Multivariate Analysis of Variance (MANOVA) for repeated measures: 1-sample case

- n repeated measures treated as $n \times 1$ response vector $\boldsymbol{y}_i$
- subjects with any missing y_{ij} (across time) are omitted from the analysis

Model

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{e}_i$$

- $\boldsymbol{\mu} = n \times 1$ mean vector for timepoints
- $\boldsymbol{e}_i = n \times 1$ vector of errors $\sim N(\mathbf{0}, \boldsymbol{\Sigma})$ in the population

- Notice, under univariate model assumptions:
  - $\boldsymbol{\Sigma} = \sigma_\pi^2 \mathbf{1}_n \mathbf{1}_n' + \sigma_e^2 \boldsymbol{I}_n$
  - $\boldsymbol{\mu} = \mu + \boldsymbol{\tau}$ (*i.e.*, grand mean and time effects)
  - $\Rightarrow$ univariate results can be extracted from multivariate model (and calculations)
- For one-sample case,
  - characterize timepoint vector $\boldsymbol{\mu}$
  - choose contrasts depending on structure and hypotheses of interest

Growth Curve Analysis - polynomial representation

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} t_1 \\ t_2 \\ . \\ . \\ t_n \end{bmatrix} \beta_1 + \begin{bmatrix} t_1^2 \\ t_2^2 \\ . \\ . \\ t_n^2 \end{bmatrix} \beta_2 + \ldots + \begin{bmatrix} t_1^{q-1} \\ t_2^{q-1} \\ . \\ . \\ t_n^{q-1} \end{bmatrix} \beta_{q-1}$$

$$\underset{n\times 1}{\boldsymbol{\mu}} = \underset{n\times q}{\boldsymbol{T}'} \; \underset{q\times 1}{\boldsymbol{\beta}}$$

- $t_1, t_2, \ldots, t_n$ represent timepoint values
- $q \leq n$ is the degree of the polynomial

Advantageous to orthogonalize $\boldsymbol{T}$

$\boldsymbol{\mu} = \boldsymbol{P}'\boldsymbol{\theta}$ where $\boldsymbol{P}$ is $q \times n$ matrix of orthogonal polynomials (including a row for the intercept or constant term)

- $\boldsymbol{P} = \boldsymbol{S}^{-1}\boldsymbol{T}$ and $\boldsymbol{S}\boldsymbol{S}' = (\boldsymbol{T}\boldsymbol{T}')$ Cholesky factorization
- $\boldsymbol{S}$ is a $q \times q$ lower triangular matrix (note: in SAS IML the `ROOT` function yields $\boldsymbol{S}'$ instead)

SAS IML example with 4 timepoints

```
TITLE 'producing orthogonal polynomial matrix';
PROC IML;
 time = { 1 1 1  1   ,
          0 1 2  3   ,
          0 1 4  9   ,
          0 1 8 27 } ;
orthpoly = INV(T(ROOT(time*T(time))))*time;
PRINT 'Time matrix', time [FORMAT = 8.4];
PRINT 'Orthonormalized', orthpoly [FORMAT=8.4];
```

Output

```
Time matrix
   TIME
 1.0000 1.0000 1.0000  1.0000
 0.0000 1.0000 2.0000  3.0000
 0.0000 1.0000 4.0000  9.0000
 0.0000 1.0000 8.0000 27.0000


Orthonormalized
   ORTHPOLY
  0.5000  0.5000  0.5000 0.5000
 -0.6708 -0.2236  0.2236 0.6708
  0.5000 -0.5000 -0.5000 0.5000
 -0.2236  0.6708 -0.6708 0.2236
```

Orthogonal Polynomial Trend Model

$$\boldsymbol{P}\boldsymbol{y}_i \;=\; \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{P}\boldsymbol{e}_i \;=\; \boldsymbol{\theta} + \boldsymbol{e}_i^*$$

- *e.g.*,

$$\boldsymbol{P}_4 \;=\; \begin{bmatrix} 1 & 1 & 1 & 1 & \vdots & \div\sqrt{4} \\ -3 & -1 & 1 & 3 & \vdots & \div\sqrt{20} \\ 1 & -1 & -1 & 1 & \vdots & \div\sqrt{4} \\ -1 & 3 & -3 & 1 & \vdots & \div\sqrt{20} \end{bmatrix} \begin{matrix} \boldsymbol{p}_0 \\ \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \boldsymbol{p}_3 \end{matrix}$$

- $\boldsymbol{\theta} = n \times 1$ vector of transformed population means with its least squares estimate given by the transformed sample mean vector, namely $\hat{\boldsymbol{\theta}} = \boldsymbol{P}\bar{\boldsymbol{y}}_.$

