STATS 205: Homework Assignment 1

Brian Liu

4/14/2019

Solution to Problem 1

First, let's check the value of `qbinom`:

```
qbinom(p=0.05, size=25, prob=1/2, lower.tail=F)
```

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

As expected, the resulting value is still 17, since nothing here was changed. Therefore,

$$\text{Reject } H_0 \text{ if } B \geq 18.$$

is still true.

Next, we will we model sign test with new Y_i and X_i:

```
# input values from Table 3.5
x_i <- c(5.8, 13.5, 26.1, 7.4, 7.6, 23.0, 10.7, 9.1, 19.3, 26.3, 17.5, 17.9, 18.3, 14.2, 55.2, 15.4, 30
# note the third value of y_i is 173 instead of 73
y_i <- c(5, 21, 173, 25, 3, 77, 59, 13, 36, 46, 9, 25, 59, 38, 70, 36, 55, 46, 25, 30, 29 ,46, 71, 31, 
```

```
library(BSDA) # required to run SIGN.test
```

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

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

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

```
SIGN.test(y_i, x_i, alt="greater")
```

```
##
##  Dependent-samples Sign-Test
##
## data:  y_i and x_i
## S = 21, p-value = 0.0004553
## alternative hypothesis: true median difference is greater than 0
## 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
```

$B^* = 3.40$ is still the same, since B (denoted S in the output) and n did not change.

Nothing from the original calculations (S [which is the same as B], p-value, median, etc.) changed, so we can still reject H_0 in favor of $\theta > 0$ at the $\alpha = 0.05$ level, since

$$B = 21 \geq 18$$

This seems to confirm that sign tests are relatively insensitive to outliers.

An example in which changing one observation has an effect on the final decision regarding rejection or acceptance of H_0 is if there were only 1 sample each for X_i and Y_i in which all were minus differences. Change one of them into a plus difference instead of a minus difference.

```r
qbinom(p=0.05, size=1, prob=1/2, lower.tail=F)
```

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

The resulting value is 1, which means

$$\text{Reject } H_0 \text{ if } B \geq 1.$$

Continuing with the Sign test,

```r
a <- c(1)
b <- c(2)
SIGN.test(a, b, alt="greater")
```

```
##
##  Dependent-samples Sign-Test
##
## data:  a and b
## S = 0, p-value = 1
## alternative hypothesis: true median difference is greater than 0
## 50 percent confidence interval:
##   -1 Inf
## sample estimates:
## median of x-y
##            -1
```

Here, $B = 0$ which is not greater than or equal to 1. Therefore **we cannot reject** H_0 at the $\alpha = 0.05$ level. However,

```r
b[1] = 0
SIGN.test(a, b, alt="greater")
```

```
##
##  Dependent-samples Sign-Test
##
## data:  a and b
## S = 1, p-value = 0.5
## alternative hypothesis: true median difference is greater than 0
## 50 percent confidence interval:
##    1 Inf
## sample estimates:
## median of x-y
##             1
```

Once we change `b[1]` to 0, $B = 1$, which is in fact greater than or equal to 1. **We can reject** H_0 at the $\alpha = 0.05$ level.

Solution to Problem 2

$$B^* = \frac{B - \frac{n}{2}}{(\frac{n}{4})^{\frac{1}{2}}}$$

$$B^* = \frac{8 - \frac{25}{2}}{(\frac{25}{4})^{\frac{1}{2}}} = -1.8$$

Now, I believe that at this point I am supposed to be able to figure out the p-value from this B^* calculation, and that that p-value could be different given that I could calculate B^* using H_0 vs. H_1, but I'm not sure how to do that.

Solution to Problem 3

```r
child_before = c(349, 400, 520, 490, 574, 427, 435)
child_after  = c(425, 533, 362, 628, 463, 427, 449)

wilcox.test(child_before, child_after, alternative="greater", paired=TRUE)
```

```
## Warning in wilcox.test.default(child_before, child_after, alternative =
## "greater", : cannot compute exact p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  child_before and child_after
## V = 9, p-value = 0.6625
## alternative hypothesis: true location shift is greater than 0
```

At the $\alpha = 0.05$ level, we **cannot reject** the null hypothesis, which is that hormone therapy has no increasing effect on heat-soluble hydroxyproline in the skin, since the p-value is $0.6625 > 0.05$.

