Two-sample Position and Scale Test

Weihao Lu

2018/3/30

Reference: Wang. J. L., Non-parametric statistical analysis,2006.4

Brown-Mood Median Test

page 55

```
BM.test=function(x,y,alternative)
{
    ##Accurate test,
    # normal approximation,
    # normal approximation after continuity correction
    ##coef alternative can be greater, less or both
    xy=c(x,y)
    md.xy=median(xy)
    t=sum(xy>md.xy)
    lx=length(x[x!=md.xy])
    ly=length(y[y!=md.xy])
    lxy=lx+ly
    A=sum(x>md.xy)  ##Test statistics A
    z1=(A-lx*t)/(lx+ly)/(lx*ly*t*(lx+ly-t)/(lx+ly)^3)^0.5
    ##Normalized Statistics for Normal Approximation
    if(A>(min(lx,t)/2)){
        z2=(A+0.5-lx*t)/(lx+ly)/(lx*ly*t*(lx+ly-t)/(lx+ly)^3)^0.5
        ##Normalized statistics for normal approximation after continuity correction
    }else{z2=(A-0.5-lx*t)/(lx+ly)/(lx*ly*t*(lx+ly-t)/(lx+ly)^3)^0.5}
    if(alternative=="greater"){
        pv1=phyper(A,lx,ly,t) ##Accurate p-value
        pv2=pnorm(z1)         ##Normal approximation p-value
        pv3=pnorm(z2)         ##Normal approximation p-value after continuity correction
    }else if(alternative=="less"){
        pv1=1-phyper(A,lx,ly,t)
        pv2=1-pnorm(z1)
        pv3=1-pnorm(z2)
    }else if(alternative=="both"){
        pv1=2*min(phyper(A,lx,ly,t),1-phyper(A,lx,ly,t))
        pv2=2*min(pnorm(z1),1-pnorm(z1))
        pv3=2*min(pnorm(z2),1-pnorm(z2))
    }
    conting.table=matrix(c(A,lx-A,lx,t-A,ly-(t-A),ly,t,lxy-t,lxy),3,3)
    col.name=c("X","Y","X+Y")
    row.name=c(">MXY","<MXY","TOTAL")
    dimnames(conting.table)=list(row.name,col.name)
    list(contingency.table=conting.table,p.value=pv1,pvnorm=pv2,pvnr=pv3)
}
```

```
a=c(698,68,675,656,655,648,640,639,620)
b=c(780,754,740,712,693,680,621)
```

```r
BM.test(a,b,alternative="greater")
```

```
## $contingency.table
##       X Y X+Y
## >MXY  2 6   8
## <MXY  7 1   8
## TOTAL 9 7  16
##
## $p.value
## [1] 0.02027972
##
## $pvnorm
## [1] 5.178433e-06
##
## $pvnr
## [1] 4.475372e-06
```

```r
BM.test(a,b,alternative="both")
```

```
## $contingency.table
##       X Y X+Y
## >MXY  2 6   8
## <MXY  7 1   8
## TOTAL 9 7  16
##
## $p.value
## [1] 0.04055944
##
## $pvnorm
## [1] 1.035687e-05
##
## $pvnr
## [1] 8.950744e-06
```

```r
women=c(28500,31000,22800,32350,30450,38200,34100,30150,33550,27350,
        25200,32050,26550,30650,35050,35600,26900,31350,28950,32900,
        31300,31350,35700,35900,35200,30450)
men=c(39700,33250,31800,38200,30800,32250,38050,34800,32750,38800,
      29900,37400,33700,36300,37250,33950,37750,36700,36100,26550,
      39200,41000,40400,35500)
BM.test(women,men,alternative="both")
```

```
## $contingency.table
##        X  Y X+Y
## >MXY   8 17  25
## <MXY  18  7  25
## TOTAL 26 24  50
##
## $p.value
## [1] 0.01012198
##
## $pvnorm
## [1] 3.61573e-13
##
## $pvnr
```

```
## [1] 3.467287e-13
```

```
BM.test(women,men,alternative="less")
```

```
## $contingency.table
##        X  Y X+Y
## >MXY   8 17  25
## <MXY  18  7  25
## TOTAL 26 24  50
##
## $p.value
## [1] 0.994939
##
## $pvnorm
## [1] 1
##
## $pvnr
## [1] 1
```

