2 <-138
p <- (m1+m2)/(n1+n2) # 0.7425
p1 <- m1/n1   # 0.795
p2 <- m2/n2   # 0.690
z <- (p1-p2)/ sqrt(p*(1-p)/n)   # 2.4013
pnorm(z)   # 0.9918
# Wartość_p
1 - pnorm(z)   # 0.082
```

5. Przeprowadzono badanie zmienności dwóch partii towaru. Liczebności próbek to $n_1 = n_2 = 12$. Testujemy hipotezę: **zmienność dwóch próbek jest różna**.

```
# Zadanie 5.
dane5 <- read.csv("w-03-05.csv", header=FALSE, sep=";")
n <- length(dane5[,1])
mean1 <- mean(dane5[,1]
mean2 <- mean(dane5[,2])
ssq1 <- sum( (dane5[,1]-mean1)^2 )   # 222.4
```

```
ssq2 <- sum( (dane5[,2]-mean2)^2 )   # 41.3
f <- ssq1/ssq2   # 5.38
pf(f,df1=n-1,df2=n-1) # 0.9952
# Wartość_p
1- pf(f,df1=n-1,df2=n-1) # 0.0048
```

6. Dane zawierają koszt ubezpieczenia w kolejnych latach. Próbki nie są skojarzone. Testujemy hipotezę: **koszt ubezpieczenia wzrósł o 200 złotych**.

```
# Zadanie 6.
dane6 <- read.csv("w-03-06.csv",header=FALSE, sep=";")
n <- nrow(dane6)
mean1 <- mean(dane6[,1])   # 671.48
mean2 <- mean(dane6[,2])   # 895.08
mu0 <- 200
licznik <- mean2 - mean1 - mu0   # 23.6
tmp <- sum( (dane6[,1]- mean1)^2) + sum( (dane6[,2]-mean2)^2)
mianownik <- sqrt( tmp * (1/n + 1/n) /(n+n-2))   # 5.053
ttest <- licznik /mianownik   # 4.670
pt(ttest,df=n-1)   # 0.9999881
# Wartość_p
1 - pt(ttest,df=n-1)   # 1.19e-05
```

7. Dane zawierają wyniki testu dla dwu grup uczniów: tradycyjny sposób nauczania, sposób eksperymentalny. Testujemy hipotezę: **nowy sposób nauczania daje na ogół lepsze wyniki**.

```
# Zadanie 7.
dane7 <- read.csv("w-03-07.csv",header=FALSE, sep=";")
n <- nrow(dane7)
mean1 <- mean(dane7[,1])   # 81.13
mean2 <- mean(dane7[,2])   # 83.07
licznik <- mean2 - mean1   # 1.93
tmp <- sum( (dane7[,1]- mean1)^2) + sum( (dane7[,2]-mean2)^2)
mianownik <- sqrt( tmp * (1/n + 1/n) /(n+n-2))   # 2.58
ttest <- licznik /mianownik   # 0.7491
pt(ttest,df=n-1)   # 0.7669
# Wartość_p
1 - pt(ttest,df=n-1)   # 0.2331
```

8. 10 poletek doświadczalnych podzielono na dwie części, w jednej z nich przeprowadzono dodatkowe czynności *agrotechniczne*. W wierszu znajduje się wydajność części poddanej dodatkowym zabiegom i części poletka uprawianej tradycyjnie. Testujemy hipotezę: **dodatkowy czynnik ma wpływ na wydajność uprawy**.

```
dane8 <- read.csv("w-03-08.csv",header=FALSE, sep=";")
n <- nrow(dane8)
dane <- dane8[,1]-dane8[,2]
mean1 <- mean(dane)   # 4.1
ssq <- sum( (dane-mean1)^2 )   # 274.9
ttest <- mean1 * sqrt( (n-1)/ssq )   # 0.7418
pt(ttest,df=n-1)   # 0.7615
# Wartość_p
1 - pt(ttest,df=n-1)   # 0.2385
```

9. (2 p.) Liczba cyklonów w południowo-wschodniej części Australii wynosiła, w kolejnych latach: 6 5 4 6   6 3 12 7   4 2 6 7. Załóżmy, że tę liczbę (cyklonów) opisuje rozkład Poissona z parametrem λ. Znaleźć oszacowanie parametru λ (na przykład metodą największej wiarygodności).

```
# Zadanie 9.
dane9 <- c(6, 5, 4, 6, 6, 3, 12, 7, 4, 2, 6, 7)
estymator <- mean(dane9)   # 5.67
```

Witold Karczewski