## Nonparametric Methods

- Most of the methods we have studied require distributional assumptions (on either joint or marginal distributions)
  - multivariate normal joint distribution (either unstructured or with a special covariance structure)
  - multinomial joint distribution
  - normal, Bernoulli, Poisson, gamma marginal distributions
- The exceptions were:
  - summary statistic approach using nonparametric measures of association
  - randomization model analyses based on CMH statistics

## Rationale for Nonparametric Methods

1. For continuous responses:

   a. the assumption of multivariate normality is not always reasonable

   b. the actual distribution may be unknown

   Thus, the use of standard parametric procedures is subject to criticism

2. For ordered categorical responses with a large number of possible outcomes:

   a. general categorical data methods may be inapplicable due to sample size limitations

   b. assumptions of specific ordinal data models may be inappropriate

3. To confirm the results of parametric analyses

## Some Nonparametric Methods

*Summary Statistic Approach*

- Analysis of a univariate function of the repeated measurements using distribution-free methods
- Ghosh et al. (1973) extension based on the use of two or more summary statistics for each subject

*Rank Correlation Methods*

- Another type of multivariate approach based on summary statistics
- Applicable when ordinal response is measured at multiple time points in several ordered groups
- Rank measures of association between group and response are constructed at each time
- Covariance matrix of these correlated measures of association is then estimated
- Carr, Hafner, and Koch (1989)

## Some Nonparametric Methods

*Multivariate Generalizations of Univariate Distribution-Free Methods*

(rank-based methods for samples from continuous multivariate distributions)

- Multivariate one-sample tests for complete data
  - Nonparametric analogues of Hotelling's $T^2$
  - Multivariate generalizations of the sign and Wilcoxon signed rank tests
  - Hettmansperger (1984, chapter 6)
  - Puri and Sen (1971, chapter 4)

- Multivariate multisample tests for complete data
  - Nonparametric analogues of MANOVA
  - Multivariate generalizations of the Kruskal-Wallis and Brown-Mood (1951) median tests
  - Puri and Sen (1971, chapter 5)

# Some Nonparametric Methods

*Counterparts of Multivariate Normal Methods*

- Nonparametric counterparts of Hotelling's $T^2$ statistic and profile analysis (Bhapkar, 1984)
- Nonparametric analogues of the Potthoff-Roy growth curve model (Sen, 1984)

*Randomization Model Approaches*

- Cochran-Mantel-Haenszel tests for one-sample repeated measures using rank scores
- Randomization analysis of growth curves Zerbe and Walker (1977), Zerbe (1979)

*Two-sample Tests for Incomplete Data*

- Wei & Lachin (1984), Wei & Johnson (1985)
- Palesch and Lachin (1994) extension to more than two groups

## Some Nonparametric Methods

*Rank Transform Methods*

- Replace the observations by their ranks and then perform standard parametric analyses
- Inappropriate for many common hypotheses in the repeated measures setting (Akritas, '91, '93)
- Thompson (1991) and Akritas & Arnold (1994) provide valid asymptotic tests for hypotheses of interest in several repeated measures models

*Nonparametric Regression Methods*

- Approaches based on kernel estimation, weighted local least squares estimation, and smoothing splines
- Müller (1988), Diggle et al. (1994, Chapter 3), Kshirsagar and Smith (1995, Chapter 10)

## Multivariate Multisample Nonparametric Tests for Complete Data

- Puri and Sen (1971) considered the problem of testing the equality of $s$ multivariate distributions $F_1, \ldots, F_s$, where $F_h$ is a $t$-variate cdf

- When the underlying distributions $F_1, \ldots, F_s$ are multivariate normal, they can differ only in their mean vectors and covariance matrices

- However, for non-normal $F_h$, differences among distributions may be due to a variety of reasons

- Equality of location vectors and covariance matrices does not imply that $F_1 = \cdots = F_s$

- Puri and Sen assumed that the cdfs $F_h$ had a common unspecified form, but differed in their location (or scale) vectors

# Multivariate Multisample Nonparametric Tests for Complete Data

- Puri & Sen considered the general null hypothesis $H_0\colon F_1(x) = \cdots = F_s(x)$ for all $x,$ where $F \in \Omega,$ and $\Omega$ is the class of continuous cdfs

- The general alternative hypothesis was that each $F_h \in \Omega$ and not all equal

- Although they considered both translation- and scale-type alternatives, we shall consider the case

$$F_h(x) = F(x + \Delta_h), \quad h = 1, \ldots, s$$

- The null hypothesis of interest is

$$H_0\colon \Delta_1 = \cdots = \Delta_s = (0, \ldots, 0)'$$

- The alternative is that $\Delta_1 \ldots, \Delta_s$ are not all equal

## Methodology for Repeated Measures

- Suppose repeated measurements at $t$ time points have been obtained from $s$ groups of subjects
- Let $n_h$ denote the number of subjects in group $h$ and let $n = \sum_{h=1}^{s} n_h$
- Let $y_{hij}$ denote the response at time $j$ from the $i$th subject in group $h$, for $h = 1, \ldots, s$, $i = 1, \ldots, n_h$, and $j = 1, \ldots, t$
- Let $F_h(x + \Delta_h)$ denote the cdf in group $h$, where $x' = (x_1, \ldots, x_t)$ and $\Delta_h = (\Delta_{h1}, \ldots, \Delta_{ht})'$
- The test of no difference among groups across all time points tests $H_0\colon \Delta_1 = \cdots = \Delta_s = 0_t$
- The omnibus alternative is that not all groups are the same at all time points

