Class Notes: STAT 501

Nonparametrics & Log-Linear Models

Friedman Rank Test

Da Kuang

University of Pennsylvania

Contents

1 Friedman Rank Test

1.1 Background

1.1.1 Two-Way Layout

In this chapter, we focus on an experimental design involving two factors, each at two or more levels. Out primary interest is in the relative location effects (medians) of the different levels of one of these factors, hereafter called the **treatment factors**, within the various levels of the second factors, hereafter called the **blocking factor**.

The blocking factor is associated quite commonly with the experimental design where subjects are first divided into more homogeneous subgroups (called blocks, e.g. schools or hospitals) and then randomly assigned to the various treatment levels within these blocks. Such a design is called a **randomized block design**.

The basic null hypothesis of interest is that of no differences in the location effects (medians) of the k treatments within each of the blocks.

For Friedman Rank test, we focus on the case of one observation per treatment-block cell, commonly known as a **randomized complete block design**.

The data can be described as follows:

- One observation per treatment-block combination.
- Within each block, there are I observations.
- Within each treatment group, there are J observations.

	Treatments			
Blocks	1	2	$\cdots$	I
1	$Y_{1,1}$	$Y_{2,1}$	$\cdots$	$Y_{I,1}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
J	$Y_{1,J}$	$Y_{2,J}$	$\cdots$	$Y_{I,J}$

Figure 1: Randomized Complete Block Design

1.2 Assumption

- The random variables $Y_{i,j}$'s all mutually independent
- For each fixed (i, j), the random variable $Y_{i,j}$ follows continuous distribution F_{ij}.
- The distribution functions $F_{11}, \ldots, F_{1k}, \ldots, F_{n1}, \ldots, F_{nk}$ are connected through the relationship

$$F_{ij}(u) = F(u - \tau_i - \beta_j), -\infty < u < \infty$$

, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where

- F is a continuous distribution function with unknown median θ.
- τ_i is the unknown additive treatment effect contributed by the i-th treatment.
- β_j is the unknown additive effect contributed by block j.

We can set up the model of each observation based on the assumptions.

$$Y_{ik} = \theta + \tau_i + \beta_j + \epsilon_{ij}$$

- θ: Overall median.
- τ_i: the i-th treatment effect.
- β_j: the j-th block effect.
- ϵ_{ij}: random error. iid from a continuous distribution with median 0.

1.3   Hypothesis

- H_0: $\tau_1 = \cdots = \tau_I$, which means that $F_1 = \cdots F_I = F$.
- H_1: $\tau_1, \ldots, \tau_I$ are not all equal.

1.3.1   Rank Based Test Statistic, S

To compute the Friedman statistic S, first order the I obervations from least to greatest separately within each of the J blocks.

Let r_{ij} denotes the rank of Y_{ij} in this joint ranking of the j-th block, $i = 1, \ldots, I$.

Let $R_i = \sum_{j=1}^{J} r_{ij}$, which is the sum of the within-blocks ranks received by the obervations in the i-th treatment.

Let $\bar{R}_i = \frac{R_i}{j}$, $i = 1, \ldots, I$, which is the **average within-blocks rank** received by the observations in the i-th treatment. Note that the overall average rank is $(I + 1)/2$.

| Blocks | Treatments | | | |
	1	2	$\cdots$	I
1	r_{11}	r_{21}	$\cdots$	r_{I1}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
J	r_{1J}	r_{2J}	$\cdots$	r_{IJ}

Figure 2: Rank Schema

Blocks	Treatments		
	1	2	3
1	5.4 (1)	5.5 (2)	5.55 (3)
2	5.85 (3)	5.7 (1)	5.75 (2)
3	5.2 (1)	5.6 (3)	5.5 (2)
	$R_1 = 5$	$R_2 = 6$	$R_3 = 7$

Figure 3: Rank Example

The Friedman statistic S is given by

$$S = (\frac{12}{I(I+1)J} \sum_{i=1}^{I} R_i^2) - 3J(I+1) \tag{1}$$

$$= \frac{12J}{I(I+1)} \sum_{i=1}^{I} (\bar{R}_i - \frac{I+1}{2})^2 \tag{2}$$

Reject H_0 if $S > s_\alpha$, the upper α percentile for the null distribution of S.

```
1  library(NSM3)
2  ?cFrd
```

We can also use CTL to estimate the critical value.

