# Biostatistics 406
## Applied Multivariate Biostatistics

**Instructor:** Professor A. A. Afifi

**Course Topics:** Various topics in applied multivariate analysis including multiple linear regression, logistic regression, log-linear models, principal components and factor analysis and survival analysis.

**Prerequisites:** Biostatistics 100B or equivalent.

**Learning Objectives:**

<table>
<thead>
<tr>
<th>COURSE LEARNING OBJECTIVES</th>
<th>HOW THESE LEARNING OBJECTIVES ALIGN WITH COMPETENCIES FOR SPECIFIC DEGREE PROGRAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Competencies for MPH in Biostatistics</td>
</tr>
<tr>
<td>1. Fit and compare different models to analyze health science data.</td>
<td><strong>G4.</strong> Consult with public health professionals and researchers helping them design research studies (using statistically rigorous methods for sample size determination and power) and analyzing data obtained from such designs.</td>
</tr>
<tr>
<td>COURSE LEARNING OBJECTIVES</td>
<td>HOW THESE LEARNING OBJECTIVES ALIGN WITH COMPETENCIES FOR SPECIFIC DEGREE PROGRAMS</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Competencies for MPH in Biostatistics</td>
</tr>
<tr>
<td>2. Use various SAS procedures such as REG, LOGISTIC, CATMOD, LIFETEST, PHREG, PRINCOMP, and FACTOR.</td>
<td>G3. Develop analytical and computational skills for the management, modeling and analysis of public health datasets with several variables that may be dependent on one another using statistically rigorous methods and models.</td>
</tr>
<tr>
<td>3. Perform a multiple linear, logistic or survival regression and correctly interpret the output, including the coefficients and their confidence intervals and related tests of hypotheses.</td>
<td>G2. Develop analytical skills and obtain broad insights involving the design and analysis of experiments to understand and model the dependence between different variables (e.g. regression), handle missing or incomplete data, and carry out rigorous statistical modeling for data obtained from a variety of public health study designs.</td>
</tr>
<tr>
<td>COURSE LEARNING OBJECTIVES</td>
<td>HOW THESE LEARNING OBJECTIVES ALIGN WITH COMPETENCIES FOR SPECIFIC DEGREE PROGRAMS</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Competencies for MPH in Biostatistics</td>
</tr>
<tr>
<td>4. Examine plots and other analyses for inconsistencies between the fitted model and patterns in the data and for outliers and high leverage observations.</td>
<td>G2. Develop analytical skills and obtain broad insights involving the design and analysis of experiments to understand and model the dependence between different variables (e.g. regression), handle missing or incomplete data, and carry out rigorous statistical modeling for data obtained from a variety of public health study designs.</td>
</tr>
<tr>
<td>5. Select the best model for the data based on model fit statistics and variable selection methods.</td>
<td>G2. Develop analytical skills and obtain broad insights involving the design and analysis of experiments to understand and model the dependence between different variables (e.g. regression), handle missing or incomplete data, and carry out rigorous statistical modeling for data obtained from a variety of public health study designs.</td>
</tr>
<tr>
<td>COURSE LEARNING OBJECTIVES</td>
<td><strong>HOW THESE LEARNING OBJECTIVES ALIGN WITH COMPETENCIES FOR SPECIFIC DEGREE PROGRAMS</strong></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td><strong>Competencies for MPH in Biostatistics</strong></td>
</tr>
<tr>
<td></td>
<td>G2. Develop analytical skills and obtain broad insights involving the design and analysis of experiments to understand and model the dependence between different variables (e.g. regression), handle missing or incomplete data, and carry out rigorous statistical modeling for data obtained from a variety of public health study designs.</td>
</tr>
<tr>
<td></td>
<td>6. Perform a principal components, factor or log-linear analysis to explore underlying associations among a group of variables.</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>COURSE LEARNING OBJECTIVES</th>
<th>HOW THESE LEARNING OBJECTIVES ALIGN WITH COMPETENCIES FOR SPECIFIC DEGREE PROGRAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Participate in a research team to formulate scientific questions, analyze them using various methods and write a report describing the analysis and conclusions.</td>
<td>Competencies for MPH in Biostatistics: G8. Develop written and oral presentations based on statistical analyses for public health professionals as well as lay audiences.</td>
</tr>
<tr>
<td></td>
<td>Competencies for MS in Biostatistics: A1. Collaborate with researchers to formulate the aims of a public health research project. A2. Formulate a public health question in statistical terms. B2. Effectively communicate statistical concepts and reasoning to public health collaborators. B3. Learn to write and disseminate substantive field publications and communicate the statistical portion of the methodology to a substantive field audience. B5. Be able to articulate interdisciplinary approaches to solving public health problems.</td>
</tr>
<tr>
<td></td>
<td>Competencies for DrPH in Biostatistics: B1. Collaborate with researchers to formulate the aims of a public health research project. B2. Formulate a public health question in statistical terms. C4. Effectively communicate statistical concepts and reasoning to public health collaborators. C5. Learn to write and disseminate substantive field publications and communicate the statistical portion of the methodology to a substantive field audience. D5. Understand and be able to effectively communicate the public health significance of the problems being addressed.</td>
</tr>
<tr>
<td></td>
<td>Competencies for PhD in CHS: 5. Formulate a research question on an important public health topic, design a rigorous and original empirical study to answer it, conduct that study, interpret the results, and draw conclusions.</td>
</tr>
<tr>
<td></td>
<td>Competencies for PhD in Epidemiology: 8. Formulate a research question on an important epidemiologic topic, design a rigorous and original empirical study to answer it, conduct that study, interpret the results, and draw conclusions.</td>
</tr>
</tbody>
</table>

**Homework Assignments:** You will be required to hand in several assignments throughout the quarter based on a combination of problems in the textbook and data analyses from the computer lab.

**Midterm and Final Examinations:** There will be one midterm and a final examination held in class.
**Group Final Paper:** You are required to do a Group Final Paper. The group final paper is to be co-authored by at least 3 (max. 4) enrolled students of the class (no auditors allowed). You will need to plan early. The final paper should summarize and extend the data analyses from the homework assignments in a more formal manner and is considered a “mini-publication”. All co-authors will receive the same grade for the Group Final Paper. Additional information on the specific requirements and the deadline for submitting it will be given in class.

**Grading:** Grades will be based on the homework assignments (20%), the midterm (20%), the final (20%) and the group final paper (40%).

**Data Sets:** The data sets that are used for homework assignments will be available for downloading on the internet. Some of the analyses in the course will use data sets described in the course text and others will be focused on a large data set supplied by the instructor. Supplementary materials describing this data set will be available on the course web site.

**Computing:** In the lab sessions you will be using the SAS language. You should have at least some familiarity with SAS; if not be prepared to spend some extra time on the computer in the first couple of weeks.
**Course Text:**

The textbook for the class is *Practical Multivariate Analysis*, 5th Ed., by A. A. Afifi, S. J. May and V. A. Clark, Chapman & Hall, New York: 2012. It is available in the Health Sciences Bookstore.

**Internet:** There is a web page for this course at //afifi.bol.ucla.edu/biostat406/

**Tentative schedule**

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Textbook Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction and preview</td>
<td>1 - 5</td>
</tr>
<tr>
<td></td>
<td>Review simple linear regression</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Review multiple linear regression</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Variable selection</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Other linear regression topics</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Survival analysis</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>Survival analysis, continued</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Logistic regression</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Logistic regression, continued</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Log-linear models</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>Principal components analysis</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>Factor analysis</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>Missing values analysis</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>Cluster analysis</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Review</td>
<td></td>
</tr>
</tbody>
</table>