## Methodology for Repeated Measures

- The data can be displayed as follows:

| | | Time Point | | | | |
|---|---|---|---|---|---|---|
| Group | Subject | 1 | $\ldots$ | $j$ | $\ldots$ | $t$ |
| 1 | 1 | $y_{111}$ | $\cdots$ | $y_{11j}$ | $\cdots$ | $y_{11t}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $i$ | $y_{1i1}$ | $\cdots$ | $y_{1ij}$ | $\cdots$ | $y_{1it}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $n_1$ | $y_{1n_11}$ | $\cdots$ | $y_{1n_1j}$ | $\cdots$ | $y_{1n_1t}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $h$ | 1 | $y_{h11}$ | $\cdots$ | $y_{h1j}$ | $\cdots$ | $y_{h1t}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $i$ | $y_{hi1}$ | $\cdots$ | $y_{hij}$ | $\cdots$ | $y_{hit}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $n_h$ | $y_{hn_h1}$ | $\cdots$ | $y_{hn_hj}$ | $\cdots$ | $y_{hn_ht}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $s$ | 1 | $y_{s11}$ | $\cdots$ | $y_{s1j}$ | $\cdots$ | $y_{s1t}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $i$ | $y_{si1}$ | $\cdots$ | $y_{sij}$ | $\cdots$ | $y_{sit}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $n_s$ | $y_{sn_s1}$ | $\cdots$ | $y_{sn_sj}$ | $\cdots$ | $y_{sn_st}$ |

## Rank Permutation Principle

- Rank the $t$ columns of the $n \times t$ data matrix $Y$ (all groups combined) in ascending order
- Let $R$ denote the $n \times t$ matrix of ranks
- Under $H_0$, each column of $R$ is a random permutation of the numbers $1, \ldots, n$
- Two such matrices are *permutationally equivalent* if one can be obtained from the other by a rearrangement of its rows
- Let $R^*$ denote the matrix that has the same row vectors as $R$, but is arranged so that its first column is ordered $1, \ldots, n$
- $R^*$ has $(n!)^{t-1}$ possible realizations

## Rank Permutation Principle

- The $t$ components of $y_{hi} = (y_{hi1}, \ldots, y_{hit})'$ are, in general, stochastically dependent

- Thus, the joint distribution of the elements of $R$ (or $R^*$) will depend on the unknown distribution $F$ (even when $H_0 \colon F_1 = \cdots = F_s = F$ is true)

- Let $\mathcal{R}^*$ denote the set of all $(n!)^{t-1}$ possible realizations of $R^*$

- The unconditional distribution of $R^*$ over $\mathcal{R}^*$ depends on $F_1, \ldots, F_s$

- When $F_1 = \cdots = F_s$, the $n$ random vectors

$$y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}, \ldots, y_{s1}, \ldots, y_{sn_s}$$

  are independent and identically distributed

## Rank Permutation Principle

- The joint distribution of the $y_{hi}$ is invariant under any permutation among themselves

- Thus, the conditional distribution of $R$ over the set of $n!$ possible permutations of the columns of $R^*$ is uniform under $H_0\colon F_1 = \cdots = F_s = F$, i.e.,

$$\Pr(R = r \mid S(R^*), H_0) = 1/n! \text{ for all } r \in S(R^*)$$

- Puri and Sen define $\mathcal{P}$ as the conditional (permutational) probability measure generated by the $n!$ equally likely possible permutations of the columns of $R^*$

- They show that any statistic which depends explicitly on $R$ has a completely specified conditional distribution under $\mathcal{P}$

## Permutation Rank Order Tests

- Let $R_{ij}$ denote the $(i, j)$th element of $R_{(n \times t)}$
- Let $E_{ij} = J(R_{ij}/(n + 1))$ for some function $J$ satisfying Puri and Sen's (1971, p. 95) conditions
- Let $E$ denote the $n \times t$ matrix of rank scores
- Let $\overline{E}_{hj}$ denote the average rank score at the $j$th time point in the $h$th sample
- Puri and Sen derive a test statistic $L$ which is a weighted sum of $s$ quadratic forms in $\overline{E}_h - \overline{E}_.$
- $\overline{E}_h$ is the $t \times 1$ vector of average rank scores from the $h$th sample
- $\overline{E}_.$ is the vector of average rank scores from all samples combined

## Permutation Rank Order Tests

- The conditional distribution of $L$ given $R^*$ is the same under $H_0$, regardless of $F$

- Under $H_0$, the $t(s-1)$ contrasts $\overline{E}_{hj} - \overline{E}_{.j}$ are stochastically small in absolute value

- The test criteria $L$ rejects $H_0$ if any of these contrasts are numerically too large

