STAT 518 --- Section 5.8: Block Designs

• Recall that in <u>paired-data</u> studies, we match up pairs of subjects so that the two subjects in a pair are alike in some sense.

• Then we randomly assign, say, treatment A to one subject within a pair and, say, treatment B to the other subject. Picture:

1st pair → B A

2nd pair → A B

A B

⋮

n-th pair → B A

or

Before After ← subject 1

maybe

Before After ← subject 2

⋮

• The rationale is that any difference we see in the measurements will likely be due to the treatments rather than intrinsic differences in the subject.

• When there are $k > 2$ treatments, we can separate the subjects into blocks, each having k subjects that are similar in some way (age, weight, health status, etc.).

Picture:

$(k=4)$

A C B D ← block 1

C A D B ← block 2

⋮

B D C A ← block b

Example 1: We compare mean crop yields for 4 different fertilizers, where the blocks are contiguous regions of a field.
Example 2: We compare mean blood pressure reductions for 3 different drugs, where the blocks are formed of patients of similar age and health.
Example 3: We compare mean aggressiveness for mice placed in 5 different environments, where the blocks are the litters the mice came from.

- Assume there are *b* blocks, and each of the *k* treatments appears exactly once in each block.

- This experimental design is called a randomized complete block design.

- We are primarily interested in testing whether the *k* treatment groups have the same population mean.

The Friedman Test

The data are arranged as follows:

		Treatment			
		1	2	$\cdots$	k
Block	1	X_{11}	X_{12}	$\cdots$	X_{1k}
	2	X_{21}	X_{22}	$\cdots$	X_{2k}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	b	X_{b1}	X_{b2}	$\cdots$	X_{bk}

- **We rank the data from 1 to *k* separately within each block (use midranks in case of ties).**
- **Let $R(X_{ij})$ be the rank within block *i* of the subject receiving treatment *j*.**

- **Then the sum of ranks for each treatment is**

$$R_j = \sum_{i=1}^{b} R(X_{ij})$$

- **If the *b* multivariate observations $(X_{11}, \ldots, X_{1k}), \ldots, (X_{b1}, \ldots, X_{bk})$ are mutually independent (i.e., independent from block to block), and the data are at least ordinal so that we may rank data with each block, then Friedman's Test can test the hypotheses:**

H_0: The k treatments have identical effects

vs.

H_1: At least one treatments tends to yield larger values than at least one other treatment

- **Friedman suggested the statistic**

$$T_1 = \frac{(k-1)\sum_{j=1}^{k}\left(R_j - \frac{b(k+1)}{2}\right)^2}{A_1 - C_1}$$

$$A_1 = \sum_{i=1}^{b}\sum_{j=1}^{k}\left[R(X_{ij})\right]^2, \quad C_1 = bk(k+1)^2/4$$

which has an approximate χ^2_{k-1} **null distribution.**

• **But it is often preferred to use**

$$T_2 = \frac{(b-1)T_1}{b(k-1) - T_1}$$

which has an approximate $F_{k-1,\,(b-1)(k-1)}$

null distribution, because the approximation is better.

• **If H_0 is rejected, a multiple comparisons procedure is available to determine which pairs of treatments differ significantly.**

Conclude treatments *i* and *j* are different if:

$$|R_j - R_i| > t_{1-\alpha/2,\,(b-1)(k-1)} \left[\frac{2\left(bA_1 - \sum R_j^2\right)}{(b-1)(k-1)} \right]^{1/2}$$

where

$$A_1 = \sum_{i=1}^{b} \sum_{j=1}^{k} \left[R(X_{ij})\right]^2$$

<u>Example 1</u> (Grass data): In a horticultural experiment for compare 4 grass types, 12 homeowners each grew the 4 grasses in parts of their yards. Since the same homeowner would provide the same degree of care to each grass, the homeowners were treated as blocks. At the end of the study, the grasses were ranked from 1 (worst) to 4 (best), separately within each block. The data are on page 372. Is there a significant difference in mean quality among the 4 grasses? (Use $\alpha = 0.05$.)

• **The sum of ranks for each grass is:**

$R_1 = 38$, $R_2 = 23.5$, $R_3 = 24.5$, $R_4 = 34$

Test statistic: $T_1 = 8.0973 \Rightarrow T_2 = \frac{(12-1)8.0973}{12(4-1)-8.0973} = 3.19$

Decision rule:

Reject H_0 if $T > F_{1-\alpha, k-1, (b-1)(k-1)}$

F-table (A22):

$F_{.95, 3, 33} \approx 2.9$

Reject H_0 if $T > 2.9$

Since $3.19 > 2.9$, reject H_0 and conclude the mean quality of the 4 grass types is not the same.

From R, P-value $\approx .044$

Multiple comparisons:

Cutoff = 11.48 (see calculations in R)

If $|R_i - R_j| > 11.48$, then grasses i and j are signif. different

$\Rightarrow$ Grasses 1 and 2 are signif. different.

Grasses 1 and 3 are signif. different.

Related Tests

• **The Quade Test tests the same hypotheses as Friedman's Test.**

• **Quade's test gives more weight to blocks whose sample range (maximum – minimum value) is largest.**

• **It is sometimes more powerful than Friedman's test, but it requires that we have actual data values, not just ranks as in Example 1.**

• **Page's Test is appropriate when our alternative hypothesis specified a particular ordering of the population means. The hypotheses are:**

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

vs. $H_1: \mu_1 \le \mu_2 \le \cdots \le \mu_k$

with at least one strict inequality.

• **A test statistic is**

$$T_4 = \sum_{j=1}^{k} j R_j = R_1 + 2R_2 + \cdots + kR_k$$

where the ordering of $R_1, \ldots, R_k$ correspond to the ordering specified in H_1.

• **The decision rule is based on a function of the test statistic that has a normal asymptotic null distribution:**

$$T_5 = \frac{T_4 - bk(k+1)^2/4}{\left[b(k^3-k)^2/144(k-1)\right]^{1/2}}$$

Example 1 again: Suppose we suspected that the ordering of the grasses was 2, 3, 4, 1. Test whether there is evidence for this ordering.

$$T_4 = 23.5 + 2(24.5) + 3(34) + 4(38) = 326.5$$

$$T_5 = \frac{326.5 - 12(4)(5)^2/4}{\left[12(4^3-4)^2/144(3)\right]^{1/2}} = \frac{26.5}{10} = 2.65$$

Reject H_0 if $T_5 > Z_{.95} = 1.645$ (Table A1)

Since $2.65 > 1.645$, reject H_0 and conclude there is evidence for the ordering.

Note: If each treatment appears $m > 1$ times (i.e., replication) within each block, Friedman's test can be adjusted to account for this. See pp. 383-384 for details.

Comparison to Competing Tests

• Note: For $k = 2$, Friedman's test is equivalent to the sign test and Quade's test is equivalent to the Wilcoxon signed-rank test.

• In general, for k samples, the A.R.E. of Friedman's test relative to the classical F-test depends on k.

• When the alternative is a simple shift in treatment means, the A.R.E. of Friedman's test relative to the classical F-test is never less than $0.864(k/k+1)$.

Efficiency of Friedman's Test

Population	A.R.E.(Friedman vs. F)
Normal	$.955\left(\frac{k}{k+1}\right)$
Uniform (light tails)	$\frac{k}{k+1}$
Double exponential (heavy tails)	$1.5\left(\frac{k}{k+1}\right)$

• Quade's test is more powerful than Friedman's when k is 3 or 4, but tends to be less powerful than Friedman's when $k \geq 5$.