Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 24, 2021

2021 Spring

Chapter 14. Nonparametric Statistics

§ 14.1 Introduction

§ 14.2 The Sign Test

§ 14.3 Wilcoxon Tests

§ 14.4 The Kruskal-Wallis Test

§ 14.5 The Friedman Test

§ 14.6 Testing for Randomness

Plan

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test

What is the nonparametric counterpart for the two-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Assume that $n_1 = \cdots = n_k$.

Samples can be further partitioned into b blocks.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Friedman Test Statistic:

Reject H_0 at the α level if

$$G = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_{.j}^2 - 3b(k+1) \geq \chi^2_{1-\alpha,k-1}.$$

where $R_{.j}$ is the within-block ranks.

The Friedman Test Statistic:

Reject H_0 at the α level if

$$G = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_{.j}^2 - 3b(k+1) \geq \chi^2_{1-\alpha,k-1}.$$

where $R_{.j}$ is the within-block ranks.

The Friedman Test Statistic:

Reject H_0 at the α level if

$$G = \frac{12}{bk(k+1)} \sum_{j=1}^{k} R_{.j}^2 - 3b(k+1) \geq \chi^2_{1-\alpha,k-1}.$$

where $R_{.j}$ is the within-block ranks.

E.g. Baseball ...

$$\text{Test if } H_0 : \widetilde{\mu}_{\text{Narrow}} = \widetilde{\mu}_{\text{Wide}} \text{ at } \alpha = 0.01$$

E.g. Baseball ...

$$\text{Test if } H_0 : \widetilde{\mu}_{\text{Narrow}} = \widetilde{\mu}_{\text{Wide}} \text{ at } \alpha = 0.01$$

Table 14.5.1 Times (sec) Required to Round First Base

Player	Narrow-Angle	Rank	Wide-Angle	Rank
1	5.50	1	5.55	2
2	5.70	1	5.75	2
3	5.60	2	5.50	1
4	5.50	2	5.40	1
5	5.85	2	5.70	1
6	5.55	1	5.60	2
7	5.40	2	5.35	1
8	5.50	2	5.35	1
9	5.15	2	5.00	1
10	5.80	2	5.70	1
11	5.20	2	5.10	1
12	5.55	2	5.45	1
13	5.35	1	5.45	2
14	5.00	2	4.95	1
15	5.50	2	5.40	1
16	5.55	2	5.50	1
17	5.55	2	5.35	1
18	5.50	1	5.55	2
19	5.45	2	5.25	1
20	5.60	2	5.40	1
21	5.65	2	5.55	1
22	6.30	2	6.25	1
		39		27

Sol. $k = 2$, $b = 22$

Compute the rank within each block (see the previous table)

Compute the g statistic:

$$g = \frac{12}{22 \times 2 \times (2+1)} \left[39^2 + 27^2\right] - 3 \times 22 \times (2+1) = \frac{72}{11} \approx 6.54.$$

Critical region is

$$C = \left\{g : g \geq \chi^2_{0.95,1} = 3.84\right\}.$$

The p-value is

$$\mathbb{P}\left(\chi^2_1 \geq \frac{72}{11}\right) = 0.01051525.$$

Conclusion: Reject. ∎

Sol. $k = 2$, $b = 22$

Compute the rank within each block (see the previous table)

Compute the g statistic:

$$g = \frac{12}{22 \times 2 \times (2+1)} \left[39^2 + 27^2\right] - 3 \times 22 \times (2+1) = \frac{72}{11} \approx 6.54.$$

Critical region is

$$C = \left\{g : g \geq \chi^2_{0.95,1} = 3.84\right\}.$$

The p-value is

$$\mathbb{P}\left(\chi^2_1 \geq \frac{72}{11}\right) = 0.01051525.$$

Conclusion: Reject. ∎

Sol. $k = 2$, $b = 22$

Compute the rank within each block (see the previous table)

Compute the g statistic:

$$g = \frac{12}{22 \times 2 \times (2+1)} \left[39^2 + 27^2\right] - 3 \times 22 \times (2+1) = \frac{72}{11} \approx 6.54.$$