- $\boldsymbol{e}_i^* \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^* = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}')$
  - note: univariate ANOVA assumes $\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}'$ is diagonal with equal values below the first element (since $\boldsymbol{P}$, as defined above, includes the zero-order term, *i.e.*, the grand mean)

MANOVA matrices

source	SSCP ($n \times n$)	E(SSCP)
Time	$\text{SST}^* = N\boldsymbol{P}\bar{\boldsymbol{y}}_.\bar{\boldsymbol{y}}_.'\boldsymbol{P}'$	$\boldsymbol{P}\left[\boldsymbol{\Sigma} + N\boldsymbol{\mu}\boldsymbol{\mu}'\right]\boldsymbol{P}'$
Residual	$\text{SSR}^* = \boldsymbol{P}\ \text{SSR}\ \boldsymbol{P}'$ $= \boldsymbol{P}(\boldsymbol{Y}'\boldsymbol{Y} - N\bar{\boldsymbol{y}}_.\bar{\boldsymbol{y}}_.')\boldsymbol{P}'$	$(N-1)\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}'$

- $\boldsymbol{Y} = N \times n$ matrix of all the data
- $\bar{\boldsymbol{y}}_. = n \times 1$ vector of timepoint means

SST*

- first diagonal element $= Nn\bar{y}_{..}^2 =$ function of grand mean
- other $n-1$ diagonal elements $=$ orthogonal polynomial decomposition of Time SS $= N \Sigma_{j=1}^{n}(\bar{y}_{.j} - \bar{y}_{..})^2$

SSR*

- first diagonal element $=$ Subject SS $= n \Sigma_{i=1}^{N}(\bar{y}_{i.} - \bar{y}_{..})^2$
- other $n-1$ diagonal elements $=$ orthogonal polynomial decomposition of Error (*i.e.*, Subject by Time) SS

Orthogonal polynomial partition of sum of squares and products (1-sample case)

	diag element
$$\text{SST}^* = \begin{bmatrix} \text{sst}_0 & & & & \\ & \text{sst}_1 & & & \\ & & \text{sst}_2 & & \\ & & & \cdots & \\ & & & & \text{sst}_{n-1} \end{bmatrix}$$	constant linear time quad time $\ldots$ $(n-1)$th time
$$\text{SSR}^* = \begin{bmatrix} \text{ssr}_0 & & & & \\ & \text{ssr}_1 & & & \\ & & \text{ssr}_2 & & \\ & & & \cdots & \\ & & & & \text{ssr}_{n-1} \end{bmatrix}$$	subjects subj $\times$ lin subj $\times$ quad $\ldots$ subj $\times$ $(n-1)$

Note: these are symmetric matrices, however the diagonals contain all of the information of interest in SS calculations

Extracting univariate repeated measures ANOVA

Source	df	SS	MS	E(MS)
Subjects	$N-1$	$\text{SS}_S = n\,\Sigma_{i=1}^{N}(\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{\text{SS}_S}{N-1}$	$\sigma_e^2 + n\sigma_\pi^2$
Time	$n-1$	$\text{SS}_T = N\,\Sigma_{j=1}^{n}(\bar{y}_{.j} - \bar{y}_{..})^2$	$\frac{\text{SS}_T}{n-1}$	$\sigma_e^2 +$ $N\,\Sigma(\tau_j - \tau_.)^2$
Residual	$(N-1)$ $\times(n-1)$	$\text{SS}_R = \Sigma_{i=1}^{N}\Sigma_{j=1}^{n}\,y_{ij}$ $-\bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$\frac{\text{SS}_R}{(N-1)(n-1)}$	σ_e^2
total	$Nn-1$	$\text{SS}_y = \Sigma_{i=1}^{N}\Sigma_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$		

- $\text{SS}_S = \text{ssr}_0$ from SSR* matrix
- $\text{SS}_T = \Sigma$ lower $n-1$ SST* diagonal elements
- $\text{SS}_R = \Sigma$ lower $n-1$ SSR* diagonal elements

notice:

$$\text{MS}_R = \frac{\Sigma \text{ lower } n-1 \text{ SSR}^* \text{ diagonal elements}}{(N-1)(n-1)}$$
$$= \frac{\text{average } n-1 \text{ SSR}^* \text{ diagonal elements}}{(N-1)}$$

$\Rightarrow$ *sometimes called "averaged" tests*

Multivariate test of Time effect

- $H_0 : \boldsymbol{\mu}$ elements all equal $\iff \quad H_0 : \boldsymbol{\tau} = \mathbf{0}$

