cv8.R

Marek

2023-04-06

```
library('latexpdf')
```

Neparametricke metody

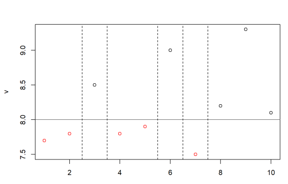
neparametricke metody pouzivame tam, kde data nie su normalne rozdelene, je ich malo, su nestandardne. Neparametricke testy vyzaduju podmienku, ze data su z nejakeho spojiteho rozdelenia. Su menej presne a menej citlive.

Hypoteza H0 je v tvare $H_0 \quad \text{data su nahodne} \quad H_1 \quad \text{nie su nahodne}$ Test serii (Wald Wofovitzov test), testujeme nahodnost dat. Test je citlivy na trend, kniznica randtests

```
library('randtests')
```

Linka MHD v rannej spicke prejde trasu priemernou rychlostou 8 km/h. Bola navrhnuta mala zmena trasy s cielom zrychlit dopravu. Pocas 10 dni sme namerali udaje o priemernej rychlosti datovy subor-v. Na hladine vyznamnosti $\alpha = 0.05$ testujte hypotezu o nahodnosti dat. Testujte nahodnost merani.

```
v<- c(7.7, 7.8, 8.5, 7.8, 7.9, 9, 7.5, 8.2, 9.3, 8.1)
runs.test(v, plot=T)
```



```
##
##  Runs Test
##
## data:  v
## statistic = 0, runs = 6, n1 = 5, n2 = 5, n = 10, p-value = 1
## alternative hypothesis: nonrandomness
```

Phodnota > 0.05, data su nahodne

```
runs.test(v, alternative = 'left.sided')
```

```
##
##  Runs Test
##
## data:  v
## statistic = 0, runs = 6, n1 = 5, n2 = 5, n = 10, p-value = 0.5
## alternative hypothesis: trend
```

```
runs.test(v, alternative = 'right.sided')
```

```
##
##  Runs Test
##
## data:  v
## statistic = 0, runs = 6, n1 = 5, n2 = 5, n = 10, p-value = 0.5
## alternative hypothesis: first-order negative autocorrelation
```

Test kritickych bodov, bodov obratu (turning point test) odhaluje periodicitu v datach

```
turning.point.test(v)
```

```
##
##  Turning Point Test
##
## data:  v
## statistic = -0.27629, n = 10, p-value = 0.7823
## alternative hypothesis: non randomness
```

P hodnota > 0.05, nezamietam hypotezu o nahodnosti dat

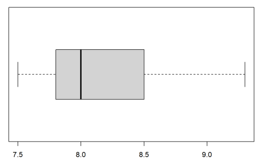
```
turning.point.test(v, alternative = 'left.sided')
```

```
##
##  Turning Point Test
##
## data:  v
## statistic = -0.27629, n = 10, p-value = 0.3912
## alternative hypothesis: positive serial correlation
```

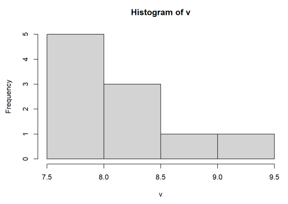
Neparametricke testy o polohe (mediane). Parametrickymi testami

sme testovali tvrdenia o strednej hodnote, boli jednovyberove t.testy a dvojvyberove t.testy - parove testy a neparove # Znamienkovy test (sign test) testujeme $H_0 \quad \text{median} = x_0 \quad H_1 \quad \text{median} \neq x_0$ Jedina podmienka kladena na data je, aby boli vyberom zo spojiteho rozdelenia. Test ma malu silu, t.j. chyba druheho druhu je velka (nezamietame H0 a pritom H0 neplati). Testujte, ze data z predosleho prikladu maju median = 8, teda ze priemerna rychlost ani po uprave sa nezmenila.

```
boxplot(v, horizontal = T)
```



```
hist(v)
```



