Lecture 3: One-sample problem II

Pratheepa Jeganathan

04/08/2019

Sign test (Fisher) - paired replicates data/one-sample data

Sign test

- $Z_1, \cdots Z_n$ random sample from a continuous population that has a common median θ.
  - If $Z_i \sim F_i$, $F_i(\theta) = F_i(Z_i \leq \theta) = F_i(Z_i > \theta) = 1 - F_i(\theta)$.
- Hypothesis testing:
  - $\mathrm{H}_0 : \theta = 0$ versus $\mathrm{H}_A : \theta \neq 0$.

Sign test (Cont.)

- Sign test statistic: $B = \sum_{i=1}^{n} \psi_i$.
- Motivation:
  - When θ is larger than 0, there will be larger number of positive Z_is $->$ big B value $->$ reject H_0 in favor of $\theta > 0$.
- Under H_0, $B \sim (n, 1/2)$
- Significance level α: probability of rejecting H_0 when it is true.
- Note
  - choices of α are limited to possible values of the $B \sim (n, 1/2)$ cdf.
  - compare the distribution of B under H_0 and the observed test statistic value.

Sign test (Cont.)

- Rejection regions
  - $\mathrm{H}_A : \theta > 0$, Reject H_0 if $B \geq b_{\alpha;n,1/2}$.
  - $\mathrm{H}_A : \theta < 0$, Reject H_0 if $B \leq n - b_{\alpha;n,1/2}$.
  - $\mathrm{H}_A : \theta \neq 0$, Reject H_0 if $B \geq b_{\alpha/2;n,1/2}$ or $B \leq n - b_{\alpha/2;n,1/2}$.

Large-Sample Approximation (Sign test)

- $B^* = \dfrac{B - \mathbb{E}_0(B)}{\mathbb{V}_0(B)^{1/2}} \sim \mathrm{N}(0,1) \quad \text{as} \quad n \to \infty$, where
- $\mathbb{E}_0(B) = \dfrac{n}{2}$ and $\mathbb{V}_0(B) = \dfrac{n}{4}$
- Rejection regions
  - $\mathrm{H}_A : \theta > 0$, Reject H_0 if $B^* \geq z_\alpha$.
  - $\mathrm{H}_A : \theta < 0$, Reject H_0 if $B^* \leq -z_\alpha$.
  - $\mathrm{H}_A : \theta \neq 0$, Reject H_0 if $B^* \geq z_{\alpha/2}$ or $B \leq -z_{\alpha/2}$.

Ties (Sign test)

- Discard zero Z values and redefine n.
- If too many zeros, choose alternative statistical procedure (Chapter 10)

Example (Sign test)

Example (HWC: Chapter 3, Example 3.5, pg. 65) - paired sample sign test

- Beak-Clapping Counts.
- Subjects: chick embryos.
- $X =$ average number of claps per minute during the dark period.
- $Y =$ average number of claps per minute during the period of illumination.
- Test responsivity of a (changes in the beak-clapping constituted a sensitive indicator of auditory responsiveness.) chick embryo to a light stimulus.
- $\mathrm{H}_A : \theta > 0$.

```r
df = data.frame(X = c(5.8,13.5,26.1,7.4,7.6,23,10.7,9.1,
  19.3,26.3,17.5,17.9,18.3,14.2,55.2,15.4,30,21.3,
  26.8,8.1,24.3, 21.3,18.2,22.5,31.1),
  Y = c(5,21,73,25,3,77,59,13,36,46,9,25,
    59,38,70,36,55,46,25,30,29,46,71,31,33))
head(df)
```

```
##      X  Y
## 1  5.8  5
## 2 13.5 21
## 3 26.1 73
## 4  7.4 25
## 5  7.6  3
## 6 23.0 77
```

```r
library(dplyr)
df = mutate(df, Z= Y-X, Psi = ifelse(Z > 0 ,1,0))
head(df)
```

```
##       X  Y    Z Psi
## 1  5.8  5 -0.8   0
## 2 13.5 21  7.5   1
## 3 26.1 73 46.9   1
## 4  7.4 25 17.6   1
## 5  7.6  3 -4.6   0
## 6 23.0 77 54.0   1
```

- lower.tail=F provides $P(B > b_{\alpha=.05}) = .05$

```
qbinom(p = .05, size = length(df$Psi),
  prob = 1/2, lower.tail = F)
```

```
## [1] 17
```

- We need $P(B \geq b) = .05$. Therefore, Reject H0 if $B \geq 18$.