- Unless $n$ and $t$ are both small, exact application of the permutation test based on $L$ is difficult

- Puri and Sen (1971) show that the asymptotic null distribution of $L$ is $\chi^2_{t(s-1)}$

- They also note that $L$ is asymptotically equivalent to the LR test based on $T^2$

## Multivariate Multisample Rank Sum Test

- For each sample at each time point, the MMRST compares the difference between the sample average rank and the combined data average rank

- Let $r_h$ denote the average rank vector $(t \times 1)$ from the $h$th group:
$$r_{hj} = \frac{\displaystyle\sum_{i=1}^{n_h} r_{hij}}{n_h},$$
where $r_{hij}$ is the rank of the $j$th response from the $i$th subject in sample $h$

- Let $\overline{r}_{.}$ denote the average rank vector $(t \times 1)$ for the combined samples:
$$\overline{r}_{.j} = \frac{\displaystyle\sum_{h=1}^{s} \sum_{i=1}^{n_h} r_{hij}}{\displaystyle\sum_{h=1}^{s} n_h}$$

## Multivariate Multisample Rank Sum Test

- The test statistic is

$$L_{RS} = \sum_{h=1}^{s} n_h (r_h - \overline{r}_.)' V^{-1} (r_h - \overline{r}_.)$$

- The covariance matrix $V$ has elements

$$V_{jl} = \left( \sum_{h=1}^{s} \sum_{i=1}^{n_h} r_{hij} r_{hil} \Big/ \sum_{h=1}^{s} n_h \right) - \overline{r}_{.j} \overline{r}_{.l}$$

- $L_{RS}$ tests the hypothesis of no differences in the multivariate response profiles from the $s$ samples

- The asymptotic null distribution of $L_{RS}$ is $\chi^2_{t(s-1)}$

- If $t = 1$, $L_{RS}$ reduces to the Kruskal-Wallis test

- Schwertman (1982) gives a FORTRAN subroutine for computing the MMRST

## Multivariate Multisample Median Test

- The MMMT compares differences between proportions less than or equal to the median to the corresponding combined data proportions

- Let $p_h$ denote the $t \times 1$ vector of proportions from the $h$th sample which are less than or equal to the median of the combined samples:

$$p_{hj} = \textstyle\sum_{i=1}^{n_h} x_{hij}/n_h, \text{ where}$$

$$x_{hij} = \begin{cases} 1 & \text{if } r_{hij} \leq \sum_{h=1}^{s} n_h/2 \\ 0 & \text{otherwise} \end{cases}$$

- Let $\overline{p}_{.}$ denote the $t \times 1$ vector of proportions from the combined samples that are less than or equal to the combined samples median:

$$\overline{p}_{.j} = \sum_{h=1}^{s} \sum_{i=1}^{n_h} x_{hij} \Big/ \sum_{h=1}^{s} n_h$$

## Multivariate Multisample Median Test

- The test statistic is

$$L_M = \sum_{h=1}^{s} n_h (p_h - \overline{p}_.)' V^{-1} (p_h - \overline{p}_.)$$

- The covariance matrix $V$ has elements

$$V_{jl} = \left( \sum_{h=1}^{s} \sum_{i=1}^{n_h} x_{hij} x_{hil} \Big/ \sum_{h=1}^{s} n_h \right) - \overline{p}_{.j} \overline{p}_{.l}$$

- $L_M$ tests the hypothesis of no differences in the multivariate response profiles from the $s$ samples