pre symetricke data je sikmost nulova, spocitame este sikmost

```
library('moments')
skewness(v)
```

```
## [1] 0.8311761
```

```
library('BSDA')
```

```
## Warning: package 'BSDA' was built under R version 4.2.3
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'BSDA'
```

```
## The following object is masked from 'package:datasets':
##
##     Orange
```

```
SIGN.test(v, md=8)
```

```
##
##  One-sample Sign-Test
##
## data:  v
## s = 5, p-value = 1
## alternative hypothesis: true median is not equal to 8
## 95 percent confidence interval:
##  7.732444 8.837778
## sample estimates:
## median of x
##           8
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level L.E.pt U.E.pt
## Lower Achieved CI     0.8906 7.8000 8.5000
## Interpolated CI       0.9500 7.7324 8.8378
## Upper Achieved CI     0.9785 7.7000 9.0000
```

P hodnota > 0.05 nezamietame hypotezu o tom, ze priemerna rychlost sa nezmenila. Tento test mozeme pouzit aj ako dvojvyberovy parovy. Testujeme, ze median rozdielov je rovny nejakemu cislu, pouzijeme ho tam, kde je asymetria dat. Ak pridame predpoklad, ze data su symetricke okolo medianu, tak radsej pouzijem jednovyberovy Wilcoxonov test -signed rank test. Otestujeme nase data, ci su symetricke a ak ano testujeme tymto testom.

```
library('lawstat')
```

```
## Warning: package 'lawstat' was built under R version 4.2.3
```

```
##
## Attaching package: 'lawstat'
```

```
## The following object is masked from 'package:randtests':
##
##     runs.test
```

```
symmetry.test(v)
```

```
##
##  m-out-of-n bootstrap symmetry test by Miao, Gel, and Gastwirth (2006)
##
## data:  v
## Test statistic = 1.3662, p-value = 0.188
## alternative hypothesis: the distribution is asymmetric.
## sample estimates:
## bootstrap optimal m
##                  10
```

p hodnota >0.05, radsej Wilcoxonov test

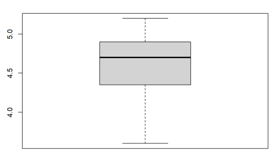
```
wilcox.test(v, mu=8)
```

```
## Warning in wilcox.test.default(v, mu = 8): cannot compute exact p-value with
## ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  v
## V = 31, p-value = 0.7593
## alternative hypothesis: true location is not equal to 8
```

P hodnota >0.05, nezamietam H0, priemerna rychlost sa nezmenila Pri tradicnom opracovani suciastok sa dosahovali priemerne hodnoty kvalitativnej vlastnosti 4.4, pokusne sa zavadza nova jednoduchsia metoda opracovania suciastok. Hodnoty kvalitativnej vlastnosti su v datovom subore x. Testujte hypotezu, ze kvalitativna vlastnost aj pri novej metode zostala rovnaka.

```
x<-c(4.5, 4.3, 4.1, 4.9, 4.6, 3.6, 4.7, 5.1, 4.8, 4, 3.7, 4.4,
     4.9, 4.9, 5.2, 5.1, 4.7, 4.9, 4.6, 4.8)
boxplot(x)
```



```
symmetry.test(x)
```

```
##
##  m-out-of-n bootstrap symmetry test by Miao, Gel, and Gastwirth (2006)
##
## data:  x
## Test statistic = -1.528, p-value = 0.108
## alternative hypothesis: the distribution is asymmetric.
## sample estimates:
## bootstrap optimal m
##                  14
```

Data su symetricke, teda jednoznacne WT.

```
wilcox.test(x, mu=4.4)
```

```
## Warning in wilcox.test.default(x, mu = 4.4): cannot compute exact p-value with
## ties
```

```
## Warning in wilcox.test.default(x, mu = 4.4): cannot compute exact p-value with
## zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x
## V = 135, p-value = 0.1114
## alternative hypothesis: true location is not equal to 4.4
```

P hodnota>0.05, nezamietam hypotezu, ze novou metodou opracovane suciastky maju rovnaku kvalitativnu vlastnost Prakticky by sme mali na zaciatku overit normalitu dat a ak su normalne rozdelene, tak t.test

```
library('nortest')
shapiro.test(x) #normalita
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.91818, p-value = 0.09142
```

```
t.test(x, mu=4.4)
```

