STA 104 Applied Nonparametric Statistics

Chapter 5: Two-Way Layout Problems: Nonparametric Two-Way Analysis of Variance

Xiner Zhou

Department of Statistics, University of California, Davis

Table of contents

The procedures of this chapter are designed for statistical analyses of data collected under an experimental design involving two factors, each at two or more levels.

	Treatments			
Blocks	1	2	...	k
1	X_{111}	X_{121}	...	X_{1k1}
	$\vdots$	$\vdots$	...	$\vdots$
	$X_{11c_{11}}$	$X_{12c_{12}}$	...	$X_{1kc_{1k}}$
2	X_{211}	X_{221}	...	X_{2k1}
	$\vdots$	$\vdots$	...	$\vdots$
	$X_{21c_{21}}$	$X_{22c_{22}}$	...	$X_{2kc_{2k}}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_{n11}	X_{n21}	...	X_{nk1}
	$\vdots$	$\vdots$	...	$\vdots$
	$X_{n1c_{n1}}$	$X_{n2c_{n2}}$	...	$X_{nkc_{nk}}$

Design Rationale.

The n blocks in this basic two-way layout design represent an effort to reduce experimental errors and prevent misleading comparisons of "apples and oranges."

Our primary interest is in the relative location effects (medians) of the different levels of one of these factors, hereafter called the treatment factor, within the various levels of the second factor, hereafter called the blocking factor.

This blocking factor is associated quite commonly with the experimental design where subjects are first divided into more homogeneous subgroups (called blocks) and then randomly assigned to the various treatment levels within these blocks.

Design Rationale.

Such a design is called a randomized block design.

As a special case, when there is only observation per treatment-block combination, it is called a randomized complete block design.

We will refer to the k levels of a treatment as the k treatments.

Setting

The data consist of $N = \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij}$ observations, with c_{ij} observations from the combination of the i th block with the j th treatment (i.e., the (i, j) th cell), for $i = 1, \ldots, n$ and $j = 1, \ldots, k$

- For each treatment-block combination $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, k\}$, the c_{ij} observations are a random sample from a continuous distribution with distribution function F_{ij}.
- The N observations are mutually independent.
- The distribution functions F_{ij} are connected through the relationship

$$F_{ij}(t) = F\left(t - \beta_i - \tau_j\right), -\infty < t < \infty,$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where F is a distribution function for a continuous distribution with unknown median θ,
β_i is the unknown additive effect contributed by block i, and
τ_j is the unknown additive treatment effect contributed by the j th treatment.

This is the usual Two-Way Analysis of Variance (ANOVA), commonly associated with normal assumptions and theory:

$$X_{ijt} = \theta + \beta_i + \tau_j + e_{ijt}, \quad i = 1, \ldots, n; \quad j = 1, \ldots, k; \quad t = 1, \ldots, c_{ij}$$

where

- θ is the overall median,
- τ_j is the treatment j effect,
- β_i is the block i effect,
- residuals e_{ijt}'s: random sample from a continuous distribution with median 0 . (Under the additional assumption of normality, the medians θ and 0 are, of course, also the respective means.)

Friedman test for General Alternatives in a Randomized Complete Block Design

Setting

The data consist of $N = \sum_{i=1}^{n} \sum_{j=1}^{k} 1 = nk$ observations, with 1 observations from the combination of the i th block with the j th treatment (i.e., the (i, j) th cell), for $i = 1, \ldots, n$ and $j = 1, \ldots, k$

	Treatments			
Blocks	1	2	$\ldots$	k
1	X_{11}	X_{12}	$\ldots$	X_{1k}
2	X_{21}	X_{22}	$\ldots$	X_{2k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_{n1}	X_{n2}	$\ldots$	X_{nk}

Hypothesis

H_0 : $\underbrace{\tau_1 = \ldots = \tau_k}$

$F_{i1},\ldots,F_{ik}$ within block i are the same; that is, $F_{i1} \equiv F_{i2} \equiv \cdots \equiv F_{ik} \equiv F_i$, for each fixed $i=1,\ldots,n$