- The asymptotic null distribution of $L_M$ is $\chi^2_{t(s-1)}$

- If $t = 1$, $L_M$ reduces to the Brown-Mood (1951) several-sample median test

- Schwertman (1982) gives a FORTRAN subroutine for computing the MMMT

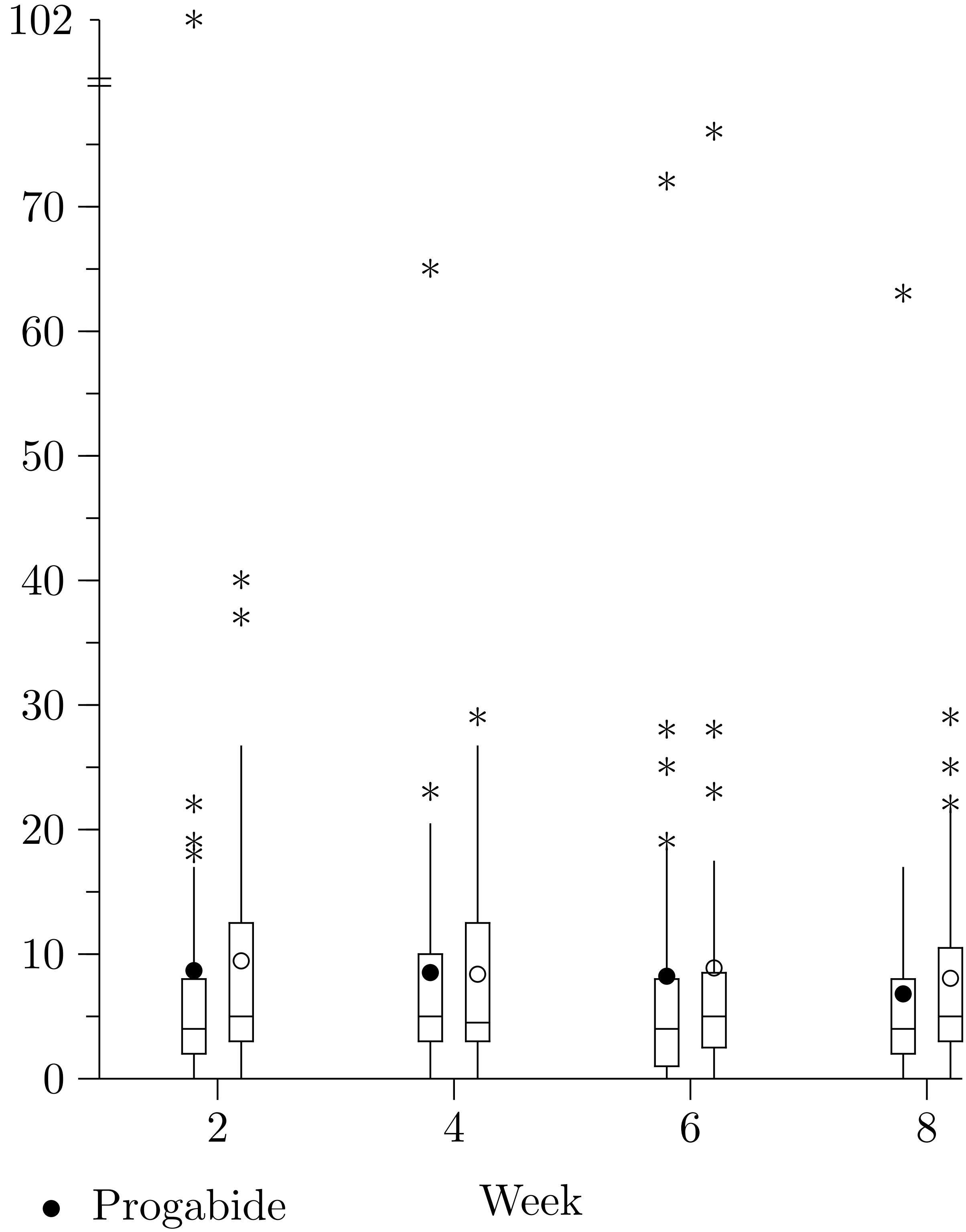
## Example

- Leppik et al. (1987) conducted a clinical trial in 59 epileptic patients
- Patients suffering from simple or complex partial seizures were randomized to receive either the antiepileptic drug progabide (31 patients) or a placebo (28 patients)
- At each of four successive postrandomization visits, the number of seizures occurring during the previous two weeks was reported
- The medical question of interest is whether or not progabide reduces the frequency of epileptic seizures

### Reference

Leppik IE, Dreifuss FE, Porter R et al. (1987). A controlled study of progabide in partial seizures: methodology and results. *Neurology* **37**, 963–968.

## Modified Box Plots of Seizure Counts



## Multivariate Approach

- During each two-week period, there appears to be a slight tendency for seizure counts to be lower in progabide-treated patients than in placebo-treated patients

- The median number of seizures in the progabide group at weeks 2, 4, 6, and 8 is 4, 5, 4, and 4, respectively

- The corresponding medians in the placebo group are 5, 4.5, 5, and 5, respectively

- Using the multivariate multisample rank sum test, the chi-square statistic is 5.47 with 4 df ($p = 0.24$)

- The multivariate multisample median test gives an even less-significant result (chi-square$= 3.46$, df$= 4$, $p = 0.48$)

## Two-Sample Nonparametric Tests for Incomplete Repeated Measures

- General methods for comparing two samples of incomplete repeated measures were studied by:

  Wei and Lachin (1984, *JASA*)

  Wei and Johnson (1985, *Biometrika*)

- The methods make no assumptions concerning the distribution of the response variable

- The missing value patterns in the two groups are allowed to be different and both “embedded” and “tail” missing observations can be accommodated

- The missing data mechanism, however, must be independent of the response

- Methods are limited to two-group comparisons

## Wei-Lachin Method

- A family of asymptotically distribution-free tests for equality of two multivariate distributions, based on censored data

- Proposed and developed for multivariate censored failure time data

- Natural generalizations of the log-rank test and the Gehan-Wilcoxon test for survival data

- Based on the commonly used random censorship model (Kalbfleisch and Prentice, 1980)
    - censoring vectors for each subject are mutually independent and also independent of the underlying failure time vectors

- The methodology is also applicable to repeated measures with missing observations

## Wei-Lachin Method

- Let $y_{hi} = (y_{hi1}, \ldots, y_{hit})'$ denote the repeated observations from subject $i$ in group $h$, for $h = 1, 2$ and $i = 1, \ldots, n_h$