Solution to Problem 4

Assume θ is referring to the difference between pre- and post-treatment:

```r
theta_values = child_before - child_after
theta_values
```

```
## [1]  -76 -133  158 -138  111    0  -14
```

```r
sort(theta_values)
```

```
## [1] -138 -133  -76  -14    0  111  158
```

The middle value is 14. Therefore, $\tilde{\theta} = 14$. For future reference, this page contains a Hodges-Lehmann function for R.

Solution to Problem 5

```r
SIGN.test(y_i, x_i, alt="greater")
```

```
##
##  Dependent-samples Sign-Test
##
## data:  y_i and x_i
## S = 21, p-value = 0.0004553
## alternative hypothesis: true median difference is greater than 0
## 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
```

According to this,

```
##                     Conf.Level L.E.pt U.E.pt
## Lower Achieved CI     0.9461 7.5000    Inf
```

the confidence interval is $(7.5000, Inf)$

Solution to Problem 6

```r
# owa(child_before, child_after)
```

Presumably, according to this link, if I didn't comment the line, the output is a list containing the ordered Walsh averages and **the value of the Hodges-Lehmann estimator, associated with the Wilcoxon signed rank test**. Unfortunately, I am not entirely sure what package installs it, and one package I tried installing ended up unable to install for some reason. I could try later.

If I figure out the output of this, I could compare $\hat{\theta}$ to $\hat{\theta}$.

However, I can do it manually like so, since $\hat{\theta}$ is calculated by getting the median value of the sums of adjacent differences if the number of sums is odd, or the average of the median two values:

```r
child_diff <- child_after - child_before
sort(child_diff)
```

```
## [1] -158 -111    0   14   76  133  138
```

The middle value is 14, which means $\hat{\theta} = 14$.

The output is the same: $\hat{\theta} = \tilde{\theta} = 14$ in this case.

Solution to Problem 7

```
depression_x = c(1.83, 0.50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30)
depression_y = c(0.878, 0.647, 0.598, 2.05, 1.06, 1.29, 1.06, 3.14, 1.29)
# owa(depression_x, depression_y)
```

Presumably, according to this, if I didn't comment the line, the output is a list containing **the ordered Walsh averages** and the value of the Hodges-Lehmann estimator, associated with the Wilcoxon signed rank test. Unfortunately, I am not entirely sure what package installs it, and one package I tried installing ended up unable to install for some reason. I could try later.

With the ordered Walsh averages, I could compare the number of positive Walsh averages with the number of positive differences, or T^+, which can be computed like so:

```
diff_depres = depression_y - depression_x
sign_diff_depres = sign(diff_depres)
sort(sign_diff_depres)
```

```
## [1] -1 -1 -1 -1 -1 -1 -1  1  1
```

```
sum(sign_diff_depres[sign_diff_depres > 0])
```

```
## [1] 2
```

The ordered Walsh averages can be computed like so:

```
#library(NSM3)
#owa(depression_x, depression_y)
```

Unfortunately, as mentioned before, the library won't install on my computer.

Solution to Problem 8

a

Since $\hat{\theta}$ is calculated by getting the median value of the sums of adjacent differences if the number of sums is odd, or the average of the median two values, when we add a number b to each of the sample values $Z_1, \ldots, Z_n$, $\hat{\theta}$ becomes $\hat{\theta} + b$.

b

Multiplying every difference by a number d would make it become $d\hat{\theta}$.

Solution to Problem 9

```
wilcox.test(child_before, child_after, alternative="greater", paired=TRUE, conf.int = TRUE, conf.level =
```

```
## Warning in wilcox.test.default(child_before, child_after, alternative =
## "greater", : cannot compute exact p-value with zeroes

## Warning in wilcox.test.default(child_before, child_after, alternative =
## "greater", : cannot compute exact confidence interval with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  child_before and child_after
## V = 9, p-value = 0.6625
## alternative hypothesis: true location shift is greater than 0
## 92.2 percent confidence interval:
##  -107  Inf
## sample estimates:
## (pseudo)median
##      -13.50001
```

According to the following,

```
## 92.2 percent confidence interval:
##  -107  Inf
```

The confidence interval at confidence coefficient 0.922 is $(-107, Inf)$.