Multivariate Analysis

BIOL 4943/5943

Who: Michael A. Patten (Oklahoma Biological Survey, room 236; mpatten@ou.edu)
When: Tuesdays and Thursdays, 10:30–11:45 (a.m.)
Where: Richards Hall, room 304

Textbook:


There will also other readings assigned through the course.

It’s easy to lie with statistics, but it’s hard to tell the truth without them. – Andrejs Dunkels

The goal of this course is to present students with a basic background in various types of multivariate statistics that are used commonly (and thus appear commonly in the scientific literature). The course will be retrospective, not comprehensive. There are numerous topics in multivariate statistics, far too many to cover in a single semester, but statistical literacy does not require complete knowledge of all techniques. Rather, knowing enough about basic groups of methods (e.g., classification vs. data reduction vs. ordination) is often sufficient to at least ask the right questions and begin an analysis properly.

The course will be structured around lecture, but the format will be open, in that I very much encourage and welcome participation—share your questions and experience with us. Often the best strategy to gain a solid understanding of a “scary” topic like statistics is to be unafraid to ask questions, to seek additional clarification, and to share ideas and thoughts. To that end, I will stress understanding of concepts and principles over understanding (typically rote memorization) of the mathematics behind the principles. Some delving into mathematics will be necessary, of course, but, to borrow a phrase from a well-known series of texts, my aim is to make this class a “gentle introduction” to a complex and fascinating topic.

One of the ways in which we will gain a deeper understanding of the statistics we go over in lectures is to read and discuss papers that deal with that technique. As your confidence grows in your ability to decipher the statistics, you will find quickly that you are able to approach published papers more critically and constructively.

Because the emphasis will be on common techniques, we will spend a seemingly inordinate amount of time of certain statistics that are used widely. This will be especially true of principal components analysis (PCA), what might well be the most commonly used truly multivariate stat. MANOVA, multiple regression (technically a univariate stat—you will come to understand why), and discriminant analysis (DA) are also used widely, so we will spend extra time on each.
**Grading**

The nature of this course makes a student's participation in discussion a critical part of his or her grade. To facilitate classroom discussion, students may be assigned the task of leading the debate, although just because a student does not leading the discussion in a given week is not an excuse to avoid participation. I will assign occasional quizzes and short exams in class. Scores on individual assignments will constitute a sizeable proportion of a student's final grade. Analytical assignments will be structured around a standard data set I develop but will require the student to use a statistical package of his or her choice (e.g., SAS, R, SPSS, Systat, Excel, etc.). The nature of course is such that in-class discussion will focus on statistical concepts not on the mechanics of any particular software package.

I will use the following point system as a guide for grading:

- **Undergraduates**
  - assignment points
    - classroom participation – 100
    - scientific paper reviews (3) – 75
    - analytical assignments (4) – 150
    - quizzes (3) – 75
  - total points: 400

- **Graduate Students**
  - assignment points
    - classroom participation – 100
    - scientific paper reviews (3) – 75
    - analytical assignments (4) – 150
    - quizzes (5) – 125
    - data analysis paper – 200
  - total points: 650

The nature of this course makes absences costly. Be warned that missed assignments typically cannot be completed later. It should go without saying, but a student’s grade will suffer drastically if he or she misses multiple classes.

Note: Because this course is cross-listed, any given class may be a mix of graduate and undergraduate students. Prerequisites for the course are the same for either group, but under the assumption that undergraduates have had less exposure to statistical analysis, chiefly because they likely have read fewer scientific papers, they will receive special dispensation on quizzes and assignments (e.g., there may be particular questions that only graduate students will be expected to answer). The point totals will be scaled accordingly.

★★★ IMPORTANT NOTICES ★★★

I will tolerate neither plagiarism nor academic misconduct. Please consult OU’s website (http://www.ou.edu/provost/integrity/) to learn the university’s stance on these policies. Any student guilty of plagiarism—including self-plagiarism (turning in your own work for more than one course) and copying information directly from a web site—will receive no credit for that assignment and may fail the course.

Please notify me if for any reason you require special accommodations to ensure that you participate fully in this course: this course will be made accessible to any student with a disability.
# Syllabus – Multivariate Analysis

**BIOLOGY 4943/5943**

<table>
<thead>
<tr>
<th>WEEK</th>
<th>TOPIC(S)</th>
<th>READING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 14 &amp; 16 Jan</td>
<td>introduction, philosophy</td>
<td>T&amp;F Ch. 1; Mecklin &amp; Mundfrom (2004)</td>
</tr>
<tr>
<td></td>
<td>general multivariate techniques</td>
<td></td>
</tr>
<tr>
<td>2. 21 &amp; 23 Jan</td>
<td>canonical correlation analysis and</td>
<td>T&amp;F Ch. 12; Rotenberry et al. (1996)</td>
</tr>
<tr>
<td>3. 28 &amp; 30 Jan</td>
<td>multiple regression</td>
<td>[T&amp;F Ch. 5]; McArule (1988)</td>
</tr>
<tr>
<td>4. 4 &amp; 6 Feb</td>
<td>Mantel test</td>
<td>Sokal (1979); Douglas &amp; Endler (1982)</td>
</tr>
<tr>
<td>5. 11 &amp; 13 Feb</td>
<td>MANOVA</td>
<td>T&amp;F Ch. 7; Anderson (2001)</td>
</tr>
<tr>
<td>6. 18 &amp; 20 Feb</td>
<td>profile analysis</td>
<td>T&amp;F Ch. 8; Simms &amp; Burdick (1988); Beachy (1997)</td>
</tr>
<tr>
<td>7. 25 &amp; 27 Feb</td>
<td>principal components analysis</td>
<td>T&amp;F Ch. 13; Rexstad et al. (1988); Jackson (1993)</td>
</tr>
<tr>
<td>8. 4 &amp; 6 Mar</td>
<td>PCA continued</td>
<td></td>
</tr>
<tr>
<td>9. 11 &amp; 13 Mar</td>
<td>cluster analysis</td>
<td>Everitt (1979); Boyce et al. (1999); Plotkin et al. (2002)</td>
</tr>
</tbody>
</table>

★★★ spring break ★★★

11. 25 & 27 Mar | regression trees | Roff & Roff (2003); Patten et al. (2011) |
| 12. 1 & 3 Apr | discriminant analysis | T&F Ch. 9; Albrecht (1980) |
| 13. 8 & 10 Apr | overview and unconstrained ordination | Gauch et al. (1977); Hill & Gauch (1980); Jackson & Somers (1991); Rotenberry & Chandler (1999) |
| 14. 15 & 17 Apr | constrained ordination | Palmer (1993); McCune (1997); Roberts (2009) |
| 15. 22 & 24 Apr | non-parametric techniques | Kenkel & Orlóci (1986) |
| 16. 29 April | wrapping up | James & McCulloch (1990) |

T&F = Tabachnick and Fidell (2013)

★★★ final exam date: 05 May, 0800–1000 ★★★