- Apart from a scale factor, the $j$th component of the W-L vector of test statistics equals

$$T_j = \sum_{i=1}^{n_1} \sum_{i'=1}^{n_2} \delta_{1ij} \delta_{2i'j} \phi(y_{1ij}, y_{2i'j}),$$

  where

$$\phi(x, y) = \begin{cases} 1 & \text{if } y > x \\ 0 & \text{if } y = x \\ -1 & \text{if } y < x \end{cases}$$

  and $\delta_{hij}$ is 1 if $y_{hij}$ is observed, 0 otherwise

- Thus, at each time point $j$, comparisons between group 1 and group 2 are made for all $i, i'$ for which $y_{1ij}$ and $y_{2i'j}$ are both observed

# Wei-Lachin Omnibus Test

- Let $F_h(x_1, \ldots, x_t)$ denote the multivariate cdf of the repeated observations from group $h$, for $h = 1, 2$

- The statistic for testing

$$H_0\colon F_1(x_1, \ldots, x_t) = F_2(x_1, \ldots, x_t)$$

  against the general alternative that $F_1 \neq F_2$ is $T'\widehat{\Sigma}_T^{-1}T$, where

  - $T' = (T_1, \ldots, T_t)$

  - $\widehat{\Sigma}_T$ is a consistent estimator of $\mathrm{Var}(T)$ (Wei and Lachin, 1984, Theorem 1)

- The asymptotic null distribution of this statistic is $\chi^2_t$

## Wei-Lachin One-Sided Test

- In many studies, the detection of stochastic ordering of the distributions $F_1$ and $F_2$ is of primary interest

- For example, the alternative hypothesis $H_1$ may be that $F_{2j}(x) \leq F_{1j}(x)$ for each marginal cdf $F_{hj}$, $j = 1, \ldots, t$

- In this case, Wei & Lachin propose the statistic

$$z = \frac{e'T}{\sqrt{e'\widehat{\Sigma}_T e}},$$

  where $e'$ is the $t$-component vector $(1, \ldots, 1)$

- The asymptotic distribution of $z$ is $N(0, 1)$

- $H_0$ is rejected when $z$ is equal to a large positive (or large negative) value

# Wei-Johnson Method

- A class of two-sample nonparametric tests for incomplete repeated measures based on two-sample $U$-statistics

- The primary focus is on optimal methods of combining dependent tests

- Motivation:

  Suppose a researcher wishes to draw an overall conclusion regarding the superiority of one treatment over another (across time)

  A univariate one-sided test that combines the results at individual time points is more appropriate than an omnibus two-sided test of

$$H_0\colon F_1(x_1, \ldots, x_t) = F_2(x_1, \ldots, x_t)$$

## One-Sample $U$-Statistics

- Let $\mathcal{F}$ denote a family of cumulative distribution functions

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with cdf $F \in \mathcal{F}$

- Let $\gamma$ denote a parameter to be estimated

- $\gamma$ is *estimable of degree* $r$ for the family $\mathcal{F}$ if $r$ is the smallest sample size for which there exists a function $h(x_1, \ldots, x_r)$ such that

$$\mathrm{E}[h(X_1, \ldots, X_r)] = \gamma$$

  for every distribution $F \in \mathcal{F}$

- $h(x_1, \ldots, x_r)$ is a statistic that does not depend on $F$ and is called the *kernel* of the parameter $\gamma$

## One-Sample $U$-Statistics

- $h(x_1, \ldots, x_r)$ is assumed to be symmetric in its arguments, that is,

$$h(x_1, \ldots, x_r) = h(x_{\alpha_1}, \ldots, x_{\alpha_r})$$

  for every permutation $(\alpha_1, \ldots, \alpha_r)$ of the integers $1, \ldots, r$

- A one-sample *$U$-statistic* for the estimable parameter $\gamma$ of degree $r$ is created with the symmetric kernel $h(x_1, \ldots, x_r)$ by forming

$$U(X_1, \ldots, X_n) = \binom{n}{r}^{-1} \sum_{\beta \in B} h(X_{\beta_1}, \ldots, X_{\beta_r}),$$

  where $B = \left\{\beta \mid \beta \text{ is one of the } \binom{n}{r} \text{ unordered subsets of } r \text{ integers chosen without replacement from the set } \{1, \ldots, n\}\right\}$

## Example of a One-Sample $U$-Statistic

- Let $\mathcal{F}$ denote the class of all univariate distributions with finite first moment $\gamma$

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with cdf $F \in \mathcal{F}$

- Since $\mathrm{E}(X_1) = \gamma$, the mean $\gamma$ is an estimable parameter of degree 1 for the family $\mathcal{F}$

- Using the kernel $h(x) = x$, the $U$-statistic estimator of the mean is

$$U(X_1, \ldots, X_n) = \binom{n}{1}^{-1} \sum_{i=1}^{n} h(X_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$= \overline{X}$$

## One-Sample $U$-Statistic Theorem

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with cdf $F \in \mathcal{F}$

- Let $\gamma$ be an estimable parameter of degree $r$ with symmetric kernel $h(x_1, \ldots, x_r)$ and let

$$U(X_1, \ldots, X_n) = \binom{n}{r}^{-1} \sum_{\beta \in B} h(X_{\beta_1}, \ldots, X_{\beta_r})$$

