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-Wilcoxon 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,

$$H_0 : t_k > t_o$$

```r
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.

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

```r
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
```

```r
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
```

The estimate for δ is

$$\hat{\delta} = P(X < Y) = 54.3$$

and the 90 confidence interval for δ is

$$\hat{\delta} = P(X < Y) = (25.9, 81.1)$$

Solution to Problem 4

```r
term = c(0.80, 0.83, 1.89, 1.04, 1.45, 1.38, 1.91, 1.64, 0.73, 1.46)
gest = c(1.15, 0.88, 0.90, 0.74, 1.21)

wilcox.test(x = term, y = gest, alternative = "greater")
```

```
##
##  Wilcoxon rank sum test
##
## data:  term and gest
## W = 35, p-value = 0.1272
## alternative hypothesis: true location shift is greater than 0
```

```r
# ks.test(x = term, y = gest, alternative = "greater")
# ks.test(x = term, y = gest)
```

The p-value for the Wilcoxon ranked test is 0.1272.

~~The p-value for the one-sided Two-sample Kolmogorov-Smirnov test is 0.9355, which is larger than the p-value for the Wilcoxon ranked test.~~

~~The p-value for the two-sided Two-sample Kolmogorov-Smirnov test is 0.1658, which is similar to that of the Wilcoxon ranked test.~~

```r
# install.packages("npsm", dependencies = TRUE, repos = "http://cran.us.r-project.org")
# install.packages("RVAideMemoire", dependencies = TRUE, repos = "http://cran.us.r-project.org")
# install.packages("robustrank", dependencies = TRUE, repos = "http://cran.us.r-project.org")
# fp.test(x = term, y = gest, alternative = 'two.sided')
# fp.test(x = term, y = gest)
```

Unfortunately, the package for `fp.test()` doesn't seem to be within reach, despite installing the following packages:

- npsm package
- RVAideMemoire package, with `fp.test()` demonstrated here
- R Documentation of npsm package including fp.test()
- robustrank package

Solution to Problem 5

Null hypothesis: "equal dispersions"

$$H_0 : p_{term} = p_{gest}$$

Alternative hypothesis: "the variation in tritiated water diffusion across human chorioamnion is different at term than at 12–26 weeks gestational age"

$$H_A : p_{term} \neq p_{gest}$$

```r
term = c(0.80, 0.83, 1.89, 1.04, 1.45, 1.38, 1.91, 1.64, 0.73, 1.46)
gest = c(1.15, 0.88, 0.90, 0.74, 1.21)

ansari.test(x = term, y = gest, alternative = "t")
```

```
##
##  Ansari-Bradley test
##
## data:  term and gest
## AB = 36, p-value = 0.1372
## alternative hypothesis: true ratio of scales is not equal to 1
```

The p-value is 0.1372, which is *not* significant at the $\alpha = 0.05$ level. There is *not enough* evidence that the variation in tritiated water diffusion across human chorioamnion is different at term than at 12–26 weeks gestational age.

Solution to Problem 6

```r
a = c(3.6, 2.6, 4.7, 8.0, 3.1, 8.8, 4.6, 5.8, 4.0, 4.6)
b = c(16.2, 17.4, 8.5, 15.6, 5.4, 9.8, 14.9, 16.6, 15.9, 5.3, 10.5)
```

The null hypothesis is that Type A subjects have the same Peak Levels of Human Plasma Growth Hormone after Arginine Hydrochloride Infusion as Type B subjects. That is,

$$H_0 : l_a = l_b$$

The alternative hypothesis is that Type A subjects have different Peak Levels of Human Plasma Growth Hormone after Arginine Hydrochloride Infusion as Type B subjects. That is,

$$H_0 : l_a > l_b$$

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

```r
wilcox.test(x = a, y = b, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(x = a, y = b, alternative = "two.sided"):
## cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  a and b
## W = 7, p-value = 0.0008201
## alternative hypothesis: true location shift is not equal to 0
```

The p-value is 0.0008201, which is significant at the $\alpha = 0.05$ level. There is *strong evidence* that Type A subjects have different Peak Levels of Human Plasma Growth Hormone after Arginine Hydrochloride Infusion as Type B subjects.

Solution to Problem 7

```r
Darwin.data = data.frame(pair = seq(1, 15), pot = c(rep(1, times=3), rep(2, times = 3), rep(3, times = 5

Darwin.data
```

```
##    pair pot cross.height self.height
## 1     1   1       23.500      17.375
## 2     2   1       12.000      20.375
## 3     3   1       21.000      20.000
## 4     4   2       22.000      20.000
## 5     5   2       19.125      18.375
## 6     6   2       21.500      18.625
## 7     7   3       22.125      18.625
## 8     8   3       20.375      15.250
## 9     9   3       18.250      16.500
## 10   10   3       21.625      18.000
## 11   11   3       23.250      16.250
## 12   12   4       21.000      18.000
## 13   13   4       22.125      12.750
## 14   14   4       23.000      15.500
## 15   15   4       12.000      18.000
```

```r
saveRDS(Darwin.data, "Darwin_data.rds"); Darwin.data
```

```
##    pair pot cross.height self.height
## 1     1   1       23.500      17.375
## 2     2   1       12.000      20.375
## 3     3   1       21.000      20.000
## 4     4   2       22.000      20.000
## 5     5   2       19.125      18.375
## 6     6   2       21.500      18.625
## 7     7   3       22.125      18.625
## 8     8   3       20.375      15.250
## 9     9   3       18.250      16.500
## 10   10   3       21.625      18.000
## 11   11   3       23.250      16.250
## 12   12   4       21.000      18.000
## 13   13   4       22.125      12.750
## 14   14   4       23.000      15.500
## 15   15   4       12.000      18.000
```