

Statistinės duomenų analizės praktinės užduotys

2017

8. Neparametriniai kriterijai.

- (a) *Spirmeno koreliacijos koeficientas. Požymių nepriklausomumo tikrinimas.* (ČM II.1.7) Kavos rūšių Atuvos rinkoje dalys (procentais) ir reklamos išlaidos pateiktos lentelėje.

Rūšis	A	B	C	D	E	F	Kita
Rinkos dalis %	15.7	3.9	10.6	9.6	12.3	26.2	21.7
Reklamos išlaidos (tūkst. eurų)	7.6	3.5	6.1	6.8	8.3	10.1	7.1

Apskaičiuokite Spirmeno koreliacijos koeficientą. Ar kintamieji yra priklausomi?

```
Rinkos_Dalis <- c(15.7, 3.9, 10.6, 9.6, 12.3, 26.2, 21.7)
Reklamos_Islaidos <- c(7.6, 3.5, 6.1, 6.8, 8.3, 10.1, 7.1)
t <- cor.test(Rinkos_Dalis, Reklamos_Islaidos, method="spearman")
pv <- t$p.value
rho <- t$estimate
t
```

```
##
## Spearman's rank correlation rho
##
## data: Rinkos_Dalis and Reklamos_Islaidos
## S = 10, p-value = 0.03413
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.8214286
```

$\hat{\rho} = 0.8214286$, $pv = 0.034127 < \alpha = 0.05 \Rightarrow$ kintamieji priklausomi.

- (b) *Dviejų nepriklausomų imčių palyginimas: Vilkoksono - Mano - Vitnio kriterijus.* (ČM II.1.3). Besiruošdamas jubiliejui, verslininkas Apolonijus Drinkūnas atliko eksperimentą. Prisipirkęs pigių ir brangių šventinių žvakučių, jis fiksavo laiką, per kurį žvakutės sudega. Ar ponas Apolonijus teisus nutaręs, kad žvakučių degimo laikas nesiskiria ir jubiliejui tiks pigiosios žvakutės? Duomenys pateikti lentelėje.

Pigios	25	27	23	28	22	20	
Brangios	31	26	30	24	29	21	33

```
Pigios <- c(25, 27, 23, 28, 22, 20)
Brangios <- c(31, 26, 30, 24, 29, 21, 33)

t <- wilcox.test(Pigios, Brangios)
pv <- t$p.value
t
```

```
##
## Wilcoxon rank sum test
##
## data: Pigios and Brangios
## W = 10, p-value = 0.1375
```

alternative hypothesis: true location shift is not equal to 0

$p_v = 0.1375291 > \alpha = 0.05 \Rightarrow$ degimo laikas statistiškai reikšmingai nesiskiria.

- (c) *Dviejų priklausomų imčių palyginimas: Vilkoksono žymėtųjų rangų kriterijus.* (ČM II.1.2) Norima patikrinti, ar tam tikri pratimai mažina sistolinį kraujo spaudimą. Atsitiktinai parinktų 15 žmonių spaudimas buvo matuotas prieš darant pratimus ir po to. Duomenys pateikti lentelėje.

	Asmuo	Prieš	Po	Asmuo	Prieš	Po
1		164	162	9	138	139
2		146	144	10	130	124
3		148	146	11	175	170
4		154	156	12	146	147
5		143	145	13	160	157
6		160	159	14	160	140
7		150	145	15	153	148
8		148	150			

Ar pratimai efektyvūs?

```
Pries <- c(164,146,148,154,143,160,150,148,138,130,175,146,160,160,153)
Po <- c(162,144,146,156,145,159,145,150,139,124,170,147,157,140,148)

t <- wilcox.test(Po, Pries, paired = TRUE, alternative = "less")
```

```
## Warning in wilcox.test.default(Po, Pries, paired = TRUE, alternative =
## "less"): cannot compute exact p-value with ties
```

```
pv <- t$p.value
t
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: Po and Pries
## V = 23.5, p-value = 0.01957
## alternative hypothesis: true location shift is less than 0
```

$p_v = 0.0195743 < \alpha = 0.05 \Rightarrow$ duomenys neprieštarauja hipotezei, kad spaudimas sumažėja.

- (d) *Kelių nepriklausomų imčių palyginimas: Kruskalio - Voliso kriterijus.* Trijose gamyklose buvo testuojami kineskopai. Jų funkcionavimo trukmės (mėnesiais iki pirmo gedimo) surašyti pateikiamoje lentelėje.

1 gamyklos kineskopai	41	70	26	89	62	54	46	77	34	51		
2 gamyklos kineskopai	30	69	42	60	44	74	32	47	45	37	52	81
3 gamyklos kineskopai	41	70	26	89	62	54	46	77	34	51	61	

```
Gamykla1 <- c(41,70,26,89,62,54,46,77,34,51)
Gamykla2 <- c(30,69,42,60,44,74,32,47,45,37,52,81)
Gamykla3 <- c(23,35,29,38,21,53,31,25,36,50,61)

data <- list(g1=Gamykla1, g2=Gamykla2, g3=Gamykla3)
t <- kruskal.test(data)
pv <- t$p.value
t
```

```
##
## Kruskal-Wallis rank sum test
##
## data: data
## Kruskal-Wallis chi-squared = 6.549, df = 2, p-value = 0.03784
```

$pv = 0.0378354 < \alpha = 0.05 \Rightarrow$ hipotezė apie vidutinių trukmių lygybę atmetama (“bet nelabai reikšmingai”).