However, the significance level is not .05.

```
1-pbinom((18-1), size = length(df$Psi),
  prob = 1/2, lower.tail = T)
```

```
## [1] 0.02164263
```

- Use the rejection region (Reject H0 if $B \geq 18$) to make the decision.
- Observed value of test statistic is

```
sum(df$Psi)
```

```
## [1] 21
```

- We reject in favor of $\theta > 0$ at the $\alpha = .05$ level.
- Didn't use actual Z_i.
- Actual magnitude of the Z_i's will be necessary for distribution-free point and interval estimates of θ associated with sign test.

Use build-in function `SIGN.test` in package `BSDA`

```r
library(BSDA)
SIGN.test(df$Y, df$X, alt = "greater")
```

```
##
##  Dependent-samples Sign-Test
##
## data:  df$Y and df$X
## S = 21, p-value = 0.0004553
## alternative hypothesis: true median difference is greate
## 95 percent confidence interval:
##  7.4519    Inf
## sample estimates:
## median of x-y
##          17.6
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level L.E.pt U.E.pt
## Lower Achieved CI     0.9461 7.5000    Inf
## Interpolated CI       0.9500 7.4519    Inf
## Upper Achieved CI     0.9784 7.1000    Inf
```

- P-value may also be found using the pbinom command: $P(B \geq 21) = 1 - P(B \leq 20)$.

```
1- pbinom((21-1), size = length(df$Psi),
  prob = 1/2, lower.tail = T)
```

```
## [1] 0.0004552603
```

- For the large-sample approximation, compute B^*.

```
B.star <- (21-25/2)/sqrt(25/4)
B.star
```

```
## [1] 3.4
```

- P-value

```
1-pnorm(B.star)
```

```
## [1] 0.0003369293
```

- Both the exact test and the large-sample approximation indicate that there is strong evidence that chick embryos are indeed responsive to a light stimulus, as measured by an increase in the frequency of beak-claps.
- To test $\mathrm{H}_0 = \theta_0$,
  - compute $Z_1^{'} = Z_1 - \theta_0, \cdots, Z_n^{'} = Z_n - \theta_0$.
  - do sign test on $Z_i^{'}$s.

Parametric t-test

- Let $Z_i \sim \mathrm{N}\left(\theta, \sigma^2\right)$.
- $\mathrm{H}_0 : \theta = 0$ versus $\mathrm{H}_A : \theta > 0$.
- Test statistic: $T = \dfrac{\bar{Z} - \theta}{s^2 / n}$.
- T is Studentized t-distribution with degrees of freedom $n - 1$.
- t_0: the observed value of test statistic.
- P-value: $P\left(T \geq t_0\right)$.

Wilcoxon signed rank test

Wilcoxon signed rank test

- Assumptions:
  - $Z_i = Y_i - X_i \sim F_i$, where F_i is symmetric about common median θ.
- Test statistic: $T^+ = \sum_{i=1}^{n} R_i \psi_i$, sum of positive signed ranks.
  - no-closed form distribution.
  - use iterative algorithms.
- Rejection regions
  - $\mathrm{H}_A : \theta > 0$, Reject H_0 if $T^+ \geq t_\alpha$.
  - $\mathrm{H}_A : \theta < 0$, Reject H_0 if $T^+ \leq \dfrac{n(n+1)}{2} - t_\alpha$.
  - $\mathrm{H}_A : \theta \neq 0$, Reject H_0 if $T^+ \geq t_{\alpha/2}$ or $T^+ \leq \dfrac{n(n+1)}{2} - t_{\alpha/2}$.

Large-sample approximation

- Read HWC page 41 -42 and comment 7 in page 48.

Ties

- Discard zero values among the Z_i's.
- If there are ties, assign each of the observations in a tied group the average of the integer ranks that are associated with the tied group.
  - not exact test

Theoretical distribution of T^+ for $n = 3$

- Comment 5 in page 46.
- Enumerate all 2^n possible outcomes for sample size three n=3:

```
library(gtools)
x <- c(0,1)
df <- permutations(n=2, r=3, v= x,
  repeats.allowed=T) %>% data.frame
df
```

```
##   X1 X2 X3
## 1  0  0  0
## 2  0  0  1
## 3  0  1  0
## 4  0  1  1
## 5  1  0  0
## 6  1  0  1
## 7  1  1  0
## 8  1  1  1
```

```r
T.plus = apply(df, 1,
  function(x){sum(x%*%seq(1,3))
    })
df = mutate(df, T.plus = T.plus)
df
```

```
##   X1 X2 X3 T.plus
## 1  0  0  0      0
## 2  0  0  1      3
## 3  0  1  0      2
## 4  0  1  1      5
## 5  1  0  0      1
## 6  1  0  1      4
## 7  1  1  0      3
## 8  1  1  1      6
```

```r
table(df$T.plus)/sum(table(df$T.plus))
```

```
##
##     0     1     2     3     4     5     6
## 0.125 0.125 0.125 0.250 0.125 0.125 0.125
```