H_1 : $\underbrace{\tau_1 \ldots \tau_k \text{ not all equal}}$

at least two of the treatment effects are not equal

Motivation

Friedman (1937) idea: within-block rank sum

$\Rightarrow$ first order the k observations from least to greatest separately within each of the n blocks

$\Rightarrow$ Let r_{ij} denote the rank of X_{ij} in the joint ranking of the observations $X_{i1}, \ldots, X_{ik}$ in the i th block and set

$$R_j = \sum_{i=1}^{n} r_{ij} \quad \text{and} \quad R_{.j} = \frac{R_j}{n}.$$

Thus, for example, R_2 is the sum (over the n blocks) of the within-blocks ranks received by the treatment 2 observations and $R_{.2}$ is the average within-blocks rank for these same observations.

If null is true:

$$E_0\left(r_{ij}\right) = \frac{k+1}{2},$$

the average rank being assigned separately in each of the blocks. Thus, we have

$$\begin{aligned} E_0\left(R_{\cdot j}\right) = E_0\left(\frac{1}{n}R_j\right) = \frac{1}{n}E_0\left(\sum_{i=1}^{n} r_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n} E_0\left(r_{ij}\right) \\ = \frac{n(k+1)}{2n} = \frac{k+1}{2}, \quad \text{for } j = 1, \ldots, k, \end{aligned}$$

and we would expect the $R_{\cdot j}$'s to be close to $(k+1)/2$ when H_0 is true.

$\Rightarrow$ The Friedman statistic S [1] is then given by

$$S = \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(R_{\cdot j} - \frac{k+1}{2} \right)^2$$

$$= \left[\frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 \right] - 3n(k+1)$$

- small values of S represent agreement with H_0
- When the τ 's are not all equal, we would expect a portion of the associated treatment average ranks to differ from their common null expectation, $(k+1)/2$, with some tending to be smaller and some larger. The net result (after squaring the observed differences to obtain the $[R_j - (k+1)/2]^2$ terms) would be a large value of S. This naturally suggests rejecting H_0 in favor of H_1 for large values of S.

[1] The Friedman S statistic also arises naturally if we apply the usual two-way layout $\mathcal{F}$ statistic to the ranks instead of the actual observations. Then S may be written as $S = [12/k(k+1)]$ SST, where SST is the treatment sum of squares applied to the ranks.

Derivation of null distribution using permutation

When H_0 is true, all possible $(k!)^n$ rank configurations for the r_{ij} 's are equally likely.

Large sample approximation of null distribution

Let $T_j = R_{.j} - E_0\left(R_{.j}\right) = R_{.j} - (k+1)/2$, for $j = 1, \ldots, k$

$\Rightarrow$ each $R_j = \sum_{i=1}^{n} r_{ij}/n$ is an average

$\Rightarrow$ properly standardized version of the vector $\mathbf{T}^* = (T_1, \ldots, T_{k-1})$ has an asymptotic (n tending to infinity) $(k-1)$-variate normal distribution when the null hypothesis H_0 is true.

Since the test statistic S is a quadratic form of $(T_1, \ldots, T_{k-1})$, therefore, quite natural that S has an asymptotic (n tending to infinity) chi-square distribution with $k-1$ degrees of freedom.

$$S \sim \chi^2_{k-1}$$

Procedure

Permutation

Reject H_0 if $S \geq s_\alpha$, otherwise do not reject

where s_α is the upper α percentile of the permutation null distribution.