Reject H_0 if $S \geq \chi^2_{I-1,\alpha}$, as $J \to \infty$. $\chi^2_{I-1,\alpha}$ is the upper α percentile point of a chi-square distribution with $I - 1$ degree of freedom.

1.4 Special Case: I = 2

When $I = 2$, the data looks like paired-sample problem. Recall that **Wilcoxon's Rank Sum Test** is for location changes in paired-sample problem.

Figure 4: Special Case, $I = 2$

In the following code, we use sign test, Friedman Rank Sum Test (CTL), and wilcox test to exam the difference of location between X and Y. We have the p-values,

- Sign Test: 0.1797
- Friedman Rank Sum Test (CTL): 0.0955807
- Friedman Rank Sum Test: 0.09558
- Wilcoxon signed rank exact test: 0.03906

We can tell that

```
1  library(BSDA)
2  x=c(1.83,0.50,1.62,2.48,1.68,1.88,1.55,3.06,1.30);
3  y=c(0.878,0.647,0.598,2.05,1.06,1.29,1.06,3.14,1.29);
4  SIGN.test(x,y);
5  # Dependent-samples Sign-Test
6  #
7  # data: x and y
8  # S = 7, p-value = 0.1797
9  # alternative hypothesis: true median difference is not equal to 0
10 # 95 percent confidence interval:
11 # -0.0730000 0.9261778
12 # sample estimates:
```

```
13  # median of x-y
14  # 0.49
15  #
16  # Achieved and Interpolated Confidence Intervals:
17  #
18  # Conf.Level L.E.pt U.E.pt
19  # Lower Achieved CI 0.8203 0.010 0.6200
20  # Interpolated CI 0.9500 -0.073 0.9262
21  # Upper Achieved CI 0.9609 -0.080 0.9520
22
23  n=length(x)
24  test.stat=(sum(x>y)-n/2)/sqrt(n/4)
25  t2=test.stat^2
26  c(2*pnorm(test.stat,lower.tail=F),t2,
27    pchisq(t2,1,lower.tail=F))
28  # [1] 0.0955807 2.7777778 0.0955807
29  da=cbind(z=c(x,y),gr=c(rep(1,n),rep(2,n)),
30          blk=rep(1:n,2))
31  da
32  # z gr blk
33  # [1,] 1.830 1 1
34  # [2,] 0.500 1 2
35  # [3,] 1.620 1 3
36  # [4,] 2.480 1 4
37  # [5,] 1.680 1 5
38  # [6,] 1.880 1 6
39  # [7,] 1.550 1 7
40  # [8,] 3.060 1 8
41  # [9,] 1.300 1 9
42  # [10,] 0.878 2 1
43  # [11,] 0.647 2 2
44  # [12,] 0.598 2 3
```

```
45  # [13,] 2.050 2 4
46  # [14,] 1.060 2 5
47  # [15,] 1.290 2 6
48  # [16,] 1.060 2 7
49  # [17,] 3.140 2 8
50  # [18,] 1.290 2 9
51  friedman.test(z~gr|blk,data=da)
52  # Friedman rank sum test
53  #
54  # data: z and gr and blk
55  # Friedman chi-squared = 2.7778, df = 1, p-value = 0.09558
56  wilcox.test(x,y,paired=T)
57  #  Wilcoxon signed rank exact test
58
59  # data: x and y
60  # V = 40, p-value = 0.03906
61  # alternative hypothesis: true location shift is not equal to 0
```

1.5 Multiple Comparison

In a Friedman Test, if we reject the null hypothesis then we admit that there is treatment effect difference. Then we usually want to know that which one makes it different. So we want to xompare the $I(I-1)/2$ pairwise group differences.