Monte Carlo Simulation

- Compare Monte Carlo simulation results with the theoretical results:

```
n = 3; nsim = 10000; Z = matrix(rnorm(n*nsim),ncol=n)
T.plus.mc = apply(Z, 1,
  function(x) {sum(rank(abs(x)) * (x>0) )
    })
table(T.plus.mc)/nsim
```

```
## T.plus.mc
##      0      1      2      3      4      5      6
## 0.1299 0.1258 0.1250 0.2492 0.1257 0.1220 0.1224
```

Example (Wilcoxon signed rank test)

- Data are from nine patients who received tranquilizer.
- X (pre) factor IV value was obtained at the first patient visit after initiation of therapy.
- Y (post) factor IV value was obtained at the second visit after initiation of therapy.
- Test improvement due to tranquilizer that corresponds to a reduction in factor IV values.

```r
pre = c(1.83, .50, 1.62, 2.48, 1.68, 1.88,
  1.55, 3.06, 1.30)
post = c(.878, .647, .598, 2.05, 1.06, 1.29,
  1.06, 3.14, 1.29)
wilcox.test(post, pre, paired=TRUE,
  alterative = "less")
```

```
##
##  Wilcoxon signed rank test
##
## data:  post and pre
## V = 5, p-value = 0.03906
## alternative hypothesis: true location shift is not equal
```

```r
df <- data.frame(X= pre, Y = post)
df <- mutate(df, Z = Y-X, R=rank(abs(Z)),
  psi = ifelse(Z>0,1,0),Rpsi = R*psi)
df
```

```
##      X     Y      Z R psi Rpsi
## 1 1.83 0.878 -0.952 8   0    0
## 2 0.50 0.647  0.147 3   1    3
## 3 1.62 0.598 -1.022 9   0    0
## 4 2.48 2.050 -0.430 4   0    0
## 5 1.68 1.060 -0.620 7   0    0
## 6 1.88 1.290 -0.590 6   0    0
## 7 1.55 1.060 -0.490 5   0    0
## 8 3.06 3.140  0.080 2   1    2
## 9 1.30 1.290 -0.010 1   0    0
```

P-value is $P\left(T^{+} \leq 5\right)$

```
psignrank(q=sum(df$Rpsi),n=9,lower.tail = T)
```

```
## [1] 0.01953125
```

- There is strong evidence that tranquilizer does lead to patient improvement at $\alpha = .05$, as measured by a reduction in the Hamilton scale factor IV values.

Point and interval estimates

- All three tests (sign test, Wilcoxon signed rank, and t-test) have an associated estimate and confidence interval for the location parameter θ.
- Order statistic:
  - $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$.
  - $Z_{(1)}$ is the minimum.
  - $Z_{(n)}$ is the maximum.
- Quantile: equally spaced splitting points of continuous intervals with equal probabilities.

The point and interval estimate of θ associated with the sign rank statistic

- median: $\tilde{\theta} = \text{median}\{Z_i, i = 1, \cdots, n\}$.
- Let $Z_{(1)}, \cdots, Z_{(n)}$ denote the ordered Z_i and if
  - n is odd, $\tilde{\theta} = Z_{(k+1)}$, where $k = (n-1)/2$.
  - n is even, $\tilde{\theta} = \dfrac{Z_{(k)} + Z_{(k+1)}}{2}$, where $k = n/2$.
- $100(1-\alpha)\%$ confidence interval associated with two-sided test: $\left(Z_{(n+1-b_{\alpha/2;n,1/2})}, Z_{(b_{\alpha/2;n,1/2})}\right)$, $b_{\alpha/2;n,1/2}$ is the upper $\alpha/2$ percentile of the null distribution of B (sign test statistic).