- Extract and compare lower $(n-1) \times (n-1)$ submatrices of SST* and SSR*

- Solve 2-matrix eigenproblem:

$$| \, \mathrm{SST}^*_{(n-1)} \; - \; \lambda \, \mathrm{SSR}^*_{(n-1)} | \; = \; 0$$

note: $\lambda = 1 \quad \textbf{if} \quad \mathrm{SST}^*_{(n-1)} = \mathrm{SSR}^*_{(n-1)}$

Motivation for 2-matrix eigenproblem

– eigenvalue = root of a matrix equation

– note: in scalar terms, we could write

$$\mathrm{MS}_T - F\ \mathrm{MS}_R = 0 \quad \Rightarrow \quad F = \ \mathrm{MS}_T /\ \mathrm{MS}_R$$

– this is akin to what we have in multivariate terms

$$|\ \mathrm{SST}^*_{(n-1)}\ -\ \lambda\ \mathrm{SSR}^*_{(n-1)}|\ =\ 0$$

– λ is the root of this equation

– function of λ yields multivariate F statistic

Simplifying the problem

- Use Cholesky factorization $\text{SSR}^*_{(n-1)} = \boldsymbol{E}\boldsymbol{E}'$ to yield a 1-matrix eigenproblem (remember SAS IML gives upper-triangular matrix $\boldsymbol{E}'$)

$$|\boldsymbol{E}^{-1}\,\text{SST}^*_{(n-1)}(\boldsymbol{E}^{-1})' - \lambda \boldsymbol{I}_{(n-1)}| = 0$$

- Multivariate test statistics
  - Roy's largest root statistic: eigenvalue λ
  - Wilk's Lambda: $\Lambda = 1/(1+\lambda)$
  - functions of these test statistics approximately follow F-distribution (under null hypothesis), though sometimes interpolation is necessary (fractional df)

Tests of specific time elements

- Univariate: extract lower $n-1$ diagonal elements of SST* and use MS_R as denominator

$$F_1 = \frac{\mathrm{sst}_1}{\Sigma_{j=1}^{n-1} \ \mathrm{ssr}_j/[(n-1)(N-1)]} \quad \text{linear}$$

$$F_2 = \frac{\mathrm{sst}_2}{\Sigma_{j=1}^{n-1} \ \mathrm{ssr}_j/[(n-1)(N-1)]} \quad \text{quadratic}$$

$$\cdot \quad \cdot \quad \cdot \qquad\qquad\qquad\qquad \cdot$$

$$F_{n-1} = \frac{\mathrm{sst}_{n-1}}{\Sigma_{j=1}^{n-1} \ \mathrm{ssr}_j/[(n-1)(N-1)]} \quad (n-1)\text{th}$$

- Multivariate: extract and compare lower $n-1$ diagonal elements of SST* and SSR*

$$
\begin{aligned}
F_1 &= \frac{\mathrm{sst}_1}{\mathrm{ssr}_1/(N-1)} && \text{linear} \\
F_2 &= \frac{\mathrm{sst}_2}{\mathrm{ssr}_2/(N-1)} && \text{quadratic} \\
&\cdots && . \\
F_{n-1} &= \frac{\mathrm{sst}_{n-1}}{\mathrm{ssr}_{n-1}/(N-1)} && (n-1)\text{th}
\end{aligned}
$$

Example 3a: Analysis of vocabulary data using MANOVA (SAS code and output)
http://tigger.uic.edu/~hedeker/bockvoc2.txt

OPTIONAL

Example 3b: Analysis of sleep data using IML to do MANOVA (SAS code and output)
http://tigger.uic.edu/~hedeker/bockvoc3.txt

MANOVA of repeated measures - s sample case

- $h = 1, \ldots, s$ groups
- $i = 1, \ldots, N_h$ subjects in group h
- $j = 1, \ldots, n$ timepoints
- $N = \sum N_h$ total number of subjects

$$\boldsymbol{y}_{hi} = \boldsymbol{\mu} + \boldsymbol{\gamma}_h + \boldsymbol{e}_{hi}$$

- $\boldsymbol{\mu} = n \times 1$ vector mean for timepoints
- $\boldsymbol{\gamma}_h = n \times 1$ vector effect for the population from with the hth group of subjects was drawn
- $\boldsymbol{e}_{hi} = n \times 1$ vector of errors $\sim N(\mathbf{0}, \mathbf{\Sigma})$ in each of the populations
  - homogeneity of variance-covariance across groups