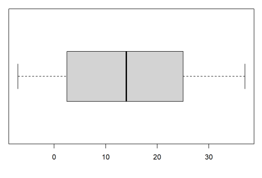
```
##
##  One Sample t-test
##
## data:  x
## t = 1.8808, df = 19, p-value = 0.07542
## alternative hypothesis: true mean is not equal to 4.4
## 95 percent confidence interval:
##  4.37856 4.80144
## sample estimates:
## mean of x
##      4.59
```

Aj parametrickym testom nam vyslo, ze kvalitativna vlastnost ostala rovnaka. Priklad pouzitia, ako parovy test. Su dane casy v sekundach, pocas ktorych vyriesili kontrolne ulohy ziaci pred a po specialnych cviceniach z pamatoveho pocitania. Zlepsili cvicenia schopnost ziakov rychlejsie riesit ulohy? Ak sa pytame, ci zlepsili, pred - po >= 0, teda negacia je <0 alternativa bude less. Testy vyberam podla toho, ci rozdiely su alebo nie su symetricke.

```
pred <- c(87,61,98,90,93,74,83,72,81,75,83)
po <- c(50,45,79,90,88,65,52,79,84,61,52)
rozdiel = pred - po
rozdiel
```

```
##  [1] 37 16 19  0  5  9 31 -7 -3 14 31
```

```
boxplot(rozdiel, horizontal = T)
```



```
skewness(rozdiel)
```

```
## [1] 0.1819471
```

```
symmetry.test(rozdiel) #symetria dat
```

```
##
##  m-out-of-n bootstrap symmetry test by Miao, Gel, and Gastwirth (2006)
##
## data:  rozdiel
## Test statistic = -0.053887, p-value = 0.908
## alternative hypothesis: the distribution is asymmetric.
## sample estimates:
## bootstrap optimal m
##                   8
```

```
wilcox.test(rozdiel, alternative = 'less')
```

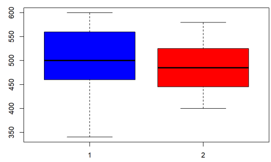
```
## Warning in wilcox.test.default(rozdiel, alternative = "less"): cannot compute
## exact p-value with ties
```

```
## Warning in wilcox.test.default(rozdiel, alternative = "less"): cannot compute
## exact p-value with zeroes
```

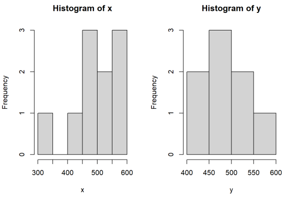
```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  rozdiel
## V = 51, p-value = 0.9928
## alternative hypothesis: true location is less than 0
```

p hodnota>0.05, nezamietame H0, ziaci sa zlepsili. Dvojvyberovy neparovy test, neparametricky, dvojvyberovy Wilcoxonov test (Mann Whitney U test). Nulova hypoteza je $H_0 \quad F_X = F_Y$. Pred testom treba overit, ci sa rozdelenia aspon priblizne rovnaju a tiez ci maju rovnaku disperziu. Ak su velke rozdiely, tak dvojvyberovy Kolmogorov Smirnov test. Neparametricky test pre rovnost disperzii Levene test. Z produkcie dvoch firiem bolo nahodne vybratych n=10 a m=8 vyrobkov. Nezavisli experti hodnotili ich kvalitu pridelenim bodov. Datovy subor x, y. Testujte hypotezu, ze kvalita vyrobkov dvoch firiem je rovnaka.

```
x<-c(420,560,600,490,550,570,340,480,510,460)
y<-c(400,420,580,470,470,500,520,530)
boxplot(x, y, col = c('blue', 'red'))
```



```
par(mfrow=c(1,2))
hist(x, breaks = 5)
hist(y, breaks = 5)
```



```
par(mfrow = c(1,1))
```

