刘程华 2018011687 计91

1、Use t test, Wilcoxon signed rank test, sign test to validate results of the 95% and 99% confidence intervals on ANSM p. 75.

Code :

```
library(DescTools)
x<-c(-2 ,4 ,8 ,25 ,-5 ,16 ,3 ,1 ,12 ,17 ,20 ,9)
x1<-c(-2 ,4 ,8 ,35 ,-5 ,16 ,3 ,1 ,12 ,17 ,20 ,9)
x2<-c(-2 ,4 ,8 ,65 ,-5 ,16 ,3 ,1 ,12 ,17 ,20 ,9)
t.test(x1)
t.test(x1,conf.level = 0.99)
wilcox.test(x1,conf.int = TRUE)
wilcox.test(x1,conf.int = TRUE,conf.level  = 0.99)
SignTest(x1)
SignTest(x1,conf.level = 0.99)

t.test(x2)
t.test(x2,conf.level = 0.99)
wilcox.test(x2,conf.int = TRUE)
wilcox.test(x1,conf.int = TRUE,conf.level  = 0.99)
SignTest(x2)
SignTest(x2,conf.level = 0.99)
```

Result :

```
> t.test(x1)

  One Sample t-test

data:  x1
t = 3.0707, df = 11, p-value = 0.01065
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  2.785007 16.881660
sample estimates:
mean of x
 9.833333

> t.test(x1,conf.level = 0.99)

  One Sample t-test

data:  x1
```

```
t = 3.0707, df = 11, p-value = 0.01065
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 -0.1125486 19.7792152
sample estimates:
mean of x
 9.833333

> wilcox.test(x1,conf.int = TRUE)

  Wilcoxon signed rank test

data:  x1
V = 71, p-value = 0.009277
alternative hypothesis: true location is not equal to 0
95 percent confidence interval:
  2.5 17.0
sample estimates:
(pseudo)median
             9

> wilcox.test(x1,conf.int = TRUE,conf.level  = 0.99)

  Wilcoxon signed rank test

data:  x1
V = 71, p-value = 0.009277
alternative hypothesis: true location is not equal to 0
99 percent confidence interval:
  0.5 20.0
sample estimates:
(pseudo)median
             9

> SignTest(x1)

  One-sample Sign-Test

data:  x1
S = 10, number of differences = 12, p-value = 0.03857
alternative hypothesis: true median is not equal to 0
96.1 percent confidence interval:
  1 17
sample estimates:
median of the differences
                      8.5

> SignTest(x1,conf.level = 0.99)

```

```
	One-sample Sign-Test

data:  x1
S = 10, number of differences = 12, p-value = 0.03857
alternative hypothesis: true median is not equal to 0
99.4 percent confidence interval:
 -2 20
sample estimates:
median of the differences 
                      8.5 

> 
> t.test(x2)

	One Sample t-test

data:  x2
t = 2.3331, df = 11, p-value = 0.03965
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  0.6983744 23.9682923
sample estimates:
mean of x 
 12.33333 

> t.test(x2,conf.level = 0.99)

	One Sample t-test

data:  x2
t = 2.3331, df = 11, p-value = 0.03965
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 -4.084738 28.751405
sample estimates:
mean of x 
 12.33333 

> wilcox.test(x2,conf.int = TRUE)

	Wilcoxon signed rank test

data:  x2
V = 71, p-value = 0.009277
alternative hypothesis: true location is not equal to 0
95 percent confidence interval:
  2.5 18.5
sample estimates:
(pseudo)median 
```

