

MTH491 Independent Study in Statistics Fall 2014

Instructor: Ryan Botts, Ph.D.

Phone Number: (619)849-2968

E-mail: ryanbotts@pointloma.edu

Text: None required, however various reference texts, provided by the instructor, may be used to review necessary statistics tools.

Content:

This course is designed to extend one's understanding of the basic principles of statistics through application. We will apply these skills to the analysis of large real world data sets. The results of this analysis will be submitted in writing.

Learning Outcomes:

- Students will understand several advanced statistical tools from the following: principal component analysis, partial least squares regression, [linear models](#), [mixed effects models](#), and [generalized linear models](#).
- Students will apply their knowledge of statistics to the analysis of a real data set.
- Students will communicate the results of statistical analysis both orally and in writing.

The best way to **learn** mathematics is by **doing** mathematics. In an attempt to further our understanding of statistics, we will perform an in depth statistical analysis of a real world data set. We will also practice communicating the results of data analysis in writing.

Grading:

Grades are based on the total number of points accumulated throughout the course. The points for each activity are:

Activity	Due Date	Points
Outline of research goals	Friday 9/12	100
Summary of relevant statistical tools	Friday 10/3	200
Preliminary model fitting and diagnostics	Friday 10/31	50
Model improvement and final analysis	Friday 11/7	50
Rough draft of final report	Tuesday 11/25	100
Final Report	Monday 12/15	400
Total		1000

Approximate minimal points required to obtain a given grade are:

	A	B	C	D
+		(875, 900)	(775, 800)	(675, 700)
	[925, 1000]	[825, 875]	[725, 775]	[625, 675]
-	[900, 925)	[800, 825)	[700, 725)	[600, 625)

Note that scores of 599 or lower will result in an F.

Deleted: Sep. 13
Deleted: Sep. 27
Deleted: GEO database tool description and download of other data
Deleted: Oct. 11
Deleted: 10
Deleted: Data Clean-up and preliminary analysis
Deleted: 1
Deleted: 20
Deleted: Friday
Deleted: Nov. 22
Deleted: Monday
Deleted: Dec. 9
Deleted: 3

Activities:

You will be given a large real world data set and asked to answer a series of research questions based on this data set. A series of activities will be submitted throughout the project in order to measure your progress towards analyzing the data. In addition to these activities it is expected that you will meet weekly with the instructor to give brief updates and discuss the analysis. Meeting times will be determined with the consensus of the student and faculty member.

Final Project:

A [written](#) final project [report](#) describing the goals of the project, an overview of the statistical techniques used in the project, and a summary of the findings of the project will be delivered on or before **Monday Dec. 9, 2012**.

Class Enrollment:

It is the student's responsibility to maintain his/her class schedule. Should the need arise to drop this course (personal emergencies, poor performance, etc.), the student has the responsibility to follow through (provided the drop date meets the stated calendar deadline established by the university), not the instructor. Simply ceasing to attend this course or failing to follow through to arrange for a change of registration (drop/add) may easily result in a grade of F on the official transcript.

Academic Accommodations:

While all students are expected to meet the minimum academic standards for completion of this course as established by the instructor, students with disabilities may require academic accommodations. At Point Loma Nazarene University, students requesting academic accommodations must file documentation with the Disability Resource Center (DRC), located in the Bond Academic Center. Once the student files documentation, the Disability Resource Center will contact the student's instructors and provide written recommendations for reasonable and appropriate accommodations to meet the individual needs of the student. This policy assists the university in its commitment to full compliance with Section 504 of the Rehabilitation Act of 1973, the Americans with Disabilities (ADA) Act of 1990, and ADA Amendments Act of 2008, all of which prohibit discrimination against students with disabilities and guarantees all qualified students equal access to and benefits of PLNU programs and activities.

Students with learning disabilities who may need accommodations should discuss options with the instructor during the first two weeks of class.

Academic Honesty:

The Point Loma Nazarene University community holds the highest standards of honesty and integrity in all aspects of university life. Academic honesty and integrity are strong values among faculty and students alike. Any violation of the university's commitment is a serious affront to the very nature of Point Loma's mission and purpose.

Academic dishonesty is the act of presenting information, ideas, and/or concepts as one's own when in reality they are the results of another person's creativity and effort. Such acts include plagiarism, copying of class assignments, and copying or other fraudulent behavior on examinations. For more details on PLNU's policy go to:

<http://www.pointloma.edu/experience/academics/catalogs/undergraduate-catalog/point-loma-education/academic-policies>

A student who is caught cheating on any item of work will receive a zero on that item and may receive an "F" for the semester. See the PLNU Catalog for a further explanation of the PLNU procedures for academic dishonesty.

Activity Guidelines

It should be noted, that the activities that you submit are an assessment of your progress towards understanding the data set. In other words these assignments are snapshots of how much progress you have made towards a thorough analysis of the data. Thus you should begin work on many of the activities long before they are due. For example the data clean-up should begin as soon as you have the data, however a final product will not be submitted until the due date.

Outline of research goals (Due Sep 9, 2014)

- provide a brief one page outline of the research aims for the project
- clearly identify the questions we wish to answer based on the data

Summary of relevant statistical tools (Due Oct 3, 2014)

- Identify the statistical tools that would be used to analyze the data that has been given to you.
- Topics include normalization, correlation, principal component analysis, partial least squares regression, [linear models](#), [mixed effects models](#), and [generalized linear models](#).
- Write a brief summary of what methods are [being used](#) and how they are used.
- [We will discuss which tools you will use, but you may need to do some research to understand more detail about the tool.](#)
- The aim of this assignment is for you to demonstrate your knowledge of the essential topics and an understanding of the procedures you are using to analyze the data.

Preliminary Model Fitting and Diagnostics (Due Oct 31, 2014)

- [Begin the analysis of your data](#) using the tools we have discussed. [Be prepared to share the preliminary results for discussion and model assessment.](#)
- In order to perform the analysis it may also be necessary to perform some data clean-up and formatting.

Model Improvement and Final Analysis (Due Nov 7, 2014)

- After discussing the results from the previous step. Modify the results to improve the performance of your model.
- Summarize this final analysis and prepare the results for interpretation. We will discuss the meaning of them together.

Rough draft of final project (Due Nov 25, 2014)

- See directions for the final project.
- Data analysis should be complete at this point and the data should be present in the paper.

Final Report (Due Dec 15, 2014)

The final report should have each of the following components:

1. Introduction into the research project and a statement of the research aims.
2. A brief summary of the statistical methods that were used.
3. Data summary with the appropriate summary statistics.
4. Results
5. Conclusions
6. Also attach the final cleaned version of the data.

Deleted: the

Deleted: the

Deleted: If there are any tools that you do not currently know that you will need, we will discuss them prior to the submission of this assignment and

Deleted: <#>you can demonstrate your understanding of the skill.¶

Deleted: Accessing the GEO database to extract other microarray data

Deleted: <#>The availability of raw data is part of what makes this an interesting topic. ¶ <#>Utilize the GEO database to find an experiment and download the microarray data that goes with it. ¶ You will use this data as part of your final analysis.

Deleted: Data clean-up and preliminary

Deleted: a

Formatted: Numbered + Level: 1 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.5"

Deleted: ¶ <#>Summary of project:¶

¶ We are going to focus on analyzing microarray and metabolomics data (these tools are very similar). We will learn about principal component analysis, partial least squares regression and possibly discriminant analysis (time permitting). We will apply these analysis techniques to the analysis of mouse eye development and one other set of data of your choice.