Uprava dat pre test

```
data <- data.frame('body' = c(x, y), 'firma' = rep(c(1,2), times = c(10,8)))
data
```

```
##    body firma
## 1   420     1
## 2   560     1
## 3   600     1
## 4   490     1
## 5   550     1
## 6   570     1
## 7   340     1
## 8   480     1
## 9   510     1
## 10  460     1
## 11  400     2
## 12  420     2
## 13  580     2
## 14  470     2
## 15  470     2
## 16  500     2
## 17  520     2
## 18  530     2
```

```
levene.test(data$body, data$firma)
```

```
##
##  Modified robust Brown-Forsythe Levene-type test based on the absolute
##  deviations from the median
##
## data:  data$body
## Test Statistic = 0.50758, p-value = 0.4865
```

Mozeme pouzit dvojvyberovy WT

```
wilcox.test(x,y)
```

```
## Warning in wilcox.test.default(x, y): cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 45.5, p-value = 0.6565
## alternative hypothesis: true location shift is not equal to 0
```

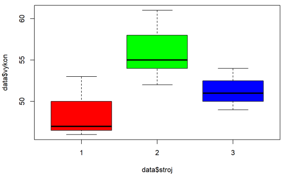
P hodnota > 0.05, nezamietame hypotezu o rovnosti DF, teda aj medianov, kvalita je rovnaka, este KS test

```
ks.test(x, y)
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.275, p-value = 0.7935
## alternative hypothesis: two-sided
```

plati tvrdenie hore # Kruskal Wallisov test, neparametricky ekvivalent k ANOVA. ANOVA mala silne podmienky, normalita dat v triedach, rovnost disperzii. Ak to nie je splnene, tak KW test. Pri ANOVE sme testovali, ze stredne hodnoty sa rovnaju, tu testujeme, ze distribucne funkcie sa rovnaju (a teda aj mediany). Zaznamenali sme vykony strojov troch znaciek. Na hladine vyznamnosti $\alpha = 0.05$ testujte hypotezu vykony strojov su rovnake.

```
data<-data.frame("vykon"=c(53,47,46,61,55,52,58,54,51,51,49,54),
                 "stroj"=c(rep(1,3),rep(2,5),rep(3,4)))
boxplot(data$vykon~data$stroj, col=c('red', 'green', 'blue'))
```



Faktorizujeme

```
data$stroj <- factor(data$stroj)
data$stroj
```

```
##  [1] 1 1 1 2 2 2 2 2 3 3 3 3
## Levels: 1 2 3
```

```
kruskal.test(data$vykon, data$stroj)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data$vykon and data$stroj
## Kruskal-Wallis chi-squared = 6.4021, df = 2, p-value = 0.04072
```

P hodnota <0.05, zamietame hypotezu o rovnakej vykonnosti Nasleduju post testy. ANOVA (Tukey, Scheffe), tu pouzijeme dunn.test (aj kniznica sa tak vola)

```
library('dunn.test')
dunn.test(data$vykon, data$stroj)
```

```
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 6.4021, df = 2, p-value = 0.04
##
##
##                            Comparison of x by group
##                                 (No adjustment)
## Col Mean-|
## Row Mean |          1          2
## ---------+----------------------
##        2 |  -2.350191
##          |    0.0094*
##          |
##        3 |  -0.652906   1.815197
##          |     0.2569     0.0347
##
## alpha = 0.05
## Reject Ho if p <= alpha/2
```

```
dunn.test(data$vykon, data$stroj, altp = T, list = T)
```

```
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 6.4021, df = 2, p-value = 0.04
##
##
##                            Comparison of x by group
##                                 (No adjustment)
## Col Mean-|
## Row Mean |          1          2
## ---------+----------------------
##        2 |  -2.350191
##          |    0.0188*
##          |
##        3 |  -0.652906   1.815197
##          |     0.5138     0.0695
##
##
## List of pairwise comparisons: Z statistic (p-value)
## ---------------------------
## 1 - 2 : -2.350191 (0.0188)*
## 1 - 3 : -0.652906 (0.5138)
## 2 - 3 :  1.815197 (0.0695)
##
## alpha = 0.05
## Reject Ho if p <= alpha
```

lisi sa prva druha trieda