Hypothesis testing in MANOVA

Steven A. Juliano & Joseph E. Fader
School of Biological Sciences
MANOVA

What is it?

• Multivariate (>1 dependent variable) tests for differences among groups
• ANOVA is a special case of MANOVA
• A very useful reference:
• NOTE: my examples use SAS 9.2
Do (medical) entomologists use MANOVA?

- Last 2 issues of 2009
- *Journal of Medical Entomology* (75 papers)
- Comparison: *Ecology* (54 papers)
Why you need MANOVA

• Measure >1 dependent variable
  – multiple correlated responses
• Probability of any type I errors increases with number of variables
• MANOVA provides a joint test for any significant effects among a set of variables at 1 legitimate $\alpha$
Why you need MANOVA

- MANOVA tests for **patterns**
- ANOVA tests for effects on individual variables
Why you need MANOVA

• MANOVA often more powerful than ANOVA
  – greater chance to detect effects

• However MANOVA ...
  – Power can be reduced by irrelevant variables
  – tests linear combinations of variables
    • biology may dictate other combinations of variables
Basis of MANOVA
(skipping lots of detail)

• MANOVA
  – COVARIANCE\textsubscript{TOTAL} = total covariation among Y s
    = WITHIN COVARIANCE MATRIX +
    BETWEEN COVARIANCE MATRIX
  – Partitioning covariance matrix
  – Tests constructed as ratios of between / within covariance matrix estimates
Basis of MANOVA

(skipping lots of detail)

• MANOVA
  – Eigen vectors: linear combinations of the original variables
  – \( E_i = a + bY_1 + cY_2 + dY_3 + \ldots \)
  – 1\(^{st}\) Eigen Vector maximizes variance between groups for the resulting value
  – Eigen value: amount of total variation accounted for by Eigen vector
  – Subsequent Eigen vectors orthogonal (i.e., perpendicular) to all previous
MANOVA test statistics

- Two common multivariate tests
  - Wilk’s Λ [most commonly used]
  - Pillai’s Trace [robust to violations of assumptions]
- usually give same result (identical for 2 groups)
MANOVA

• Data requirements
  – Multivariate normality
  – Homogeneous covariance matrix
  – cases with missing values of $Y_j$ deleted
  – inference depends on relatively large sample size
    • 20 / group
    • 20 * number of variables
Significant MANOVA
now what?

• Which groups differ?
  – multivariate question
  – homologous to ANOVA follow-up

• Which variables contribute most to any difference?
  – new kind of question
Significant MANOVA

• Multivariate contrasts
  – CONTRAST statement comes before MANOVA statement
  – Also for pairwise comparisons
  – correction for multiple tests (e.g., Bonferroni)

• MANOVA Statement
  – Use “... H=ALL ...” option
Significant MANOVA

• Which variables contribute to the difference?
• Two approaches
  – Univariate: follow MANOVA with univariate ANOVAs on each variable
  – Multivariate
Significant MANOVA

- Multivariate: follow MANOVA with **Canonical Variate Analysis**
- Canonical variates
  - Eigen vectors scaled to unit variance
  - Standardized canonical coefficients describe contribution of each dependent variable to a function describing differences among groups.
Canonical coefficients

• Potential problems
  – Depend on which $Y$ variables are included
    • different results & interpretation if different variables are omitted
  – Fewer readers immediately know what you are doing
    • cite Scheiner
  – “greater contribution” is subjective
Example:
Effect of aggregation on competition

- *Aedes albopictus* (dominant competitor)
- *Aedes aegypti* (poorer competitor)
Effect of aggregation on competition

• Theory: As dominant competitor becomes more aggregated in space, effect of competition on poorer competitor declines
• Ideal: measure effect of aggregation on \( dN/N_{dt} \)
• Practical: MANOVA on life history correlates of \( dN/N_{dt} \)
  – Survivorship
  – Adult female size
  – Development time
Effect of aggregation on competition

- Replicate = 8 containers
- Cohort = 80 Larvae
- Standard food, temperature
- Determine response of *Aedes aegypti*
## MANOVA

### part 1. Basic MANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Df, Df</th>
<th>Pillai’s Trace</th>
<th>P</th>
<th>Eigen value</th>
<th>% Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>3, 28</td>
<td>0.855</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>12, 90</td>
<td>1.317</td>
<td>&lt;0.0001</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>9.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>0.12</td>
</tr>
<tr>
<td>Interaction</td>
<td>12,90</td>
<td>0.207</td>
<td>0.8721</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MANOVA

part 2. Which variables contribute to effects?

