

STATS 205: Homework Assignment 4

Brian Liu

6/3/2019

Solution to Problem 1

```
allergics = c(1651.0, 1112.0, 102.4, 100.0, 67.6, 65.9, 64.7, 39.6, 31.0)
nonallergics = c(48.1, 48.0, 45.5, 41.7, 35.4, 34.3, 32.4, 29.1, 27.3, 18.9, 6.6, 5.2, 4.7)
allergics; nonallergics
```

```
## [1] 1651.0 1112.0 102.4 100.0 67.6 65.9 64.7 39.6 31.0
```

```
## [1] 48.1 48.0 45.5 41.7 35.4 34.3 32.4 29.1 27.3 18.9 6.6 5.2 4.7
```

The null hypothesis is that allergic smokers have the same sputum histamine levels as nonallergic smokers. That is,

$$H_0 : p_a = p_n$$

The alternative hypothesis is that allergic smokers have higher sputum histamine levels than nonallergic smokers. That is,

$$H_0 : p_a > p_n$$

To test the null hypothesis against the alternative hypothesis, we will use the Mann-Whitney-Wilcoxin test, since the two samples are independent.

Two data samples are independent if they come from distinct populations and the samples do not affect each other. – Mann-Whitney-Wilcoxin Test

```
wilcox.test(x = allergics, y = nonallergics, alternative = "greater")
```

```
##
```

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

```
##
```

```
## data: allergics and nonallergics
```

```
## W = 106, p-value = 0.000386
```

```
## alternative hypothesis: true location shift is greater than 0
```

The p -value is 0.000386, which is significant at the $\alpha = 0.05$ level. There is strong evidence that allergic smokers have higher sputum histamine levels than nonallergic smokers.

Solution to Problem 2

Original problem statement

```
karate = c(37, 39, 30, 7, 13, 139, 45, 25, 16, 146, 94, 16, 23, 1, 290, 169, 62, 145, 36, 20, 13)
olympics = c(12, 44, 34, 14, 9, 19, 156, 23, 13, 11, 47, 26, 14, 33, 15, 62, 5, 8, 0, 154, 146)
karate; olympics
```

```
## [1] 37 39 30 7 13 139 45 25 16 146 94 16 23 1 290 169 62
```

```
## [18] 145 36 20 13
```

```
## [1] 12 44 34 14 9 19 156 23 13 11 47 26 14 33 15 62 5
## [18] 8 0 154 146
```

The null hypothesis is that children who viewed the violent TV take the same amount of time to seek help (were as tolerant) as the children who viewed the nonviolent sports-action TV. That is,

$$H_0 : t_k = t_o$$

The alternative hypothesis is that children who viewed the violent TV take longer to seek help (were more tolerant) than the children who viewed the nonviolent sports-action TV. That is,

$$H_0 : t_k > t_o$$

```
wilcox.test(x = karate, y = olympics, alternative = "greater")
```

```
## Warning in wilcox.test.default(x = karate, y = olympics, alternative =
## "greater"): cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
##
## data: karate and olympics
## W = 276.5, p-value = 0.08126
## alternative hypothesis: true location shift is greater than 0
```

The p -value is 0.08126, which is *not* significant at the $\alpha = 0.05$ level. There is *not enough* evidence that children who viewed the violent TV take longer to seek help (were more tolerant) than the children who viewed the nonviolent sports-action TV.

Solution to Problem 3

Let X be the nonallergics and Y be the allergics.

$$\delta = P(X < Y)$$

```
allergics = c(1651.0, 1112.0, 102.4, 100.0, 67.6, 65.9, 64.7, 39.6, 31.0)
nonallergics = c(48.1, 48.0, 45.5, 41.7, 35.4, 34.3, 32.4, 29.1, 27.3, 18.9, 6.6, 5.2, 4.7)
allergics; nonallergics
```

```
## [1] 1651.0 1112.0 102.4 100.0 67.6 65.9 64.7 39.6 31.0
## [1] 48.1 48.0 45.5 41.7 35.4 34.3 32.4 29.1 27.3 18.9 6.6 5.2 4.7
wilcox.test(x = allergics, y = nonallergics, conf.int=TRUE, conf.level=.90)

##
## Wilcoxon rank sum test
##
## data: allergics and nonallergics
## W = 106, p-value = 0.000772
## alternative hypothesis: true location shift is not equal to 0
## 90 percent confidence interval:
## 25.9 81.1
## sample estimates:
## difference in location
```

```
##                               54.3
wilcox.test(x = allergics, y = nonallergics, alternative = "less", conf.int=TRUE, conf.level=.90)

##
## Wilcoxon rank sum test
##
## data:  allergics and nonallergics
## W = 106, p-value = 0.9997
## alternative hypothesis: true location shift is less than 0
## 90 percent confidence interval:
##  -Inf    70
## sample estimates:
## difference in location
##                               54.3
```

$$\delta = P(X < Y) \simeq$$

The p -value stands for the probability that sample data like this would occur given that the null hypothesis (that allergic smokers have the same sputum histamine levels as nonallergic smokers) is true. Therefore,