

# 2015-2016 LONG SIGNATURE SHEET

**RECEIVED**  
10-21-15

Revised ✓  
**RECEIVED**  
10-22-15

Proposal Number: MATH 10-20-15

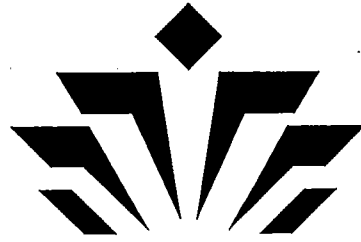
**UNC CHARLOTTE**

Proposal Title: Establishment of New Courses STAT 5110 and STAT 6115

Originating Department: Mathematics and Statistics

TYPE OF PROPOSAL: UNDERGRADUATE \_\_\_\_\_ GRADUATE X UNDERGRADUATE & GRADUATE \_\_\_\_\_  
(Separate proposals sent to UCC and Grad. Council)

DATE RECEIVED	DATE CONSIDERED	DATE FORWARDED	ACTION	SIGNATURES
10/20/2015	10/20/2015	10/20/2015	Approved	<u>DEPARTMENT CHAIR</u> Yuanan Diao <i>Yuan Diao</i>
			Approved	<u>COLLEGE CURRICULUM COMMITTEE CHAIR</u> <i>Janet E. Leux</i> [print name here:] Janet E Leux
		<i>11</i>	Approved <i>11-13-15</i>	<u>COLLEGE FACULTY CHAIR (if applicable)</u> <i>Elizabeth Stearns</i> [print name here:] Elizabeth Stearns
			Approved <i>11-16-15</i>	<u>COLLEGE DEAN</u> <i>Shawn Lewis</i> [print name here:] Shawn Lewis
			Approved	<u>GENERAL EDUCATION</u> (if applicable; for General Education courses) [print name here:]
			Approved	<u>HONORS COLLEGE</u> (if applicable; for Honors courses & programs) [print name here:]
			Approved	<u>UNDERGRADUATE COURSE &amp; CURRICULUM COMMITTEE CHAIR</u> (for undergraduate content)
<i>11/16/15</i>	<i>12/1/15</i>		Approved	<u>GRADUATE COUNCIL CHAIR</u> (for graduate content) <i>Dennis Lnesay</i>
				<u>FACULTY GOVERNANCE ASSISTANT</u> (Faculty Council approval on Consent Calendar)
				<u>FACULTY EXECUTIVE COMMITTEE</u> (if decision is appealed)



# UNC CHARLOTTE

## LONG FORM COURSE AND CURRICULUM PROPOSAL

---

To: Dr. Dennis Livesay, Chair of The Graduate Council

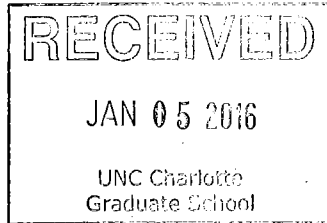
From: Department of Mathematics and Statistics

Date: October 20, 2015

Re: Establishment of New Courses STAT 5110 and STAT 6115

---

Please find the attached proposal to establish two new stat courses STAT 5110 – Applied Regression Analysis and STAT 6115 – Statistical Learning with Big Data.



MATH 10-20-15

MATH 10-20-15

**University of North Carolina at Charlotte**

**New Graduate**

**Course and Curriculum Proposal from: Department of Mathematics and Statistics**

**Title: Establishment of New Courses STAT 5110 and STAT 6115**

## **II. CONTENT OF PROPOSAL**

### **A. PROPOSAL SUMMARY**

#### **1. PROPOSAL SUMMARY**

The Department of Mathematics and Statistics proposes to create two new graduate-level statistics courses STAT 5110 - Applied Regression Analysis and STAT 6115 - Statistical Learning with Big Data. The two statistics courses are intended for graduate students in the Professional Science Master's in Data Science and Business Analytics (PSM-DSBA) program.