- (e) *Kelių priklausomų imčių palyginimas: Frydmano kriterijus*. Penki nepriklausomi ekspertai vertino 3 rūšių (A, B ir C) alų. Duomenys pateikti lentelėje.

Ekspertas	A	B	C
Pirmas	10	7	8
Antras	5	2	4
Trečias	6	8	6
Ketvirtas	3	4	6
Penktas	9	8	10

```
A<-c(10,5,6,3,9)
B<-c(7,2,8,4,8)
C<-c(8,4,6,6,10)
data <- matrix(rbind(A,B,C), nrow=5,byrow=TRUE)

t <- friedman.test(data)
pv <- t$p.value
t
```

```
##
## Friedman rank sum test
##
## data: data
## Friedman chi-squared = 1.3684, df = 2, p-value = 0.5045
```

$pv = 0.5044884 > 0.05 \Rightarrow$ duomenys neprieštarauja hipotezei, jog skirstiniai vienodi.

- (f) *Kelių priklausomų imčių palyginimas: Frydmano kriterijus*. (ČM II.1.5). Lentelėje pateikti duomenys apie 3 tiekėjų siūlomų 12 skirtingų tipų spausdintuvų kainas.

Tipas	1 tiekėjas	2 tiekėjas	3 tiekėjas
1	660	673	658
2	790	799	785
3	590	580	599
4	950	945	960
5	1290	1280	1295
6	1550	1500	1499
7	1980	1950	1970
8	2300	2295	2310
9	2500	2480	2490
10	2190	2190	2210
11	5590	5500	5550
12	6000	6100	6090

Ar tiekėjų siūlomos spausdintuvų kainos skiriasi?

```
Tiekejas1 <- c(660,790,590,950,1290,1550,1980,2300,2500,2190,5590,6000)
Tiekejas2 <- c(673,799,580,945,1280,1500,1950,2295,2480,2190,5500,6100)
Tiekejas3 <- c(658,785,599,960,1295,1499,1970,2310,2490,2210,5550,6090)
data <- matrix(rbind(Tiekejas1,Tiekejas2,Tiekejas3), nrow=12, byrow=TRUE)
t <- friedman.test(data)
pv <- t$p.value
t
```

```
##
## Friedman rank sum test
##
## data: data
## Friedman chi-squared = 2.5957, df = 2, p-value = 0.2731
```

$p = 0.2731123 > \alpha = 0.05 \Rightarrow$ duomenys neprieštaruoja hipotezei, jog skirstiniai vienodi.

- (g) *Proporcijų palyginimas, kai imtys priklausomos: Maknemaros kriterijus.* Prieš rinkiminius debatus 1000 žmonių iš įvairių visuomenės sluoksnių atsakė į klausimą: “ar balsuosite už kandidatą N?”. Po rinkiminių debatų tie patys žmonės atsakė į tą patį klausimą. Rezultatai pateikti lentelėje.

	Taip (po)	Ne (po)
Taip (prieš)	421	115
Ne (prieš)	78	386

Ar rinkiminiai debatai pakeitė rinkėjų nuomonę?

```
data <- matrix(
  c(421, 78, 115, 386),
  nrow=2,
  dimnames=list(
    "Pries"=c("Taip", "Ne"),
    "Po"=c("Taip", "Ne")
  )
)

t <- mcnemar.test(data)
pv <- t$p.value
t
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: data
## McNemar's chi-squared = 6.715, df = 1, p-value = 0.00956
```

$p = 0.0095604 < \alpha = 0.05 \Rightarrow$ Atmetame hipotezę, jog debatai nepakeitė rinkėjo nuomonės, t.y. rinkėjai po debatų pakeitė nuomonę.

- (h) *Serių kriterijus.* (ČM II.1.3 pvz.) Degalinėje yra A-95 (A) ir A-98 (B) oktaninio skaičiaus benzino. Vieną dieną 50 automobilių A ir B prisipylė tokia tvarka:

AABAABABBAAABBABBABBABBAABABBABBAABBBBAABABABAAABA

Ar galima teigti, kad benzino rūšies pasirinkimas yra atsitiktinis (t.y. paros metas neturi įtakos benzino rūšies pasirinkimui)?

```
data <- factor(
  c(1, 1, 2, 1, 1, 2, 1,
    2, 2, 1, 1, 1, 2, 2,
    1, 2, 2, 1, 2, 2, 1,
    2, 2, 1, 1, 2, 1, 2,
    2, 1, 2, 2, 1, 1, 2,
    2, 2, 2, 1, 1, 2, 1,
    2, 1, 2, 1, 1, 1, 2,
    1
  )
)
```

```
# install.packages("tseries")
library(tseries)
packageVersion("tseries")
```

```
## [1] '0.10.42'
```

```
t <- runs.test(data)
pv <- t$p.value
t
```

```
##
## Runs Test
##
## data: data
## Standard Normal = 1.4289, p-value = 0.153
## alternative hypothesis: two.sided
```

$p_v = 0.1530419 > \alpha = 0.05 \Rightarrow$ hipotezės neatmetame, t.y. paros metas neturi įtakos benzino rūšies pasirinkimui.

Padaryta su R version 3.4.2 (2017-09-28), x86_64-pc-linux-gnu.