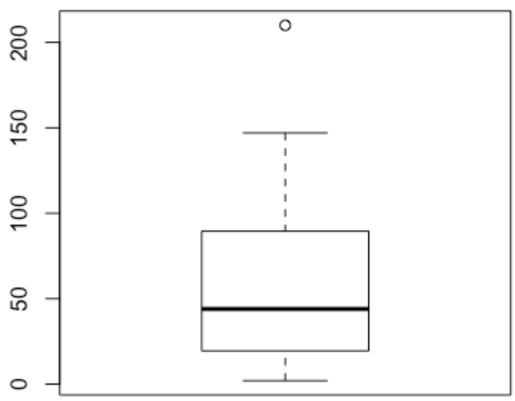
```
               9

> wilcox.test(x1,conf.int = TRUE,conf.level  = 0.99)

  Wilcoxon signed rank test

data:  x1
V = 71, p-value = 0.009277
alternative hypothesis: true location is not equal to 0
99 percent confidence interval:
  0.5 20.0
sample estimates:
(pseudo)median
             9

> SignTest(x2)

  One-sample Sign-Test

data:  x2
S = 10, number of differences = 12, p-value = 0.03857
alternative hypothesis: true median is not equal to 0
96.1 percent confidence interval:
  1 17
sample estimates:
median of the differences
                      8.5

> SignTest(x2,conf.level = 0.99)

  One-sample Sign-Test

data:  x2
S = 10, number of differences = 12, p-value = 0.03857
alternative hypothesis: true median is not equal to 0
99.4 percent confidence interval:
 -2 20
sample estimates:
median of the differences
                      8.5
```

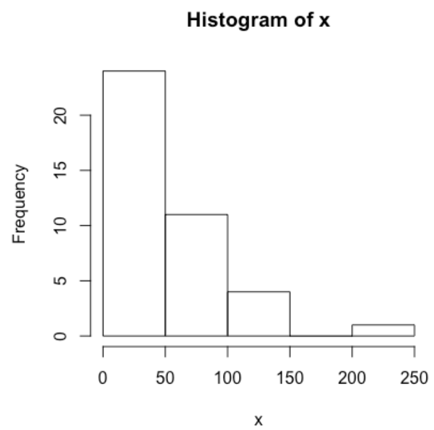
2、ANSM 3.16

A pathologist counts the numbers of diseased plants in randomly selected areas each 1 metre square on a large field. For 40 such areas the numbers of diseased plants are:

21	18	42	29	81	12	94	117	88	210
44	39	11	83	42	94	2	11	33	91
141	48	12	50	61	35	111	73	5	44
6	11	35	91	147	83	91	48	22	17

Use histograms and boxplots to decide whether there is evidence of skewness or outliers. Use nonparametric tests to find whether it is reasonable to assume the median number of diseased plants per square metre might be 50 (i) without assuming population symmetry, (ii) assuming population symmetry. For these data and the evidence provided by a boxplot and histograms do you consider the latter assumption reasonable?





firstthere exist an outlier and it is reasonable to assume the median number of diseased plants per square metre might be 50.secondly, assuming population symmetry is unreasonable.

Code :

```r
x<-c(21 , 18 , 42 , 29 , 81 , 12 , 94 , 117 , 88 , 210 ,
     44 , 39 , 11 , 83 , 42 , 94 , 2 , 11 , 33 , 91 ,
     141 , 48 , 12 , 50 , 61 , 35 , 111 , 73 , 5 , 44 ,
     6 , 11 , 35 , 91 , 147 , 83 , 91 , 48 , 22 , 17)
boxplot(x)
hist(x)
```

3、ANSM 4.8

The journal Biometrics $1985, 41$, p. 830, gives data on numbers of medical papers published annually in that journal for the period $1971-81$. These data are extended below to cover (in order) the period $1969-85$. Is there evidence of a monotonic trend in numbers of medical papers published?