Again, with orthogonal transformation for time effects

$$\boldsymbol{P}\boldsymbol{y}_{hi} = \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{P}\boldsymbol{\gamma}_h + \boldsymbol{P}\boldsymbol{e}_{hi}$$

$$\boldsymbol{e}^*_{hi} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^* = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}')$$

test $\boldsymbol{\Sigma}^*$ for sphericity to use univariate "averaged" tests

MANOVA matrices

source	SSCP $(n \times n)$
Time	$\text{SST}^* = \boldsymbol{P}\ \text{SST}\ \boldsymbol{P}' = N\boldsymbol{P}\bar{\boldsymbol{y}}_{..}\bar{\boldsymbol{y}}'_{..}\boldsymbol{P}'$
Group	$\text{SSG}^* = \boldsymbol{P}\ \text{SSG}\ \boldsymbol{P}' = \boldsymbol{P}(\Sigma_h N_h \bar{\boldsymbol{y}}_{h.}\bar{\boldsymbol{y}}'_{h.} - \text{SST})\boldsymbol{P}'$
Residual	$\text{SSR}^* = \boldsymbol{P}\ \text{SSR}\ \boldsymbol{P}' = \boldsymbol{P}(\text{SSY} - \text{SSG} - \text{SST})\boldsymbol{P}'$
Total	$\text{SSY}^* = \boldsymbol{P}\ \text{SSY}\ \boldsymbol{P}' = \boldsymbol{P}(\Sigma_h \Sigma_i \boldsymbol{y}_{hi}\boldsymbol{y}'_{hi})\boldsymbol{P}'$

Results only depend on summary statistics

- Cross-product matrix from overall mean vector
  $\bar{\boldsymbol{y}}_{..}\bar{\boldsymbol{y}}'_{..}$

- Sum of cross-product matrices from group mean vectors
  $\Sigma_h N_h \bar{\boldsymbol{y}}_{h.}\bar{\boldsymbol{y}}'_{h.}$

- sum of cross-product matrices from subject data vectors
  $\Sigma_h \Sigma_i \boldsymbol{y}_{hi}\boldsymbol{y}'_{hi}$

Orthogonal Polynomial partition of sum of squares and products (s-sample case) Diagonal elements

	diag element
$$\text{SST}^* = \begin{bmatrix} \text{sst}_0 & & & & \\ & \text{sst}_1 & & & \\ & & \text{sst}_2 & & \\ & & & \cdots & \\ & & & & \text{sst}_{n-1} \end{bmatrix}$$	constant linear time quad time ... $(n-1)$th time
$$\text{SSG}^* = \begin{bmatrix} \text{ssg}_0 & & & & \\ & \text{ssg}_1 & & & \\ & & \text{ssg}_2 & & \\ & & & \cdots & \\ & & & & \text{ssg}_{n-1} \end{bmatrix}$$	groups grps $\times$ lin grps $\times$ quad ... grps $\times$ $(n-1)$
$$\text{SSR}^* = \begin{bmatrix} \text{ssr}_0 & & & & \\ & \text{ssr}_1 & & & \\ & & \text{ssr}_2 & & \\ & & & \cdots & \\ & & & & \text{ssr}_{n-1} \end{bmatrix}$$	subjects s(g) $\times$ lin s(g) $\times$ quad ... s(g) $\times$ $(n-1)$

Extracting univariate ANOVA results

		source
SS Time	$= \text{sst}_1 + \text{sst}_2 + \ldots + \text{sst}_{n-1}$	SST*
SS Group	$= \text{ssg}_0$	SSG*
SS GT	$= \text{ssg}_1 + \text{ssg}_2 + \ldots + \text{ssg}_{n-1}$	SSG*
SS Subj (w/in groups)	$= \text{ssr}_0$	SSR*
SS Residual	$= \text{ssr}_1 + \text{ssr}_2 + \ldots + \text{ssr}_{n-1}$	SSR*

For denominator:

- Time, G by T, & Subj: SS Residual$/(N-s)(n-1)$

  $\Rightarrow$ average of $n-1$ residual terms

- Group: $\text{ssr}_0 \; /(N-s)$

Multivariate model

- Group test is the same
- Each time-related term has its own denominator
  - no averaging to obtain one error term

e.g., after multivariate test of Group by Time, then

$$F_{\mathrm{GT}_1} = \frac{\mathrm{ssg}_1/(s-1)}{\mathrm{ssr}_1/(N-s)} \quad \text{group by linear}$$

$$F_{\mathrm{GT}_2} = \frac{\mathrm{ssg}_2/(s-1)}{\mathrm{ssr}_2/(N-s)} \quad \text{group by quadratic}$$

$$\cdot \quad \cdot \quad \cdot \qquad\qquad\qquad \cdot$$

$$F_{\mathrm{GT}_{n-1}} = \frac{\mathrm{ssg}_{n-1}/(s-1)}{\mathrm{ssr}_{n-1}/(N-s)} \quad \text{group by } (n-1)\text{th}$$

each on $s-1$ and $N-s$ degrees of freedom

Multivariate test of Group by Time

- Extract $(n-1) \times (n-1)$ submatrices of SSG* and SSR*

- Solve 2-matrix eigenproblem for $\min(s-1, n-1)$ roots:

$$|\ \text{SSG}^*_{(n-1)} \ - \ \boldsymbol{\lambda}\ \text{SSR}^*_{(n-1)}| \ = \ 0$$