<table>
<thead>
<tr>
<th>Source</th>
<th>Eigen value</th>
<th>% Var.</th>
<th>Prop. Surviving</th>
<th>Wing length</th>
<th>Devel. time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; 9.30 94%</td>
<td>-0.21</td>
<td>-2.92</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; 0.44 4%</td>
<td>1.12</td>
<td>-0.24</td>
<td>-0.09</td>
<td></td>
</tr>
</tbody>
</table>
MANOVA
part 3. comparing groups

Female median time to adult (d)
Female mean wing Length (mm)
Medium aggregation
No Aggregation
High aggregation
Low aggregation
Control
A
B
C

Control

Medium aggregation

High aggregation

Low aggregation

No Aggregation

C

B

A
MANOVA
part 3. comparing groups

Proportion survivorship vs. Female mean wing length (mm)
Is MANOVA best?

• MANOVA combines variables in linear combinations
• Biological hypotheses may predict effects on nonlinear, non-additive combinations
• Biological meaning should take precedence over statistical convenience
Effect of aggregation on competition

• Theory: predicts effect of aggregation on $dN/N_{dt}$
• Estimate from demographic measurements
• Do MANOVA and estimated $dN/N_{dt}$ yield same conclusions?
Estimating rate of increase from demography

\[
\frac{dN}{Ndt} = \frac{\ln \sum_{x=1} \ln \frac{R_0}{T_c}}{\ln \left( \frac{\sum_{x=1} l_x m_x}{\sum_{x=1} l_x m_x} \right)} = \frac{\ln(R_0 \big)}{T_c}
\]

\(N\) = number of females

\(l_x\) = probability that female survives to day \(x\)

\(m_x\) = number of female offspring per female on day \(x\)

\(R_0\) = basic reproductive rate

= expected number of surviving offspring per female

\(T_c\) = cohort generation time

= mean time between birth of mother and birth of young
Index of performance (est. rate of increase)


\[ r' = \frac{\ln \left( \frac{1}{N_0} \sum_{x=1} A_x f(w_x) \right)}{D + \left[ \frac{\sum_{x=1} x A_x f(w_x)}{\sum_{x=1} A_x f(w_x)} \right]} \]

- \( N_0 \) = initial number of females (assumed \( \frac{1}{2} \) cohort)
- \( A_x \) = number females eclosing on day \( x \)
- \( x \) = days since hatching of cohort
- \( w_x \) = mean size of females eclosing on day \( x \)
- \( f(w_x) \) = function predicting female eggs based on size \( w_x \)
- \( D \) = days from eclosion to oviposition
$r'$ from the aggregation experiment

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>1</td>
<td>36.48</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>4</td>
<td>16.46</td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td>Block * Treatment</td>
<td>4</td>
<td>0.26</td>
<td>0.9027</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANOVA on $r'$ index of population performance
Conclusions

• Analysis of biologically-derived index vs. MANOVA yield similar, not identical conclusions
  – MANOVA: linear combinations, based on statistics
  – Index: nonlinear combinations, based on biology
  – When possible, using biologically-derived synthesis of multiple variables is desirable
  – In the absence of *a priori* synthesis of variables, MANOVA desirable
Acknowledgements

- Joe Fader
- Paul O’Neal
- Ebony Murrell
- Colleen Stephens
- Jen Breaux
- Scott Chism
- Stefani Brandt
- L. Philip Lounibos
- George F. O’Meara
- Cynthia Lord
TITLE 'MULTIVARIATE ANALYSIS OF PERFORMANCE VARIABLES';
proc glm data=fader1 /* GLM for unbalance designs */;
   class block treat /* treat = treatments */;
   model psurv meaned meanwl = block treat block*treat / ss3 /* type III sums of squares for unbalanced designs */;
   lsmeans treat block*treat / stderr pdiff;
   contrast 'con vs uniform' treat 1 -1 0 0 0 /* contrast statements test pairwise differences */;
   contrast 'con vs half' treat 1 0 -1 0 0;
   contrast 'con vs quart' treat 1 0 0 -1 0;
   contrast 'con vs one' treat 1 0 0 0 -1;
   contrast 'unif vs half' treat 0 1 -1 0 0;
   contrast 'unif vs quart' treat 0 1 0 -1 0;
   contrast 'unif vs one' treat 0 1 0 0 -1;
   contrast 'half vs quart' treat 0 0 1 -1 0;
   contrast 'half vs one' treat 0 0 1 0 -1;
   contrast 'quart vs one' treat 0 0 0 1 -1;
   manova h=_ALL_ / canonical /* yields multivariate analysis & canonical coefficients */;
run;
References for MANOVA

