Assignment 15

Sachin Shubham

5/1/2021

1 Task

You are working with the results of the 2-way unreplicated block design experiment (Table 1). The data is balanced. The first column (count) represents the number of individuals captured. Year represents the year when the survey was conducted. The Month shows that for each year there were 5 survey events in each year. Each year is a "treatment" (or "group") and the month variable represents a "block". This is a common sort of experimental design; the blocks are set up to take care of any possible variation and to provide replication for the treatment. In this instance you wish to know if there is any significant difference due to the year.

1. Apply Friedman test. Comment on the results.
2. Apply post-hoc tests (pairwise.wilcox.test). Comment on the results.

Solution:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
```

```
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.0.5

##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
count<-c(2,48,40,3,120,81,2,16,36,7,21,17,2,14,17)
month<-c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5)
year<-c(2004,2005,2006,2004,2005,2006,2004,2005,2006,2004,2005,2006,2004,2005,2006)
df<-data.frame(count,month,year)
df
```

```
##    count month year
## 1      2     1 2004
## 2     48     1 2005
## 3     40     1 2006
## 4      3     2 2004
## 5    120     2 2005
## 6     81     2 2006
## 7      2     3 2004
## 8     16     3 2005
## 9     36     3 2006
## 10     7     4 2004
## 11    21     4 2005
## 12    17     4 2006
## 13     2     5 2004
## 14    14     5 2005
## 15    17     5 2006
```

```
attach(df)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     count, month, year
```

```
friedman.test(count, groups = year, blocks = month)
```

```
##
##  Friedman rank sum test
##
## data:  count, year and month
## Friedman chi-squared = 7.6, df = 2, p-value = 0.02237
```

```
friedman.test(count ~ year | month, data= df)
```

```
##
##  Friedman rank sum test
##
## data:  count and year and month
## Friedman chi-squared = 7.6, df = 2, p-value = 0.02237
```

```r
wb <- aggregate(df$count,
                by = list(w = df$year,
                          t = df$month),
                FUN = mean)
wb
```

```
##       w t   x
## 1  2004 1   2
## 2  2005 1  48
## 3  2006 1  40
## 4  2004 2   3
## 5  2005 2 120
## 6  2006 2  81
## 7  2004 3   2
## 8  2005 3  16
## 9  2006 3  36
## 10 2004 4   7
## 11 2005 4  21
## 12 2006 4  17
## 13 2004 5   2
## 14 2005 5  14
## 15 2006 5  17
```

```r
friedman.test(wb$x, wb$w, wb$t)
```

```
##
##  Friedman rank sum test
##
## data:  wb$x, wb$w and wb$t
## Friedman chi-squared = 7.6, df = 2, p-value = 0.02237
```

```r
friedman.test(x ~ w | t, data = wb)
```

```
##
##  Friedman rank sum test
##
## data:  x and w and t
## Friedman chi-squared = 7.6, df = 2, p-value = 0.02237
```

```r
p<-pairwise.wilcox.test(df$count, g = df$year, exact=FALSE)
print(p)
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  df$count and df$year
##
##      2004  2005
## 2005 0.033 -
## 2006 0.033 0.834
##
## P value adjustment method: holm
```

CONCLUSION:

The 2005 and 2006 years data do not appear to be significant but 2004 data appears to be weakly statistically significant to 2005 and 2006 with a p-value of 0.033. I would normally divide alpha threshold over the number of categories/channels in each pairing to set a stronger threshold that the Wilcoxon test must pass. In this case alpha = 0.05/3 categories = 0.0167. Overall, I would be hesitant to call this statistically significant even though it falls under the alpha=0.05 threshold using a Wilcoxon test.