  – Roy's largest root statistic = 1st eigenvalue λ_1

  – Wilk's Lambda $\Lambda \ = \ \Pi_{h=1}^{s-1} \ 1 \, / \, (1 + \lambda_h)$

- same test can be made for $n - q - 1$ terms if only a $q < n$ degree trend is considered

  – *e.g.*, only test for linear and quad trends even if $n > 3$

If Group by Time is not significant, may want to pool Group $\times$ Time into Residual for (multivariate) testing of Time

- pooling not easy to do with software
- pooling increases power, but if interaction df are small relative to residual df then loss in power by not pooling is not appreciable

Example 3c: MANOVA analysis of Prozac weight data. Includes use of CONTRAST statement to obtain estimates of a-priori contrasts for the multi-group situation. (SAS code and output)
http://tigger.uic.edu/~hedeker/prozwtc.txt

More General Model

- A and B are between-subjects factors (*e.g.*, sex and group)
- T is the within-subjects factor (time)

within-subjects	between-subjects I (1)	A $(a-1)$	B $(b-1)$	AB $(a-1)(b-1)$
I (1)	grand mean	sex	group	sex $\times$ group
T $(n-1)$	time	sex $\times$ time	group $\times$ time	sex $\times$ group $\times$ time

$\Rightarrow$ may want to test time-related terms with less than $(n-1)$ df, especially with orthogonal polynomials (hard to interpret polynomials beyond cubic)

Design on repeated measures

e.g., two-period crossover study

period 1		period 2	
drug	placebo	drug	placebo
drug	placebo	placebo	drug
placebo	drug	drug	placebo
placebo	drug	placebo	drug

WS factors: Period and Treatment

BS factor: Order

within-subjects	between-subjects I(1)	Order $(c-1)$
I (1)	grand mean	order
Treatment $(a-1)$	treatment	order $\times$ treatment
Period $(b-1)$	period	order $\times$ period
Group $\times$ Period $(a-1)(b-1)$	group $\times$ period	order $\times$ treatment $\times$ period

with $n-1=(a-1)+(b-1)+(a-1)(b-1)$
where n = total number of timepoints

Summary - ANOVA and MANOVA approches

Pros

- well-understood and developed methods
- readily available software
- results are based on relatively simple calculations (non-iterative)

Cons

- ANOVA assumes sphericity
- MANOVA allows general $\mathbf{\Sigma}$, but no missing data
- categorical treatment of time
- handling of covariates (continuous-type) is limited, especially for time-varying covariates

More Recent Methods for Longitudinal Data

- representation as regression model, $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}$
  - more flexibility for continuous predictors
  - time can be treated continuously (or not)
  - easy to include time-varying predictors
- Maximum likelihood (ML) estimation (or variant of ML estimation) for unbalanced data
  - **subjects can be missing across time**

- variety of forms for $\boldsymbol{\Sigma}$ ($= V(\boldsymbol{y})$)
  - mixed-effects models: add multiple random subject effects, $\boldsymbol{Z}_i\boldsymbol{v}_i$, to model, thus $\boldsymbol{\Sigma} = \boldsymbol{Z}\boldsymbol{\Sigma}_{\upsilon}\boldsymbol{Z}' + \sigma^2\boldsymbol{I}$
    - *e.g.*, random subject intercepts and time trends
  - covariance pattern models for parsimonious modeling of $\boldsymbol{\Sigma}$
    - *e.g.*, unstructured, AR(1), or banded
  - hybrid: mixed-effects models with autocorrelated errors, $\boldsymbol{\Sigma} = \boldsymbol{Z}\boldsymbol{\Sigma}_{\upsilon}\boldsymbol{Z}' + \sigma^2\boldsymbol{\Omega}$
  - $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}$ is right model, but tests are incorrect assuming $\boldsymbol{e}$ are independent
    - get "robust" standard errors for tests of $\boldsymbol{\beta}$ via generalized estimating equations (GEE)