Mann-Whitney U Test & Wilcoxon Rank Sum Test

page 60 / page 75
As wilcoxon rank sum statistic is equivalent to Man-Whitney statistic, we only use one of them to test the hypothesis
Also called Wilcoxon-Mann-Whitney U test

```
weight=c(70,83,85,94,97,101,104,107,112,113,118,119,123,124,129,132,134,146,161)
group=c(1,2,1,1,2,1,2,2,1,2,1,2,2,2,2,1,2,2,2)
by(weight, group, median)
```

```
## group: 1
## [1] 101
## ------------------------------------------------------------
## group: 2
## [1] 121
```

```
wilcox.test(weight~group, exact=F,correct=F)
```

```
##
##  Wilcoxon rank sum test
##
## data:  weight by group
## W = 22, p-value = 0.09097
## alternative hypothesis: true location shift is not equal to 0
```

```
women=c(28500,31000,22800,32350,30450,38200,34100,30150,33550,27350,
        25200,32050,26550,30650,35050,35600,26900,31350,28950,32900,
        31300,31350,35700,35900,35200,30450)
men=c(39700,33250,31800,38200,30800,32250,38050,34800,32750,38800,
      29900,37400,33700,36300,37250,33950,37750,36700,36100,26550,
      39200,41000,40400,35500)
wilcox.test(women, men, exact = F, correct = F)
```

```
##
##  Wilcoxon rank sum test
##
```

```
## data:  women and men
## W = 127, p-value = 0.0003272
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(women, men, exact = F, correct = F, alternative = "less")
```

```
##
##  Wilcoxon rank sum test
##
## data:  women and men
## W = 127, p-value = 0.0001636
## alternative hypothesis: true location shift is less than 0
```

Ansari-Bradley Test

page 82

```r
manmade=c(4.5,6.5,7,10,12)
machine=c(6,7.2,8,9,9.8)

ansari.test(manmade, machine, alternative="two.sided")
```

```
##
##  Ansari-Bradley test
##
## data:  manmade and machine
## AB = 11, p-value = 0.1429
## alternative hypothesis: true ratio of scales is not equal to 1
```

```r
stock=c(1149,1152,1176,1149,1155,1169,1182,1160,1120,1171,
        1116,1130,1184,1194,1184,1147,1125,1125,1166,1151)
fmonth=factor(rep(0:1, c(10,10)), labels = c("nov", "dec"))
ansari.test(stock~fmonth, alternative="two.sided", exact=F)
```

```
##
##  Ansari-Bradley test
##
## data:  stock by fmonth
## AB = 68, p-value = 0.04749
## alternative hypothesis: true ratio of scales is not equal to 1
```

```r
weight=c(4.95,5.45,5.50,5.75,5.48,5.26,5.33,5.20,5.12,
         5.16,5.20,5.22,5.38,5.90,5.36,5.25,5.28,5.35,
         5.20,5.22,4.98,5.45,5.70,5.34,5.18,
         5.22,5.30,5.34,5.28,5.29,5.25,5.30,5.27,5.16,
         5.38,5.34,5.35,5.19,5.35,5.05,5.36,5.28,5.33,
         5.30,5.28,5.30,5.20)
fmachine=factor(rep(1:2, c(25,22)))
by(weight, fmachine, sd)
```

```
## fmachine: 1
## [1] 0.2211086
## --------------------------------------------------------
## fmachine: 2
## [1] 0.07681991
```

```r
library(coin)
```

```
## Loading required package: survival
```

```r
ansari_test(weight~fmachine, alternative="two.sided")
```

```
##
##  Asymptotic Two-Sample Ansari-Bradley Test
##
## data:  weight by fmachine (1, 2)
## Z = -3.2944, p-value = 0.0009864
## alternative hypothesis: true ratio of scales is not equal to 1
```

```r
ansari_test(weight~fmachine, alternative="g")
```

```
##
##  Asymptotic Two-Sample Ansari-Bradley Test
##
## data:  weight by fmachine (1, 2)
## Z = -3.2944, p-value = 0.0004932
## alternative hypothesis: true ratio of scales is greater than 1
```