### **B. JUSTIFICATION**

#### **1. Identify the need addressed by the proposal and explain how the proposed action meets the need.**

The STAT 5110 course "Applied Regression Analysis" will address the need for training DSBA professionals in statistical modeling used in data science and business analytics. This proposed course will provide students theoretical and practical training in statistical modeling with particular emphasis on the application of linear regression and multivariate statistical analysis methods. The course will introduce associated ideas, basic concepts, statistical inference techniques, and how to apply these techniques to extract information from data generated in various application areas. This course is necessary to help students learn statistical techniques for model building, estimation and statistical inference under the proposed model as well as model adequacy checking. In addition, statistical software, either R or SAS, will be introduced and used throughout the course so students will get hands-on experience by using the software for data analysis. Students who enroll into the PSM-DSBA program are highly recommended to take this course first to be prepared for the rest of the curriculum such as Statistical Learning.

The STAT 6115 course "Statistical Learning with Big Data" will address the need for training DSBA professionals in statistical learning used in data science and business analytics. This proposed course will provide students a survey of major statistical learning concepts and methods for big data analysis. In this course students will learn

how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms in case studies in financial investment, gene identification, and feature selection in high-dimensional spaces. Students completing this course should be able to apply basic statistical learning methods to build predictive models, to properly tune and select statistical learning models, to correctly assess model fit and error, and to build an ensemble of learning algorithms for data. Software R will also be used for simulations and for real data analysis in this course.

The two proposed courses STAT 5110 and STAT 6115 are part of a joint effort to redesign the curriculum for the PSM-DSBA program to provide students in this program some statistical training that is applicable to DSBA. In this new curriculum, STAT 6115 will be a core (required) course for all students in this program, while STAT 5110 will be treated as an elective intended for students without proper prior training/knowledge in statistics needed for STAT 6115.

**2. Discuss prerequisites/corequisites for course(s) including class-standing, admission to the major, GPA, or other factors that would affect a student's ability to register.**

The prerequisites for STAT 5110 is MATH 2164 and STAT 2122 or STAT 3128 or equivalent, or permission of the department.

The prerequisite for STAT 6115 is STAT 5110 or STAT 5123 - Applied Statistics I, or permission of the department.

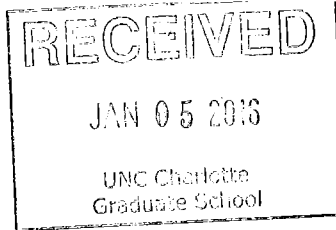
**3. Demonstrate that course numbering is consistent with the level of academic advancement of students for whom it is intended.**

The UNC Charlotte course numbering guidelines have been followed for the proposed new course numbers STAT 5110 and STAT 6115 which position the courses as graduate courses for master's students. In particular, STAT 5110 is assigned for the proposed Applied Regression Analysis course because (1) this course is in comparable with STAT 5123 – Applied Statistics I, a required core course for our own stat students in MS and PhD programs, and (2) if needed in the future, this course may be cross-listed at the 4000 level, just like STAT 5123 is cross-listed as STAT 4123.

In addition, the course numbering is also consistent with the department's course numbering practice that a course number (5|6|8)11X should be used for a course designed for non-major graduate students.

**4. In general, how will this proposal improve the scope, quality and/or efficiency of programs and/or instruction?**

These two new courses are intended for graduate students in the PSM-DSBA program. On the other hand, the STAT 6115 course will also be a useful elective course for students in our own MS-MATH and PhD-Applied Math programs.



- 5. If course(s) has been offered previously under special topics numbers, give details of experience including number of times taught and enrollment figures.**

None of the two proposed stat courses STAT 5110 and STAT 6115 has been previously offered as special topics.

**C. IMPACT.**

- 1. What group(s) of students will be served by this proposal? (Undergraduate and/or graduate; majors and/or non-majors, others? Explain). Describe how you determine which students will be served.**

Generally speaking, the two new stat courses will primarily serve graduate students in the PSM-DSBA program, although students in our own MS-MATH and PhD-Applied Math programs could also choose STAT 6115 as an elective course.

- 2. What effect will this proposal have on existing courses and curricula?**

- a. When and how often will added course(s) be taught?**

We expect to offer STAT 5110 every fall and STAT 6115 every spring, respectively, although based on the PSM-DSBA program's need, they could be offered in summer terms too.

- b. How will the content and/or frequency of offering of other courses be affected?**

The two new stat courses are intended primarily for graduate students in the PSM-DSBA program. The content of STAT 6115 doesn't overlap with other stat courses offered by the math department. On the other hand, there are some overlapping contents between STAT 5110 and STAT 5123 - Applied Statistics I, the latter being a required core course for our own stat students in MS and PhD programs. Therefore, we will not allow our major students to take STAT 5110 for any credits and we will allow our major students to take STAT 6115 only as an elective. As such, the content and the frequency of offering of the existing graduate-level courses in our own graduate programs will not be affected by the two proposed stat courses.