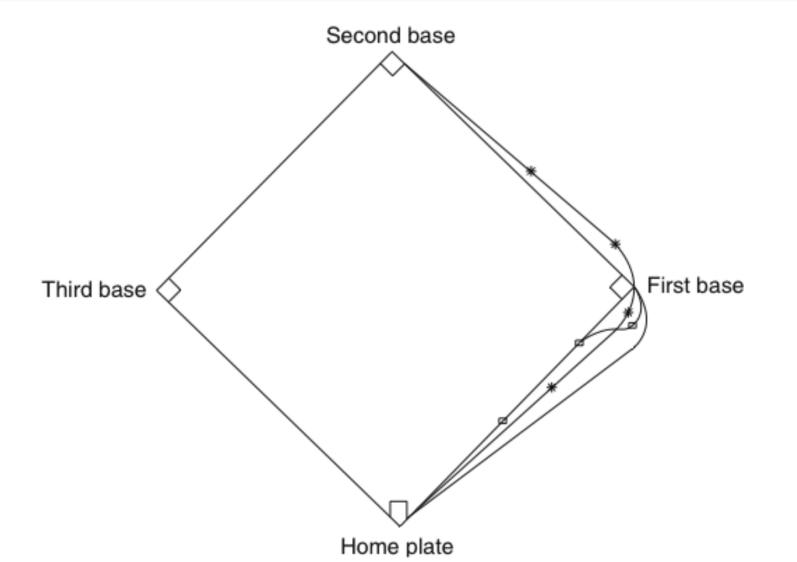
Large-sample approximation

Reject H_0 if $S \geq \chi^2_{k-1,\alpha}$, otherwise do not reject

where $\chi^2_{k-1,\alpha}$ is the upper α percentile point of a chi-square distribution with k-1 degrees of freedom.

Example:Rounding First Base

Woodward (1970) in a study to determine which, if any, of three methods of rounding first base is best, in the sense that it minimizes, on the average, the time to reach second base. The three methods are "round out," "narrow angle," and "wide angle".



Twenty-two baseball players participated in the study, and each of them ran from home plate to second base six times. Using a randomized order, these six trials per player were evenly divided (two each) among the three methods (round out, narrow angle, and wide angle). The entries are average times of the two runs per method. The within-blocks (players) ranks and the treatment (running method) rank sums $(R_1, R_2, \text{and } R_3)$ are are provided.

Players	Methods		
	Round out	Narrow angle	Wide Angle
1	5.40(1)	5.50(2)	5.55(3)
2	5.85(3)	5.70(1)	5.75(2)
3	5.20(1)	5.60(3)	5.50(2)
4	5.55(3)	5.50(2)	5.40(1)
5	5.90(3)	5.85(2)	5.70(1)
6	5.45(1)	5.55(2)	5.60(3)
7	5.45(2)	5.50(3)	5.35(1)
8	5.25(3)	5.15(2)	5.00(1)
9	5.85(3)	5.80(2)	5.70(1)
10	5.25(3)	5.20(2)	5.10(1)
11	5.65(3)	5.55(2)	5.45(1)
12	5.60(3)	5.35(1)	5.45(2)
13	5.05(3)	5.00(2)	4.95(1)
14	5.45(1)	5.55(3)	5.50(2)
15	5.45(1)	5.50(2)	5.55(3)
16	5.50(3)	5.45(2)	5.25(1)
17	5.65(3)	5.60(2)	5.40(1)
18	5.70(3)	5.65(2)	5.55(1)
	$R_1 = 43$	$R_2 = 37$	$R_3 = 28$

$$
\begin{aligned}
S &= \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1) \\
&= \frac{12}{18 \times 3(4)} (43^2 + 37^2 + 28^2) - 3 \times 18(4) \\
&= 6.33
\end{aligned}
$$

For the large-sample approximation, we compare the value of S to the chi-square distribution with $k - 1 = 2$ degrees of freedom.

```
> pchisq(6.33,df=2,lower.tail = F)
[1] 0.04221414
```

Hence, there is strong evidence here to reject the hypothesis that the methods are equivalent with respect to time to reach second base.

Check with built-in function: Agreed!

```
> RoundingTimes <- matrix(c(5.40, 5.50, 5.55,
+                           5.85, 5.70, 5.75,
+                           5.20, 5.60, 5.50,
+                           5.55, 5.50, 5.40,
+                           5.90, 5.85, 5.70,
+                           5.45, 5.55, 5.60,
+                           5.45, 5.50, 5.35,
+                           5.25, 5.15, 5.00,
+                           5.85, 5.80, 5.70,
+                           5.25, 5.20, 5.10,
+                           5.65, 5.55, 5.45,
+                           5.60, 5.35, 5.45,
+                           5.05, 5.00, 4.95,
+                           5.45, 5.55, 5.50,
+                           5.45, 5.50, 5.55,
+                           5.50, 5.45, 5.25,
+                           5.65, 5.60, 5.40,
+                           5.70, 5.65, 5.55),
+                         nrow = 18,
+                         byrow = TRUE,
+                         dimnames = list(1 : 18, c("Round Out", "Narrow Angle", "Wide Angle")))
>
> friedman.test(RoundingTimes)

        Friedman rank sum test

data:  RoundingTimes
Friedman chi-squared = 6.3333, df = 2, p-value =
0.04214
```