- If $\mathrm{E}[h^2(X_1, \ldots, X_r)] < \infty$, and if

$$\zeta_1 = \mathrm{E}[h(X_1, \ldots, X_r)h(X_1, X_{r+1}, \ldots, X_{2r-1})] - \gamma^2$$

is positive, then

$$\sqrt{n}[U(X_1, \ldots, X_n) - \gamma]$$

has a limiting $N(0, r^2\zeta_1)$ distribution

- Hoeffding (1948), Randles and Wolfe (1979)

## Two-Sample $U$-Statistics

- Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from populations with cdf's $F(x)$ and $G(y)$, respectively, from a family of cumulative distribution functions $\mathcal{F}$

- A parameter $\gamma$ is estimable of degree $(r, s)$ for distributions $(F, G)$ in a family $\mathcal{F}$ if $r$ and $s$ are the smallest sample sizes for which there exists a function $h(x_1, \ldots, x_r, y_1, \ldots, y_s)$ such that

$$\mathrm{E}[h(X_1, \ldots, X_r, Y_1, \ldots, Y_s)] = \gamma$$

  for all distributions $(F, G) \in \mathcal{F}$

- $h(x_1, \ldots, x_r, y_1, \ldots, y_s)$ is called the two-sample kernel of the parameter $\gamma$

## Two-Sample $U$-Statistics

- The kernel $h(x_1, \ldots, x_r, y_1, \ldots, y_s)$ is assumed to be symmetric separately in its $x_i$ components and in its $y_i$ components

- A two-sample $U$-statistic for the estimable parameter $\gamma$ of degree $(r, s)$ is created with the kernel $h(x_1, \ldots, x_r, y_1, \ldots, y_s)$ by forming

$$U(X_1, \ldots, X_m, Y_1, \ldots, Y_n) =$$

$$\left[\binom{m}{r}\binom{n}{s}\right]^{-1} \sum_{\alpha \in A} \sum_{\beta \in B} h(X_{\alpha_1}, \ldots, X_{\alpha_r} Y_{\beta_1}, \ldots, Y_{\beta_s}),$$

where $A$ $(B)$ is the collection of subsets of $r$ $(s)$ integers chosen without replacement from the integers $\{1, \ldots, m\}$ $(\{1, \ldots, n\})$

- Note that sample sizes $m \geq r$, $n \geq s$ are required

## Example of a Two-Sample $U$-Statistic

- Let $\mathcal{F}$ be the class of univariate distributions with finite first moment $\gamma$

- Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from distributions with cdfs $F$ and $G$, respectively, where $F, G \in \mathcal{F}$

- Since $\mathrm{E}(X_1) = \mu_X$ and $\mathrm{E}(Y_1) = \mu_Y$, the mean diference $\gamma = \mu_Y - \mu_X$ is an estimable parameter of degree (1,1) for the family $\mathcal{F}$

- Using the kernel $h(x, y) = y - x$, the $U$-statistic estimator of the mean difference is

$$U(X_1, \ldots, Y_n) = \left[\binom{m}{1}\binom{n}{1}\right]^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} h(X_i, Y_j)$$

$$= \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (Y_j - X_i) = \overline{Y} - \overline{X}$$

**Two-Sample $U$-Statistic Theorem ($r = s = 1$)**

- Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from distributions with cdfs $F$ and $G$, respectively, where $F, G \in \mathcal{F}$

- Let $h(\cdot)$ be a symmetric kernel for an estimable parameter $\gamma$ of degree $(1,1)$ and let $U$ be the $U$-statistic estimator of $\gamma$

- Let $N = m + n$ and let $0 < \lambda = \lim_{N \to \infty} \dfrac{m}{N} < 1$

- Let $\zeta_{1,0} = \mathrm{E}[h(X_1, Y_1)h(X_1, Y_2)] - \gamma^2$ and let $\zeta_{0,1} = \mathrm{E}[h(X_1, Y_1)h(X_2, Y_1)] - \gamma^2$

- If $\mathrm{E}[h^2(X_1, Y_1)] < \infty$, and if

$$\sigma^2 = \frac{\zeta_{1,0}}{\lambda} + \frac{\zeta_{0,1}}{1 - \lambda} > 0,$$

the limiting distribution of $\sqrt{N}(U - \gamma)$ is $N(0, \sigma^2)$

## Joint Limiting Distribution of Correlated Two-Sample $U$-Statistics

- Special case of several two-sample $U$-statistics, each of degree (1,1)

- Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from distributions with $t$-variate cdfs $F$ and $G$, respectively

$$X_i' = (X_{i1}, \ldots, X_{it}), \quad Y_j' = (Y_{j1}, \ldots, Y_{jt})$$

- Let $U_1, \ldots, U_t$ be two-sample $U$-statistics with symmetric kernel $h(x, y)$, where $U_k$ estimates $\gamma_k$ of degree (1,1) and is given by

$$U_k = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} h(X_{ik}, Y_{jk}), \quad k = 1, \ldots, t$$

