STATS 205: Homework Assignment 4 (Spring 2019)

4/29/2019

Solve problems 1 - 6 from the textbook *HWC* available here.

Send your Rmd and PDF files to pjeganat [at] stanford [dot] edu.

Due on 5/10/2019 at 1.30 p.m.

Friendly reminder: **Mid-term project proposal due on 5/3/2019 at 1.30 p.m.**

1) *HWC ** Page 133, Problem 1 (rank sum test - large-sample approximation)
2) *HWC ** Page 134, Problem 5 (rank sum test - Wilcoxon)
3) *HWC ** Page 141, Problem 21 (estimate δ and confidence interval for δ).
4) *HWC ** Page 149, Problem 41. (two sample location when unequal variances)
5) *HWC ** Page 168, Problem 1. (dispersion problem)
6) *HWC ** Page 198, Problem 33 (general distribution test)
7) (Hypothesis testing). This question is taken from **The design of experiment by Sir Ronald A. Fisher (8th edition, 1971). Chapter 3.** link here.

Darwin raised cross- and self-fertilized corn (Zea mays) plants. He planted equal numbers of each in four different pots, but not same number in each pot. Darwin measured heights of plants. Download and read `Darwin_data.rds` file data from Canvas @ Stanford.

```
Darwin.data = data.frame(pair = seq(1, 15),
  pot = c(rep(1, times=3), rep(2, times = 3),
    rep(3, times = 5), rep(4, times = 4)),
  cross.height = c(23.500, 12.000, 21.00, 22.000, 19.125,
    21.500, 22.125, 20.375, 18.250, 21.625,
    23.250, 21.000, 22.125, 23.000, 12.000),
  self.height = c(17.375, 20.375,20.000, 20.000,
    18.375, 18.625, 18.625, 15.250, 16.500,
    18.000, 16.250, 18.000, 12.750, 15.500, 18.000))
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
```

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

(i) Test that there is no difference between heights of crossed and self-fertilized plants using paired sample t-test. What are the assumptions for doing this test? [*Hint: use t.test()*]

(ii) Test that there is no difference between heights of crossed-fertilized and self-fertilized plants using permutation test (test statistic is sum of the difference). What are the assumptions for doing this test?

a) Let X be height of crossed-fertilized plants and Y be height of self-fertilized plants. Compute the diffrence $Z = X - Y$.

b) Test statistic is $T = \sum_{i=1}^{n=15} Z_i = \sum_{i=1}^{n=15} |Z_i|\delta_i$, where $\delta_i = 1$ or $\delta_i = -1$. Under H_0: there is no difference between heights of crossed and self-fertilized plants, δ_i is a fair coin flip $\{-1, 1\}$. There are $2^{15} = 32768$ ways to swap the plus and minus signs, all equaly likely under H_0. Use Monte Carlo method to approximate the exact p-value. [*Hint: Choose nperm = 10,000. Create a matrix of nperm* $\times$ *n. Sample using sample(c(-1,1), size = n, replace = TRUE) for each row. Multiply each row with absoulte value of* Z*. For each row, compute test statistic value. This is a two-sided test* H_A *: there is difference between heights of crossed-fertilized and self-fertilized plants. Thus, p-value* $= 2 \times P(T \geq t_0)$.]

(iii) Test that there is no difference between heights of crossed-fertilized and self-fertilized plants using Wilcoxon signed rank test. What are the assumptions for doing this test?

(iv) Test that there is no mean difference between heights of crossed-fertilized and self-fertilized plants using nonparamteric bootstrap method and studentized statistic. What are the assumptions for doing this test? [*Hint: Compute* $Z_i^{'} = Z_i - \bar{Z}$*. Compute studentized statistic* $\frac{mean}{\frac{standard\ deviation}{\sqrt{n}}}$ *for each bootstrap sample of* $Z^{'}$*. Compute p-value.*]