Two-Sided All-Treatments Multiple Comparisons for General Alternative in a Randomized Complete Block Design

After rejection of

H_0 : $\underbrace{\tau_1 = \ldots = \tau_k}$

$F_{i1},\ldots,F_{ik}$ within block i are the same; that is, $F_{i1} \equiv F_{i2} \equiv \cdots \equiv F_{ik} \equiv F_i$, for each fixed $i=1,\ldots,n$

H_1 : $\underbrace{\tau_1 \ldots \tau_k \text{ not all equal}}$

at least two of the treatment effects are not equal

with the Friedman test based on within block ranks, it is important to reach conclusions about all $\begin{pmatrix} k \\ 2 \end{pmatrix} = k(k-1)/2$ pairs of treatment effects and these conclusions are naturally two-sided.

Setting

The data consist of $N = \sum_{i=1}^{n} \sum_{j=1}^{k} 1 = nk$ observations, with 1 observations from the combination of the i th block with the j th treatment (i.e., the (i, j) th cell), for $i = 1, \ldots, n$ and $j = 1, \ldots, k$

	Treatments			
Blocks	1	2	$\ldots$	k
1	X_{11}	X_{12}	$\ldots$	X_{1k}
2	X_{21}	X_{22}	$\ldots$	X_{2k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_{n1}	X_{n2}	$\ldots$	X_{nk}

Hypothesis

$$\left\{ \begin{array}{ll} H_0 : \tau_1 = \tau_2 & H_1 : \tau_1 \neq \tau_2 \\ H_0 : \tau_1 = \tau_3 & H_1 : \tau_1 \neq \tau_3 \\ \dots & \\ H_0 : \tau_{k-1} = \tau_k & H_1 : \tau_{k-1} \neq \tau_k \end{array} \right\}$$

$\frac{k(k-1)}{2}$ simultaneous tests/multiple comparisons

Motivation

$|R_u - R_v|$ tend to be small when null $\tau_u = \tau_v$ is true, and tend to be large when alternative is $\tau_u \neq \tau_v$ is true

$\Rightarrow$ So we use absolute difference in within block rank sums as the test statistics

$\Rightarrow$ $(R_1, \ldots, R_k)$ as n tends to infinity, an asymptotic multivariate normal distribution $(Z_1 \ldots Z_k)$

$\Rightarrow$

$$\max_{1 \leq u < v \leq k} |R_u - R_v| \sim range(Z_1 \ldots Z_k)$$

$\Rightarrow$ To get the null distribution for the simultaneous tests, it is equivalent to know the null distribution of $\max_{1 \leq u < v \leq k} |R_u - R_v|$

$\Rightarrow$ It is then equivalent to the distribution of the range when we draw k independent $N(0, 1)$

Procedure

For each pair of treatments (u, v), for $1 \leq u < v \leq k$,

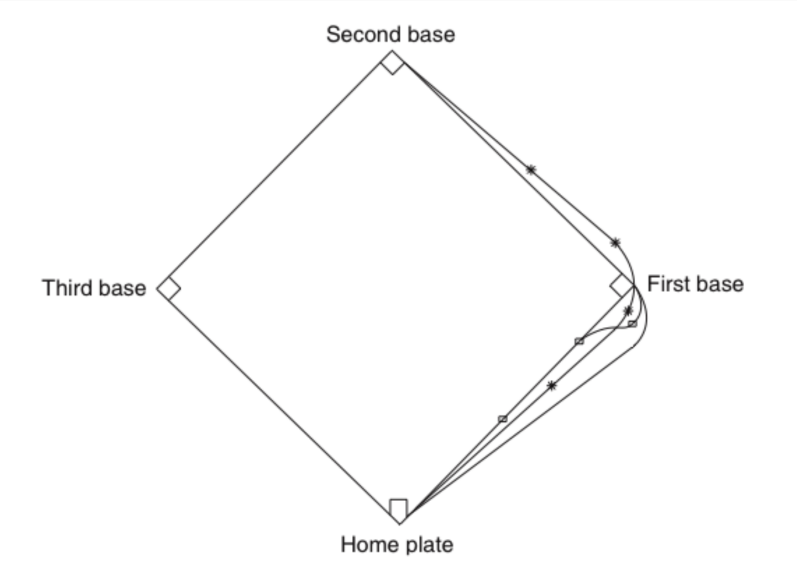
Decide $\tau_u \neq \tau_v$ if $|R_u - R_v| \geq q_\alpha \left[\frac{nk(k+1)}{12}\right]^{1/2}$