- Let $N = m + n$ and let $\lambda = \lim_{N \to \infty} \frac{m}{N}$

## Joint Limiting Distribution of Correlated Two-Sample $U$-Statistics

- The joint limiting distribution of

$$\sqrt{N}(U_1 - \gamma_1), \ldots, \sqrt{N}(U_t - \gamma_t)$$

is $t$-variate normal with zero mean vector and covariance matrix $\Sigma$ with elements

$$\sigma_{k,k'} = \frac{\zeta_{1(k,k')}}{\lambda} + \frac{\zeta_{2(k,k')}}{1-\lambda}$$

- The quantities $\zeta_{1(k,k')}$ and $\zeta_{2(k,k')}$ are given by

$$\begin{aligned}\mathrm{Cov}\big[\big(h(X_{1k}, Y_{1k}) - \gamma_k\big), \big(h(X_{1k'}, Y_{2k'}) - \gamma_{k'}\big)\big] \\ = \mathrm{E}[h(X_{1k}, Y_{1k})h(X_{1k'}, Y_{2k'})] - \gamma_k\gamma_{k'}\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}\big[\big(h(X_{1k}, Y_{1k}) - \gamma_k\big), \big(h(X_{2k'}, Y_{1k'}) - \gamma_{k'}\big)\big] \\ = \mathrm{E}[h(X_{1k}, Y_{1k})h(X_{2k'}, Y_{1k'})] - \gamma_k\gamma_{k'},\end{aligned}$$

for $k, k' = 1, \ldots, t$

## Wei-Johnson Class of Nonparametric Tests

- For each time point $j = 1, \ldots, t$, let

$$U_j = \frac{\sqrt{N}}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{i'=1}^{n_2} \delta_{1ij}\, \delta_{2i'j}\, \phi(y_{1ij}, y_{2i'j})$$

  - $N = n_1 + n_2$ is the total sample size
  - $y_{hij}$ is the observation from subject $i$ in group $h$ at time $j$, for $h = 1, 2$, $i = 1, \ldots, n_h$, and $j = 1, \ldots, t$
  - $\delta_{hij}$ is 1 if $y_{hij}$ is observed, 0 otherwise
  - $\phi(x, y)$ is a kernel function, e.g.,

$$\phi(x, y) = \begin{cases} 1 & \text{if } y > x, \\ 0 & \text{if } y = x, \\ -1 & \text{if } y < x. \end{cases}$$

$$\phi(x, y) = y - x$$

## Wei-Johnson Class of Nonparametric Tests

- If $\mathrm{E}[\phi^2(y_{1ij}, y_{2i'j}] < \infty$ and $n_1/(n_1+n_2) \to c$ $(0 < c < 1)$ as $N \to \infty$, then $U = (U_1, \ldots, U_t)'$ has an asymptotic null $N(0, \Sigma)$ distribution

- If $\mathrm{E}[\phi^4(y_{1ij}, y_{2i'j}] < \infty$, for $j = 1, \ldots, t$, the elements of the covariance matrix $\Sigma = (\sigma_{jk})$ can be consistently estimated by
$$\widehat{\sigma}_{jk} = \frac{N}{n_1}\,\widehat{\sigma}_{1jk} + \frac{N}{n_2}\,\widehat{\sigma}_{2jk},$$
where $\widehat{\sigma}_{1jk} = (n_1 n_2(n_2-1))^{-1} \times$
$$\sum_{i=1}^{n_1}\sum_{l \neq l'}^{n_2} \delta_{1ij}\delta_{1ik}\delta_{2lj}\delta_{2l'k}\,\phi(y_{1ij}, y_{2lj})\phi(y_{1ik}, y_{2l'k})$$
and $\widehat{\sigma}_{2jk} = (n_2 n_1(n_1-1))^{-1} \times$
$$\sum_{l=1}^{n_2}\sum_{i \neq i'}^{n_1} \delta_{1ij}\delta_{1i'k}\delta_{2lj}\delta_{2lk}\,\phi(y_{1ij}, y_{2lj})\phi(y_{1i'k}, y_{2lk})$$

## Test Statistics

- Let $\widehat{\Sigma}$ denote the estimated covariance matrix of the vector of test statistics $U$

- Since $U \approx N(0, \widehat{\Sigma})$, the hypothesis $H_0\colon F_1 = F_2$ can be tested against a general alternative using the statistic

$$Q = U'\widehat{\Sigma}^{-1}U,$$

  which is asymptotically $\chi_t^2$

- A univariate one-sided test that combines the results at individual time points can be based on the linear combination $w'U = \sum_{k=1}^{t} w_k U_k$

- Under $H_0$, the statistic

$$z = \frac{w'U}{\sqrt{w'\widehat{\Sigma}w}}$$

  is asymptotically $N(0, 1)$

## Choice of Weights

- The simplest choice is to weight each component equally, i.e. $w' = (1, \ldots, 1)$
  - Bloch and Moses (1988, *Amer. Statist.*) show that the use of equal weights often results in little loss of efficiency

- Another possibility is to weight by the reciprocals of the variances, i.e.,

$$w' = (1/\widehat{\Sigma}_{11}, \ldots, 1/\widehat{\Sigma}_{tt})$$

- Under the assumption that the test statistics at the individual time points are estimates of a common effect, the optimal weights are given by

$$w' = (1, \ldots, 1)\widehat{\Sigma}^{-1}$$

  - This assumption may not be reasonable

## Example

- Leppik et al. (1987) conducted a clinical trial in 59 epileptic patients
- Patients suffering from simple or complex partial seizures were randomized to receive either the antiepileptic drug progabide (31 patients) or a placebo (28 patients)
- At each of four successive postrandomization visits, the number of seizures occurring during the previous two weeks was reported
- The medical question of interest is whether or not progabide reduces the frequency of epileptic seizures

**Reference**

Leppik IE, Dreifuss FE, Porter R et al. (1987). A controlled study of progabide in partial seizures: methodology and results. *Neurology* **37**, 963–968.