Siegel-Tukey Test

page 83
See R-statistics blog
Notice: cannot compute exact p-value with ties

```r
source("https://raw.githubusercontent.com/talgalili/R-code-snippets/master/siegel.tukey.r")
weight=c(4.95,5.45,5.50,5.75,5.48,5.26,5.33,5.20,5.12,
         5.16,5.20,5.22,5.38,5.90,5.36,5.25,5.28,5.35,
         5.20,5.22,4.98,5.45,5.70,5.34,5.18,
         5.22,5.30,5.34,5.28,5.29,5.25,5.30,5.27,5.16,
         5.38,5.34,5.35,5.19,5.35,5.05,5.36,5.28,5.33,
         5.30,5.28,5.30,5.20)
machine=rep(0:1, c(25,22))
siegel.tukey(weight~machine)
```

```
##
## Median of group 1 = 5.28
## Median of group 2 = 5.295
##
## Testing median differences...
```

```
## Warning in wilcox.test.default(data$x[data$y == 0], data$x[data$y == 1]):
## cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data$x[data$y == 0] and data$x[data$y == 1]
## W = 291.5, p-value = 0.7327
## alternative hypothesis: true location shift is not equal to 0
##
## Performing Siegel-Tukey rank transformation...
##
```

```
##    sort.x sort.id unique.ranks
## 1    4.95       0      1.00000
## 2    4.98       0      4.00000
## 3    5.05       1      5.00000
## 4    5.12       0      8.00000
## 5    5.16       0     10.50000
## 6    5.16       1     10.50000
## 7    5.18       0     13.00000
## 8    5.19       1     16.00000
## 9    5.20       0     20.50000
## 10   5.20       0     20.50000
## 11   5.20       0     20.50000
## 12   5.20       1     20.50000
## 13   5.22       0     27.33333
## 14   5.22       0     27.33333
## 15   5.22       1     27.33333
## 16   5.25       0     32.50000
## 17   5.25       1     32.50000
## 18   5.26       0     36.00000
## 19   5.27       1     37.00000
## 20   5.28       0     42.50000
## 21   5.28       1     42.50000
## 22   5.28       1     42.50000
## 23   5.28       1     42.50000
## 24   5.29       1     47.00000
## 25   5.30       1     42.50000
## 26   5.30       1     42.50000
## 27   5.30       1     42.50000
## 28   5.30       1     42.50000
## 29   5.33       0     36.50000
## 30   5.33       1     36.50000
## 31   5.34       0     31.66667
## 32   5.34       1     31.66667
## 33   5.34       1     31.66667
## 34   5.35       0     25.33333
## 35   5.35       1     25.33333
## 36   5.35       1     25.33333
## 37   5.36       0     20.50000
## 38   5.36       1     20.50000
## 39   5.38       0     16.50000
## 40   5.38       1     16.50000
## 41   5.45       0     12.50000
## 42   5.45       0     12.50000
## 43   5.48       0     10.00000
## 44   5.50       0      7.00000
## 45   5.70       0      6.00000
## 46   5.75       0      3.00000
## 47   5.90       0      2.00000
##
## Performing Siegel-Tukey test...
##
## Mean rank of group 0: 17.88667
## Mean rank of group 1: 30.94697
##
```

```
##  Wilcoxon rank sum test with continuity correction
##
## data:  ranks0 and ranks1
## W = 122.5, p-value = 0.00114
## alternative hypothesis: true location shift is not equal to 0
```