- c. What is the anticipated enrollment in course(s) added (for credit and auditors)?**

Enrollment in the proposed courses is anticipated to be around 15-20 students per offering.

- d. How will enrollment in other courses be affected? How did you determine this?**

It is expected that the enrollment in the existing graduate-level courses in our own graduate curriculum will not be affected since, as mentioned above in b, we will not

allow our MS-Math and PhD-Applied Math students to take STAT 5110 for credits and we will allow our major students to take STAT 6115 only as an elective.

- a. Identify other areas of catalog copy that would be affected, including within other departments and colleges (e.g., curriculum outlines, requirements for the degree, prerequisites, articulation agreements, etc.)**

At this point, no area of graduate catalog copy would be affected by this proposal. However, it is our understanding that the two new stat courses, once approved by the Graduate Council, might be included in the PSM-DSBA curriculum; at that point a separate long form will be submitted by the PSM-DSBA program.

### **III. RESOURCES REQUIRED TO SUPPORT PROPOSAL.**

#### **A. PERSONNEL**

Dr. Shaoyu Li from Mathematics and Statistics drafted the course proposal for STAT 5110 and is interested in teaching this course. Dr. Li received her dual major Ph.D degree in Statistics and Quantitative Biology from Michigan State University in 2011. Before she joined the faculty in the Department of Mathematics and Statistics at UNC Charlotte in 2014, Dr. Li has worked at St Jude Children's Research Hospital as an Assistant Faculty Member in the Department of Biostatistics. Dr. Li's research interests focus on the development of efficient statistical/computational methods for high-throughput genomics/epigenomics data analysis as well as statistical issues in clinical trial design. She has taught STAT 3110: Applied Regression, STAT 3128: Probability and Statistics for Engineers, and STAT 4123/5123: Applied Statistics I at UNC Charlotte.

Other stat faculty from Mathematics and Statistics such as Dr. Jiancheng Jiang, Dr. Yang Li, Yanqing Sun, Zhiyi Zhang, and Weihua Zhou, all are qualified to teach this course too.

Dr. Jiancheng Jiang from Mathematics and Statistics drafted the course proposal for STAT 6115 and is interested in teaching this course. Dr. Jiang has published many important research works in leading journals on statistics and econometrics. In the past 10 years he has been working on big data. He has taught many statistical courses at levels 1000-8000 for undergraduates and graduates at UNC Charlotte.

#### **B. PHYSICAL FACILITY.**

No additional facilities required.

#### **C. EQUIPMENT AND SUPPLIES:**

No additional equipment and supplies required.

#### **D. COMPUTER.**

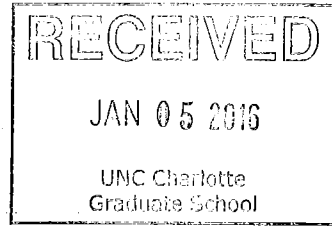
No additional computer resources required since the proposed courses will use the free software R.

**E. AUDIO-VISUAL.**

No additional audio and visual resources required.

**F. OTHER RESOURCES.**

Not applicable



**G. SOURCE OF FUNDING**

Not applicable

**IV. CONSULTATION WITH THE LIBRARY AND OTHER DEPARTMENTS OR UNITS**

**A. LIBRARY CONSULTATION.**

The Department consulted with Alison Bradley at the J. Murrey Atkins Library and was ensured that the present library holdings are adequate to support the proposed courses. Copies of Consultation on Library Holdings can be found in **Attachment I**.

**B. CONSULTATION WITH OTHER DEPARTMENTS OR UNITS**

The proposed courses were developed in consultation with faculty in the PSM-DSBA program for which the courses are primarily designed. A copy of the memo from Dr. Mirsad Hadzikadic, Director of PSM-DSBA, documenting their support of the courses can be found in **Attachment II**. In particular, they agree to cross list STAT 5110 as DSBA 5110 and STAT 6115 as DSBA 6115, respectively.