The point and interval estimate of θ associated with the Wilcoxon signed rank statistic

- Hodges–Lehmann estimator:
  $\hat{\theta} = \text{median}\left\{\frac{Z_i + Z_j}{2}; i \leq j = 1, \cdots, n\right\}$.
  - Walsh averages $\frac{Z_i + Z_j}{2}; i \leq j = 1, \cdots, n$.
  - $M = \frac{n(n+1)}{2}$ Walsh averages.
  - $W_{(1)} \leq \cdots \leq W_{(M)}$ denote the ordered values of $(Z_i + Z_j)/2$.
- If
  - M is odd, $\hat{\theta} = W_{(k+1)}$, where $k = (M-1)/2$.
  - M is even, $\hat{\theta} = \frac{W_{(k)} + W_{(k+1)}}{2}$, where $k = M/2$.

- $100(1-\alpha)\%$ confidence interval associated with two-sided test: $\left(W_{\left(\frac{n(n+1)}{2}+1-t_{\alpha/2}\right)}, W_{(t_{\alpha/2})} \right)$, $t_{\alpha/2}$ is the upper $\alpha/2$ percentile of the null distribution of T^+.
- $t_{\alpha/2}$, the percentile points can be found using the R function `psignrank`.

Relaionship between Wilcoxon signed rank test statistic and Walsh averages (Tukey (1949))

- **HWC** page 57, comment 17.
- Wilcoxon test statistic: $T^+ = \sum_{i=1}^{n} R_i \psi_i$
- Number of Walsh averages greater than θ:
  $W^+ = \#\left\{ \frac{Z_i + Z_j}{2} > \theta \right\}$.
- Prove $T^+ = W^+$ by induction.
- Base of the Induction:
  - Assume that θ is greater than all $Z_1, \cdots, Z_n$.
    - then, θ is greater than all Walsh averages. Thus, $W^+ = 0$.
    - then, $Z_i - \theta$ are all negative. Thus, $T^+ = 0$.

Relaionship between Wilcoxon signed rank test statistic and Walsh averages

- Induction Steps:
  - Move θ to the left passing through $Z_1, \cdots, Z_n$ one and two at the time and show that
  - W^+ changes value when moving past an Walsh average by the same amount
  - T^+ changes value when
    - ranks of some $|Z_i - \theta|$ change or
    - sign of some rank change by the same amount
- See the complete proof here.

Comparison

- Power of a statistical test: the probability of rejecting the null hypothesis when it is false.
- The power of the sign test can be low relative to t-test.
- The power of signed-rank Wilcoxon test is nearly that of the t-test for normal distributions and generally greater than that of the t-test for distributions with heavier tails than the normal distribution.

Note: Read HWC page 71, comment 35 (power results for sign test)

Empirical power calculation $\theta = 0$

```
power.compute <- function(n = 30,
  df = 2,
  nsims = 1000,
  theta = 0){
  wil.sign.rank = rep(0, nsims)
  ttest = rep(0,nsims)
  Z = matrix((rt(n*nsims,df) + theta),
    ncol = n,nrow = nsims)
  wil.sign.rank = apply(Z, 1, function(x){
    wilcox.test(x)$p.value})
  ttest = apply(Z, 1, function(x){t.test(x)$p.value})
  pow.wil.sign.rank = mean(wil.sign.rank <=.05)
  pow.ttest = mean(ttest <=.05)
  rt = c(pow.wil.sign.rank, pow.ttest)
  names(rt) = c("Wilcoxon.signed.rank.power",
    "t.test.power")
  return(rt)
}
```

Empirical power calculation $\theta = 0$

```
power.compute.val = power.compute(n=30, df =2,
  nsims =1000, theta = 0)
power.compute.val
```

```
## Wilcoxon.signed.rank.power                     t.test.power
##                      0.059                            0.038
```

Empirical power calculation $\theta = 0.5$

```
power.compute.val = power.compute(n=30, df =2,
  nsims =1000, theta = 0.5)
power.compute.val
```

```
## Wilcoxon.signed.rank.power                        t.test.power
##                      0.453                               0.279
```

Empirical power calculation $\theta = 1$

```
power.compute.val = power.compute(n=30, df =2,
  nsims =1000, theta = 1)
power.compute.val
```

```
## Wilcoxon.signed.rank.power                  t.test.power
##                      0.920                         0.685
```

Summary

- Assumptions on F_i
  - Sign Test: any continuous distribution.
  - Signed-Rank Test: any symmetric continuous distribution.
  - t-test: any normal distribution.
- The continuity assumption assures that ties are impossible: With probability one we have $Z_i \neq Z_j$ when $i \neq j$.
- The continuity assumption is only necessary for exact hypothesis tests not for estimates and confidence intervals.

References

References for this lecture

HWC: Chapter 3.4-3.6, 3.8, 3.1-3.3, 3.7

KM: Chapter 2, page 21, Example 2.3.2. (empirical power)