otherwise decide $\tau_u = \tau_v$

- q_α is the upper α quantile of the range of k normal variates.

Example:Rounding First Base

We have had found that there is strong evidence to conclude that the three methods of running to first base are not equivalent with respect to time to reach second base. Here we want to determine which of the three running methods differ in median times to second base.



Players	Methods		
	Round out	Narrow angle	Wide Angle
1	5.40(1)	5.50(2)	5.55(3)
2	5.85(3)	5.70(1)	5.75(2)
3	5.20(1)	5.60(3)	5.50(2)
4	5.55(3)	5.50(2)	5.40(1)
5	5.90(3)	5.85(2)	5.70(1)
6	5.45(1)	5.55(2)	5.60(3)
7	5.45(2)	5.50(3)	5.35(1)
8	5.25(3)	5.15(2)	5.00(1)
9	5.85(3)	5.80(2)	5.70(1)
10	5.25(3)	5.20(2)	5.10(1)
11	5.65(3)	5.55(2)	5.45(1)
12	5.60(3)	5.35(1)	5.45(2)
13	5.05(3)	5.00(2)	4.95(1)
14	5.45(1)	5.55(3)	5.50(2)
15	5.45(1)	5.50(2)	5.55(3)
16	5.50(3)	5.45(2)	5.25(1)
17	5.65(3)	5.60(2)	5.40(1)
18	5.70(3)	5.65(2)	5.55(1)
	$R_1 = 43$	$R_2 = 37$	$R_3 = 28$

```
> library(NSM3)
> cRangeNor(0.05,k=3)
[1] 3.315
```

Decide $\tau_u \neq \tau_v$ if $|R_u - R_v| \geq (3.315)\left[\frac{18(3)(4)}{12}\right]^{1/2} = 14.06435.$

Using the treatments sums of within-runners ranks given, we find that

$$|R_2 - R_1| = 6 < 14.06 \Rightarrow \text{ decide } \tau_2 = \tau_1$$
$$|R_3 - R_1| = 15 \geq 14.06 \Rightarrow \text{ decide } \tau_3 \neq \tau_1$$
$$|R_3 - R_2| = 9 < 14.06 \Rightarrow \text{ decide } \tau_3 = \tau_2$$

Thus, at an approximate experimentwise error rate of .05, we have reached the conclusion that only the round out (treatment 1) and wide angle (treatment 3) running methods yield significantly different median times to second base.

Mack-Skillings test for General Alternatives in a Randomized Block Design with Equal Number of Replications Per treatment-Block Combination

It is often the case in two-way layout settings that we have more than one observation for some of the treatment-block combinations.

These multiple observations in a given cell are referred to as replications for that treatment-block combination.

Here we focus on the setting where we have a common, equal number $c > 1$ of replications for every treatment-block combination.