## Example

- The Wei–Lachin vector of test statistics is $W' = (-0.4700, -0.0375, -0.2008, -0.3685)$ with estimated covariance matrix

$$\widehat{\Sigma}_W = \begin{pmatrix} 0.0788 & 0.0529 & 0.0460 & 0.0509 \\ 0.0529 & 0.0804 & 0.0538 & 0.0556 \\ 0.0460 & 0.0538 & 0.0789 & 0.0501 \\ 0.0509 & 0.0556 & 0.0501 & 0.0775 \end{pmatrix}$$

- The Wei–Lachin omnibus test of equality of distributions is $X_W^2 = W'\widehat{\Sigma}_W^{-1} W = 5.66$ with 4 df ($p = 0.23$)

- Using equal weights, the Wei–Johnson univariate statistic

$$\frac{c'U}{\sqrt{c'\widehat{\Sigma}_U c}},$$

with $c' = (1, \ldots, 1)$, is equal to $-1.09$

- The two-sided $p$-value is 0.14

**Example**

- A clinical trial comparing two treatments for maternal pain relief during labor
- 83 women in labor were randomized to receive an experimental pain medication (43 subjects) or placebo (40 subjects)
- Treatment was initiated when the cervical dilation was 8 cm
- At 30-minute intervals, the amount of pain was self-reported by placing a mark on a 100-mm line (0 = no pain, 100 = very much pain)
- The repeated pain scores are both nonnormal and incomplete
- Seems appropriate to compare treatments using the Wei–Lachin or the Wei–Johnson procedures

## Example

- Based on the data from minutes 30, 60, 90, 120, 150, and 180, the Wei–Lachin vector $W'$ and covariance matrix $\widehat{\Sigma}_W$ are, respectively,

$$(-0.394, -0.602, -0.755, -0.729, -0.497, -0.298)$$

$$\begin{pmatrix} .0794 & .0479 & .0284 & .0178 & .0114 & .0057 \\ .0479 & .0585 & .0316 & .0208 & .0155 & .0064 \\ .0284 & .0316 & .0368 & .0197 & .0111 & .0036 \\ .0178 & .0208 & .0197 & .0265 & .0148 & .0054 \\ .0114 & .0155 & .0111 & .0148 & .0132 & .0057 \\ .0057 & .0064 & .0036 & .0054 & .0057 & .0052 \end{pmatrix}$$

- Wei–Johnson $U'$ and covariance matrix $\widehat{\Sigma}_U$:

$$(-1.578, -2.410, -3.024, -2.918, -1.992, -1.192)$$

$$\begin{pmatrix} 1.3298 & 0.9268 & .6557 & .4182 & .2429 & .1433 \\ 0.9268 & 1.1120 & .7783 & .5576 & .3625 & .2114 \\ 0.6557 & 0.7783 & .9337 & .7511 & .4985 & .2555 \\ 0.4182 & 0.5576 & .7511 & .7790 & .5016 & .2528 \\ 0.2429 & 0.3625 & .4985 & .5016 & .4189 & .2234 \\ 0.1433 & 0.2114 & .2555 & .2528 & .2234 & .1819 \end{pmatrix}$$

## Example

- Standardized statistics at each time point:

| | Standardized Statistic | |
|---|---|---|
| Time point (minute) | Wei–Lachin | Wei–Johnson |
| 30 | −1.40 | −1.37 |
| 60 | −2.49 | −2.28 |
| 90 | −3.94 | −3.13 |
| 120 | −4.47 | −3.31 |
| 150 | −4.33 | −3.08 |
| 180 | −4.11 | −2.79 |

- At each time, pain scores are lower (better) in the experimental group

- Although both methods yield similar conclusions, the Wei–Lachin standardized statistic is larger in absolute value at every time point

**Example**

- Omnibus chi-square statistics for testing equality of distributions:
  - $X_W^2 = W'\widehat{\Sigma}_W^{-1}W = 30.1$ with 6 df, $p < 0.001$
  - $X_U^2 = U'\widehat{\Sigma}_U^{-1}U = 11.9$ with 6 df, $p = 0.065$
- Linear combinations:

| | Standardized Statistic | |
|---|---|---|
| | Wei–Lachin | Wei–Johnson |
| Equal weights | −3.88 | −3.06 |
| Reciprocals of variances | −4.85 | −3.28 |
| Optimal | −4.42 | −2.11 |

- With respect to the $N(0,1)$ reference distribution, all statistics indicate a significant difference between the two groups