- Calculate $|R_u - R_v|$, $1 \leq u < v \leq I$.
- Decide $\tau_u \neq \tau_v$ if $|R_u - R_v| \geq r_\alpha$, otherwise decide $\tau_u = \tau_v$.

```
1  warpbreaks
2  wb= aggregate(warpbreaks$breaks,
3     by = list(gr = warpbreaks$wool,
4     bk = warpbreaks$tension),FUN = mean)
```

```r
 5  wb
 6  # gr bk x
 7  # 1 A L 44.55556
 8  # 2 B L 28.22222
 9  # 3 A M 24.00000
10  # 4 B M 28.77778
11  # 5 A H 24.55556
12  # 6 B H 18.77778
13  friedman.test(wb$x, wb$gr, wb$bk)
14  friedman.test(x ~ gr| bk, data = wb)
15  # Friedman rank sum test
16
17  # data: wb$x, wb$gr and wb$bk
18  # Friedman chi-squared = 0.33333, df = 1, p-value = 0.5637
19  install.packages("agricolae")
20  library(agricolae)
21  # friedman(wb$bk,wb$gr,wb$x,console=T)
22  # Study: wb$x ~ wb$bk + wb$gr
23
24  # wb$gr, Sum of the ranks
25
26  # wb.x r
27  # A 5 3
28  # B 4 3
29
30  # Friedman's Test
31  # ===============
32  # Adjusted for ties
33  # Critical Value: 0.3333333
34  # P.Value Chisq: 0.5637029
35  # F Value: 0.25
36  # P.Value F: 0.6666667
```

```
37
38  # Post Hoc Analysis
39
40  # Alpha: 0.05 ; DF Error: 2
41  # t-Student: 4.302653
42  # LSD: 8.605305
43
44  # Treatments with the same letter are not significantly different.
45
46  # Sum of ranks groups
47  # A 5 a
48  # B 4 a
```

1.6   Extension: Mack-Skillings Test

There can be more than one observation for some of the treatment-block combinations, i.e. replications in a given cell. To be more specific, we can have

- Some cells with no observation.
- Some cells with one observation.
- Some cells with replications.

The data can be represented with I treatments and J blocks. n_{ij} subjects in the (i, j)-th cell. $N = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$ observations. The Outcome $Y_{i,j,k}$, $k = 1, \ldots, n_{ij}$.

	Treatments			
Blocks	1	2	$\cdots$	I
	$Y_{1,1,1}$	$Y_{2,1,1}$	$\cdots$	$Y_{I,1,1}$
1	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	$Y_{1,1,n_{11}}$	$Y_{2,1,n_{21}}$	$\cdots$	$Y_{I,1,n_{I1}}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	$Y_{1,J,1}$	$Y_{2,J,1}$	$\cdots$	$Y_{I,J,1}$
J	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	$Y_{1,J,n_{1J}}$	$Y_{2,J,n_{2J}}$	$\cdots$	$Y_{I,J,n_{IJ}}$

Figure 5: Data Strucutre

The assumptions

- The random variables $Y_{i,j,k}$'s all mutually independent
- For each fixed (i,j), the random variable $Y_{i,j,1}, \ldots, Y_{i,j,n_{ij}}$ follow continuous distribution F_{ij}.
- The distribution functions $F_{11}, \ldots, F_{1k}, \ldots, F_{n1}, \ldots, F_{nk}$ are connected through the relationship

$$F_{ij}(u) = F(u - \tau_i - \beta_j), -\infty < u < \infty$$

, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where

- F is a continuous distribution function with unknown median θ.
- τ_i is the unknown additive treatment effect contributed by the i-th treatment.
- β_j is the unknown additive effect contributed by block j.

We can set up the model of each observation based on the assumptions.

$$Y_{ik} = \theta + \tau_i + \beta_j + \epsilon_{ijk}$$

- θ: Overall median.
- τ_i: the i-th treatment effect.

- β_j: the j-th block effect.
- ϵ_{ijk}: random error. iid from a continuous distribution with median 0.

The hypothesis are

- H_0: $\tau_1 = \cdots = \tau_I$, which means that $F_{1j} = \cdots = F_{Ij} = F_j$, $j = 1, \ldots, J$.
- H_1: $\tau_1, \ldots, \tau_I$ are not all equal.