Setting

The data consist of $N = \sum_{i=1}^{n} \sum_{j=1}^{k} c = nkc$ observations, with $c > 1$ observations from the combination of the i th block with the j th treatment (i.e., the (i, j) th cell), for $i = 1, \ldots, n$ and $j = 1, \ldots, k$

	Treatments			
Blocks	1	2	$\ldots$	k
1	X_{111}	X_{121}	$\ldots$	X_{1k1}
	$\vdots$	$\vdots$	$\ldots$	$\vdots$
	X_{11c}	X_{12c}	$\ldots$	X_{1kc}
2	X_{211}	X_{221}	$\ldots$	X_{2k1}
	$\vdots$	$\vdots$	$\ldots$	$\vdots$
	X_{21c}	X_{22c}	$\ldots$	X_{2kc}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_{n11}	X_{n21}	$\ldots$	X_{nk1}
	$\vdots$	$\vdots$	$\ldots$	$\vdots$
	X_{n1c}	X_{n2c}	$\ldots$	X_{nkc}

Hypothesis

H_0 : $\underbrace{\tau_1 = \ldots = \tau_k}$

$F_{i1},\ldots,F_{ik}$ within block i are the same; that is, $F_{i1} \equiv F_{i2} \equiv \cdots \equiv F_{ik} \equiv F_i$, for each fixed $i=1,\ldots,n$

H_1 : $\underbrace{\tau_1 \ldots \tau_k \text{ not all equal}}$

at least two of the treatment effects are not equal

Motivation

Mack-Skillings idea (Friedman-type):

$\Rightarrow$ first order the k observations from least to greatest separately within each of the n blocks

$\Rightarrow$ Let r_{ijq} denote the within block rank of X_{ijq} (the q th replication from the j th treatment in the i th block) among the kc total observations in the i th block and set

$$S_j = \sum_{i=1}^{n} \left[\sum_{q=1}^{c} r_{ijq}/c \right], \qquad \text{for } j = 1, \ldots, k$$

Thus, S_j is the sum (across blocks) of the cellwise averages of the within-blocks ranks assigned to the c observations from treatment j, for $j = 1, \ldots, k$.

$\Rightarrow$ The Mack-Skillings statistic for equal replications [2]

$$MS = \left[\frac{12}{k(N+n)}\right] \sum_{j=1}^{k} \left[S_j - \frac{N+n}{2}\right]^2$$

$$= \left[\frac{12}{k(N+n)}\right] \left\{\sum_{j=1}^{k} S_j^2\right\} - 3(N+n)$$

where $\frac{N+n}{2}$ is expected sum (across blocks) of the cellwise averages for each of the k treatments when H_0 is true; that is, $(N+n)/2$ is the expected value of S_j, for each $j = 1, \ldots, k$, when the null hypothesis H_0 is true.

[2] Special Case of $c = 1$: When we have a single observation for every treatment-block combination (i.e., $c = 1$), we are dealing with data from a complete randomized block design. In this setting, the Mack-Skillings statistic is equivalent to the Friedman statistic. Thus, the Mack-Skillings test represent natural extensions of the Friedman test, to the case of an equal number $c > 1$ of replications per cell.

$\Rightarrow$ Since MS is a constant times a sum of squared differences between the observed treatment sums of cellwise average ranks, S_j, and their common null expected value

- small values of MS represent agreement with H_0
- When the τ 's are not all equal, we would expect a portion of the associated treatment average ranks to differ from their common null expectation , with some tending to be smaller and some larger. The net result would be a large value of MS. This naturally suggests rejecting H_0 in favor of H_1 for large values of MS.

Derivation of null distribution using permutation

When H_0 is true, all possible $[(ck)!]^n$ permutations of the within-blocks ranks are equally likely. Thus, to obtain the exact null distribution of MS, we compute its value for each of these $[(ck)!]^n$ block rank configurations and then tabulate the collected outcomes. [3]

[3] he number $[(ck)!]^n$ of configurations for which we need to compute the value of MS can get large rather quickly, as either k or c is moderately increased.

Large sample approximation of null distribution

Define the centered treatment sums of cellwise average ranks
$S_j^* = S_j - E_0\left(S_j\right) = S_j - (N+n)/2$, for $j = 1, \ldots, k$

$\Rightarrow$ each S_j is an average

$\Rightarrow$ properly standardized version of the vector $\left(S_1^*, \ldots, S_{k-1}^*\right)$ has an asymptotic (nc tending to infinity) $(k-1)$-variate normal distribution when the null hypothesis H_0 is true.