$$11 \quad 6 \quad 14 \quad 13 \quad 18 \quad 14 \quad 11 \quad 22 \quad 19 \quad 19 \quad 25 \quad 24 \quad 38 \quad 19 \quad 25 \quad 31 \quad 19$$

```r
cox.stuart.test = function(x) {
  method = "Cox-Stuart test for trend analysis"
  leng = length(x)
  apross = round(leng)%%2
  if (apross == 1) {
    delete = (length(x) + 1)/2
    x = x[-delete]
  }
  half = length(x)/2
  x1 = x[1:half]
  x2 = x[(half + 1):(length(x))]
  difference = x1 - x2
  signs = sign(difference)
  signcorr = signs[signs != 0]
  pos = signs[signs > 0]
  neg = signs[signs < 0]
  if (length(pos) < length(neg)) {
    prop = pbinom(length(pos), length(signcorr), 0.5)
    names(prop) = "Increasing trend, p-value"
    rval <- list(method = method, statistic = prop)
    class(rval) = "htest"
    return(rval)
  } else {
    prop = pbinom(length(neg), length(signcorr), 0.5)
    names(prop) = "Decreasing trend, p-value"
    rval <- list(method = method, statistic = prop)
    class(rval) = "htest"
    return(rval)
  }
}
```

```
> x<-c(11 , 6 , 14 , 13 , 18 , 14 , 11 , 22 , 19 , 19 , 25 , 24 , 38 , 19 ,
25 , 31 , 19)
> cox.stuart.test(x)

  Cox-Stuart test for trend analysis

data:
Increasing trend, p-value = 0.035156

```

We think the numbers of medical papers published is in Increasing trend.

4、ANSM 4.11

A psychologist is testing 16 applicants for a job one at a time. Each has to perform a series of tests and the psychologist awards an overall point score to each applicant. A high score indicates a good performance. As each applicant may discuss the tests with later applicants before the latter are tested it is suggested that those tested later may have an unfair advantage. Do the applicants' scores (in order of testing) given below support this assertion?

$$62 \quad 69 \quad 55 \quad 71 \quad 64 \quad 68 \quad 72 \quad 75 \quad 49 \quad 74 \quad 81 \quad 83 \quad 77 \quad 79 \quad 89 \quad 42$$

Use an appropriate runs test. Do you consider the Cox-Stuart test (Section 3.2.3) may also be appropriate? Give reasons for your decision.

Code :

```
x<-c(62 , 69 , 55 , 71 , 64 , 68 , 72 , 75 , 49 , 74 , 81 , 83 , 77 , 79 ,
89 , 42)
cox.stuart.test(x)
l=length(x)
xs=sort(x)
x1=c()
x2=c()
t1=1
t2=1
for( i in 1:16){
  if(x[i]<=71){
    x1[t1]=x[i]
    t1=t1+1;
  }else{
    x2[t2]=x[i]
    t2=t2+1;
  }
}
cox.stuart.test(x1)
cox.stuart.test(x2)
```

Result :

```
> cox.stuart.test(x)

  Cox-Stuart test for trend analysis

data:
Increasing trend, p-value = 0.14453

> cox.stuart.test(x2)

  Cox-Stuart test for trend analysis

data:
Increasing trend, p-value = 0.0625

> cox.stuart.test(x1)

  Cox-Stuart test for trend analysis

data:
Decreasing trend, p-value = 0.3125

```

We divide the data into two parts and do Cox-Stuart test separately. We found that the results are quite different, Cox-Stuart test may be less powerful as monotonicity assumption is unlikely to be justified.

5、ANSM 5.6

One hundred general practitioners attend a health promotion workshop. At the start of the workshop they are asked to indicate whether they are in favour of routinely asking patients about alcohol consumption. They are then shown a video on the health and social problems caused by the excessive consumption of alcoholic drinks. The video is followed by discussion in small groups. After the video and discussion they are asked the original question again. Do the results given below indicate a significant change in attitudes as a result of the video and group discussion?

		Before video and discussion	
		In favour	Against
After video	In favour	41	27
and discussion	Against	16	58

```
> x<-matrix(c(41,27,16,58),2,2)
> mcnemar.test(x)

  McNemar's Chi-squared test with continuity
  correction

data:  x
McNemar's chi-squared = 2.3256, df = 1, p-value= 0.1273
```

No strong evidence of change in attitudes.