Example code

```
1  ## Skillings and Mack (1981), Table 1 page 173
2  ## Comparison of four methods of assembling a product
3  B <- rep(c(1,2,3,4,5,6,7,8,9),rep(4,9))
4  G <- rep(c('A','B','C','D'),9)
5  y <- c(3.2,4.1,3.8,4.2, 3.1,3.9,3.4,4.0, 4.3,3.5,4.6,4.8,
6  3.5,3.6,3.9,4.0, 3.6,4.2,3.7,3.9, 4.5,4.7,3.7, NA,
7  NA ,4.2,3.4,NA , 4.3,4.6,4.4,4.9, 3.5, NA,3.7, 3.9)
8  Ski.Mack(y,groups = G,blocks = B)
9  # Skillings-Mack Statistic = 15.493049 , p-value = 0.001440
10 # Note: the p-value is based on the chi-squared distribution with d.f. =
       3
11
12 # $Nblocks
13 # [1] 9
14
15 # $Ntreatments
16 # [1] 4
17
18 # $rawdata
19 # [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
20 # [1,] 3.2 3.1 4.3 3.5 3.6 4.5 NA 4.3 3.5
21 # [2,] 4.1 3.9 3.5 3.6 4.2 4.7 4.2 4.6 NA
22 # [3,] 3.8 3.4 4.6 3.9 3.7 3.7 3.4 4.4 3.7
```

```
23  # [4,]  4.2 4.0 4.8 4.0 3.9 NA NA 4.9 3.9
24
25  # $rankdata
26  #     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
27  # [1,]  1   1   2   1   1   2   1.5  1   1
28  # [2,]  3   3   1   2   4   3   2.0  3   2
29  # [3,]  2   2   3   3   2   1   1.0  2   2
30  # [4,]  4   4   4   4   3   2   1.5  4   3
31
32  # $varCovarMatrix
33  #     [,1] [,2] [,3] [,4]
34  # [1,]  22  -7  -8  -7
35  # [2,]  -7  21  -8  -6
36  # [3,]  -8  -8  23  -7
37  # [4,]  -7  -6  -7  20
38
39  # $adjustedSum
40  #     [,1]       [,2]        [,3]        [,4]
41  # [1,] -14.12559752 4.281244146 -4.281244146 14.12559752
42  ## y is a matrix
43
44  maty <- matrix(
45  c(3.2,4.1,3.8,4.2,3.1,3.9,3.4,4.0, 4.3,3.5,4.6,4.8,
46  3.5,3.6,3.9,4.0, 3.6,4.2,3.7,3.9, 4.5,4.7,3.7, NA,
47  NA ,4.2,3.4,NA , 4.3,4.6,4.4,4.9, 3.5, NA,3.7, 3.9),
48  ncol=9,byrow=FALSE)
49  Ski.Mack(maty, simulate.p.value = TRUE, B = 1000)
50  # Skillings-Mack Statistic = 15.493049 , p-value = 0.001440
51  # Note: the p-value is based on the chi-squared distribution with d.f. =
        3
52  # Based on B = 1000 , Simulated p-value = 0.000000
53
```

```
54  # $Nblocks
55  # [1] 9
56
57  # $Ntreatments
58  # [1] 4
59
60  # $rawdata
61  # [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
62  # [1,] 3.2 3.1 4.3 3.5 3.6 4.5 NA 4.3 3.5
63  # [2,] 4.1 3.9 3.5 3.6 4.2 4.7 4.2 4.6 NA
64  # [3,] 3.8 3.4 4.6 3.9 3.7 3.7 3.4 4.4 3.7
65  # [4,] 4.2 4.0 4.8 4.0 3.9 NA NA 4.9 3.9
66
67  # $rankdata
68  # [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
69  # [1,] 1 1 2 1 1 2 1.5 1 1
70  # [2,] 3 3 1 2 4 3 2.0 3 2
71  # [3,] 2 2 3 3 2 1 1.0 2 2
72  # [4,] 4 4 4 4 3 2 1.5 4 3
73
74  # $varCovarMatrix
75  # [,1] [,2] [,3] [,4]
76  # [1,] 22 -7 -8 -7
77  # [2,] -7 21 -8 -6
78  # [3,] -8 -8 23 -7
79  # [4,] -7 -6 -7 20
80
81  # $adjustedSum
82  # [,1] [,2] [,3] [,4]
83  # [1,] -14.12559752 4.281244146 -4.281244146 14.12559752
```