Since the test statistic MS is a quadratic form of $\left(S_1^*, \ldots, S_{k-1}^*\right)$, therefore, quite natural that MS has an asymptotic (nc tending to infinity) chi-square distribution with $k-1$ degrees of freedom.

$$MS \sim \chi^2_{k-1}$$

Procedure

Permutation

Reject H_0 if $MS \geq ms_\alpha$, otherwise do not reject

where ms_α is the upper α percentile point of the permutation distribution.

Large-sample approximation

Reject H_0 if $MS \geq \chi^2_{k-1,\alpha}$, otherwise do not reject

where $\chi^2_{k-1,\alpha}$ is the upper α percentile point of a chi-square distribution with k-1 degrees of freedom.

Example:Determination of Niacin in Bran Flakes

In a study to investigate the precision and homogeneity of a procedure for assessing the amount of niacin in bran flakes, Campbell and Pelletier (1962) prepared homogenized samples of bran flakes enriched with 0,4 , or $8\mathrm{mg}$ niacin per 100 g of cereal. Portions of the homogenized samples were sent to different laboratories, which were asked to carry out the specified procedure for each of three separate samples. The resulting data (in milligrams per 100 g bran flakes) for a subset (4 out of 12) of the laboratories included in the study are presented.

	Amount of niacin enrichment (milligrams per 100 g bran flakes)		
Laboratory	0	4	8
1	7.58(3)	11.63(7)	15.00(2)
	7.87(8)	11.87(11)	15.92(9)
	7.71(6)	11.40(3)	15.58(4)
2	7.95(9)	12.20(12)	16.60(12)
	8.27(12)	11.6(8)	16.40(11)
	8.05(10)	11.80(10)	15.90(7)
3	7.60(4)	11.04(2)	15.87(6)
	7.30(1)	11.45(5)	15.91(8)
	7.82(7)	11.49(4)	16.28(10)
4	8.03(11)	11.50(6)	15.10(3)
	7.35(2)	10.10(1)	14.80(1)
	7.66(5)	11.70(9)	15.70(5)

Of primary interest here is the precision of the laboratory procedure for determining niacin content in bran flakes. The actual amount of niacin enrichment in the prepared bran flakes serves only as a "nuisance" blocking factor in our evaluation of the consistency of the results across the four laboratories.

Hence, we have data from a two-way layout with $k = 4$ treatments (laboratories), $n = 3$ blocks (amounts of niacin enrichment), and $c = 3$ replications (individual bran flake samples) per laboratory/enrichment combination.

$$S_1 = \frac{3+8+6+7+11+3+2+9+4}{3} = 17.67,$$
$$S_2 = \frac{9+12+10+12+8+10+12+11+7}{3} = 30.33,$$
$$S_3 = \frac{4+1+7+2+5+4+6+8+10}{3} = 15.67,$$
$$S_4 = \frac{11+2+5+6+1+8.5+3+1+5}{3} = 14.17$$
$$\Rightarrow MS = \left[\frac{12}{4(36+3)}\right]\{1678.476\} - 3(36+3)$$
$$= 12.11$$

Large-sample p-value:

```
> pchisq(12.11,df=3,lower.tail = F)
[1] 0.007015691
```

We can reject H_0 at the $\alpha = .05$ level, providing rather strong evidence that the studied process for assessing niacin content in bran flakes does not produce consistent results across a variety of laboratories and is therefore not reliable as an evaluative procedure.

Two-Sided All-Treatments Multiple Comparisons for General Alternative in a Randomized Block Design with Equal Number of Replications Per treatment-Block Combination

After rejection of

H_0 : $\underbrace{\tau_1 = \ldots = \tau_k}$

$F_{i1},\ldots,F_{ik}$ within block i are the same; that is, $F_{i1} \equiv F_{i2} \equiv \cdots \equiv F_{ik} \equiv F_i$, for each fixed $i=1,\ldots,n$

H_1 : $\underbrace{\tau_1 \ldots \tau_k \text{ not all equal}}$

at least two of the treatment effects are not equal

with Mack-Skillings test, it is important to reach conclusions about all $\binom{k}{2} = k(k-1)/2$ pairs of treatment effects and these conclusions are naturally two-sided.