Critical region is

$$C = \left\{g : g \geq \chi^2_{0.95,1} = 3.84\right\}.$$

The p-value is

$$\mathbb{P}\left(\chi^2_1 \geq \frac{72}{11}\right) = 0.01051525.$$

Conclusion: Reject. ∎

Sol. $k = 2$, $b = 22$

Compute the rank within each block (see the previous table)

Compute the g statistic:

$$g = \frac{12}{22 \times 2 \times (2+1)} \left[39^2 + 27^2\right] - 3 \times 22 \times (2+1) = \frac{72}{11} \approx 6.54.$$

Critical region is

$$C = \left\{g : g \geq \chi^2_{0.95,1} = 3.84\right\}.$$

The p-value is

$$\mathbb{P}\left(\chi^2_1 \geq \frac{72}{11}\right) = 0.01051525.$$

Conclusion: Reject. ∎

Sol. $k = 2$, $b = 22$

Compute the rank within each block (see the previous table)

Compute the g statistic:

$$g = \frac{12}{22 \times 2 \times (2+1)} \left[39^2 + 27^2\right] - 3 \times 22 \times (2+1) = \frac{72}{11} \approx 6.54.$$

Critical region is

$$C = \left\{g : g \geq \chi^2_{0.95,1} = 3.84\right\}.$$

The p-value is

$$\mathbb{P}\left(\chi^2_1 \geq \frac{72}{11}\right) = 0.01051525.$$

Conclusion: Reject. ■

Sol. $k = 2$, $b = 22$

Compute the rank within each block (see the previous table)

Compute the g statistic:

$$g = \frac{12}{22 \times 2 \times (2+1)} \left[39^2 + 27^2\right] - 3 \times 22 \times (2+1) = \frac{72}{11} \approx 6.54.$$

Critical region is

$$C = \left\{g : g \geq \chi^2_{0.95,1} = 3.84\right\}.$$

The p-value is

$$\mathbb{P}\left(\chi^2_1 \geq \frac{72}{11}\right) = 0.01051525.$$

Conclusion: Reject. ■

R Code for this problem:

```
C1 <- c(
5.50, 5.70, 5.60, 5.50, 5.85, 5.55, 5.40, 5.50, 5.15, 5.80, 5.20,
5.55, 5.35, 5.00, 5.50, 5.55, 5.55, 5.50, 5.45, 5.60, 5.65, 6.30)
C2 <- c(
5.55, 5.75, 5.50, 5.40, 5.70, 5.60, 5.35, 5.35, 5.00, 5.70, 5.10,
5.45, 5.45, 4.95, 5.40, 5.50, 5.35, 5.55, 5.25, 5.40, 5.55, 6.25)
angles <- matrix(
  cbind(C1, C2),
  nrow = 22,
  byrow = FALSE,
  dimnames = list(1:22, c("Narrow", "Wide"))
)
friedman.test(angles)
```

Here is the output:

```
  > C1 <- c(
+ 5.50, 5.70, 5.60, 5.50, 5.85, 5.55, 5.40, 5.50, 5.15, 5.80, 5.20,
+ 5.55, 5.35, 5.00, 5.50, 5.55, 5.55, 5.50, 5.45, 5.60, 5.65, 6.30)
> C2 <- c(
+ 5.55, 5.75, 5.50, 5.40, 5.70, 5.60, 5.35, 5.35, 5.00, 5.70, 5.10,
+ 5.45, 5.45, 4.95, 5.40, 5.50, 5.35, 5.55, 5.25, 5.40, 5.55, 6.25)
> angles <- matrix(
+ cbind(C1, C2),
+ nrow = 22,
+ byrow = FALSE,
+ dimnames = list(1:22, c("Narrow", "Wide"))
+ )
> friedman.test(angles)

        Friedman rank sum test

data:  angles
Friedman chi-squared = 6.5455, df = 1, p-value = 0.01052
```