Hypothesis

$$\left\{ \begin{array}{ll} H_0 : \tau_1 = \tau_2 & H_1 : \tau_1 \neq \tau_2 \\ H_0 : \tau_1 = \tau_3 & H_1 : \tau_1 \neq \tau_3 \\ \dots & \\ H_0 : \tau_{k-1} = \tau_k & H_1 : \tau_{k-1} \neq \tau_k \end{array} \right\}$$

$\frac{k(k-1)}{2}$ simultaneous tests/multiple comparisons

Motivation

$\Rightarrow S_1, \ldots, S_k$: the treatment sums of cellwise averages of within-blocks ranks

$\Rightarrow |R_u - R_v|$ tend to be small when null $\tau_u = \tau_v$ is true, and tend to be large when alternative $\tau_u \neq \tau_v$ is true

$\Rightarrow$ So we use absolute difference in within block rank sums as the test statistics

$\Rightarrow (R_1, \ldots, R_k)$ as N tends to infinity, an asymptotic multivariate normal distribution $(Z_1 \ldots Z_k)$

$\Rightarrow$

$$\max_{1 \leq u < v \leq k} |R_u - R_v| \sim range(Z_1 \ldots Z_k)$$

$\Rightarrow$ To get the null distribution for the simultaneous tests, it is equivalent to know the null distribution of $\max_{1 \leq u < v \leq k} |R_u - R_v|$

$\Rightarrow$ It is then equivalent to the distribution of the range when we draw k independent $N(0, 1)$

Procedure

For each pair of treatments (u, v), for $1 \leq u < v \leq k$,

Decide $\tau_u \neq \tau_v$ if $|S_u - S_v| \geq [k(N+n)/12]^{1/2} q_c$

otherwise decide $\tau_u = \tau_v$

- q_α is the upper α quantile of the range of k normal variates.

Example:Determination of Niacin in Bran Flakes

We have found rather strong evidence that the studied process for assessing niacin content in bran flakes does not produce consistent results across a variety of laboratories.

To determine which of the laboratories differ in median detected niacin content in the bran flakes.

	Amount of niacin enrichment (milligrams per 100 g bran flakes)		
Laboratory	0	4	8
1	7.58(3)	11.63(7)	15.00(2)
	7.87(8)	11.87(11)	15.92(9)
	7.71(6)	11.40(3)	15.58(4)
2	7.95(9)	12.20(12)	16.60(12)
	8.27(12)	11.6(8)	16.40(11)
	8.05(10)	11.80(10)	15.90(7)
3	7.60(4)	11.04(2)	15.87(6)
	7.30(1)	11.45(5)	15.91(8)
	7.82(7)	11.49(4)	16.28(10)
4	8.03(11)	11.50(6)	15.10(3)
	7.35(2)	10.10(1)	14.80(1)
	7.66(5)	11.70(9)	15.70(5)

```
> library(NSM3)
> cRangeNor(0.05,k=4)
[1] 3.634
```

Decide $\tau_u \neq \tau_v$ if $|S_u - S_v| \geq [4(36+3)/12]^{1/2}(3.634) = 13.1$

$$
\begin{aligned}
|S_2 - S_1| &= 12.83 < 13.1 \Rightarrow \text{ decide } \tau_2 = \tau_1, \\
|S_3 - S_1| &= 1.84 < 13.1 \quad \Rightarrow \quad \text{decide } \tau_3 = \tau_1, \\
|S_4 - S_1| &= 3.67 < 13.1 \Rightarrow \text{ decide } \tau_4 = \tau_1, \\
|S_3 - S_2| &= 14.67 > 13.1 \Rightarrow \text{ decide } \tau_3 \neq \tau_2, \\
|S_4 - S_2| &= 16.5 > 13.1 \quad \Rightarrow \quad \text{decide } \tau_4 \neq \tau_2, \\
|S_4 - S_3| &= 1.83 < 13.1 \Rightarrow \text{ decide } \tau_4 = \tau_3.
\end{aligned}
$$

Thus, at an approximate experimentwise error rate of .05, we see that Laboratory 2 yielded significantly different median detected niacin content than either Laboratory 3 or Laboratory 4.

These multiple comparison decisions help to focus the rationale for the original rejection of the Mack-Skillings test, as it now seems reasonable to question the reliability of Laboratory 2 in conducting this niacin content process.