**C. HONORS COUNCIL CONSULTATION**

The proposal does not involve Honors programs as well as any Honors courses. As such, no consultation with the Honors Council is conducted.

**V. INITIATION, ATTACHMENTS AND CONSIDERATION OF THE PROPOSAL**

**A. ORIGINATING UNIT**

Discussion regarding the creation of two stat courses for the PSM-DSBA program initiated between Dr. Mirsad Hadzikadic, Chair of the Software and Information System Department, and Dr. Yuanan Diao, Chair of the Mathematics and Statistics Department.

Dr. Shaoyu Li and Dr. Jiancheng Jiang drafted the course proposal for STAT 5110 and STAT 6115, respectively. The proposal was presented to the department Graduate Curriculum Committee (GCC) for review on September 10, 2015. The GCC fully supported the proposal with some minor recommendations that have been incorporated into the latest proposal. The GCC includes Drs. Shaozhong Deng (Chair), Joel Avrin, Wei Cai, Victor

Cifarelli, Anthony Fernandes, Jiancheng Jiang, Shaoyu Li, and Weihua Zhou. A copy of the committee's supporting letter can be found in **Attachment III**.

**B. CREDIT HOUR (Mandatory if new and/or revised course in proposal)**

- The appropriate faculty committee has reviewed the course outline/syllabus for each of the two courses (STAT 5110 and STAT 6115) and has determined that the assignments are sufficient to meet the University definition of a credit hour.

**C. ATTACHMENTS**

1. CONSULTATION

Copies of Consultation on Library Holdings can be found in **Attachment I**.

A copy of Consultation with the PSM-DSBA program can be found in **Attachment II**.

A copy of the Graduate Curriculum Committee's supporting letter can be found in **Attachment III**.

2. COURSE OUTLINE/SYLLABUS

Outlines/syllabi for the proposed courses can be found in **Attachment IV**.

3. PROPOSED CATALOG COPY

The proposed catalog copy for each of the two new courses (STAT 5110, STAT 6115) can be found in **Attachment IV**.

a. For STAT 5110, check all the statements that apply:

This course will be cross listed with another course. (Namely, DSBA 5110)

There are prerequisites for this course.

There are corequisites for this course.

This course is repeatable for credit.

This course will increase/decrease the number of credits hours currently offered by its program.

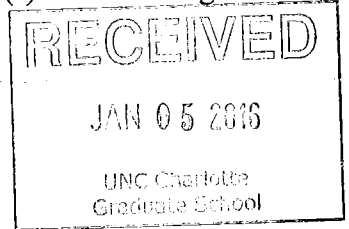
This proposal results in the deletion of an existing course(s) from the degree program and/or catalog.

a. For STAT 6115, check all the statements that apply:

This course will be cross listed with another course. (Namely, DSBA 6115)



- There are prerequisites for this course.
- There are corequisites for this course.
- This course is repeatable for credit.
- This course will increase/decrease the number of credits hours currently offered by its program.
- This proposal results in the deletion of an existing course(s) from the degree program and/or catalog.



4. ACADEMIC PLAN OF STUDY (UNDERGRADUATE ONLY)

- Yes.
- No.

5. STUDENT LEARNING OUTCOMES (UNDERGRADUATE & GRADUATE)

- Yes.
- No.

6. TEXTBOOK COSTS

- Yes. Briefly explain below.
- No. Briefly explain below.

There is no required textbook for either STAT 5110 or STAT 6115. On the other hand, the university's library has purchased full access to the electronic version of the reference book for STAT 6115.

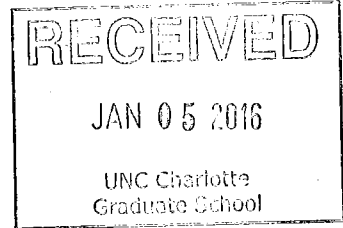
**Attachment I:**

**Consultation on Library Holdings**



J. Murrey Atkins Library

Consultation on Library Holdings



To: Shaozhong Deng
From: Allison Bradley
Date: 10/6/15
Subject: STAT 5110 Applied Regression Analysis

Summary of Librarian's Evaluation of Holdings:

Evaluator: Allison Bradley Date: 10/6/15

Check One:

- 1. Holdings are superior
2. Holdings are adequate (checked)
3. Holdings are adequate only if Dept. purchases additional items.
4. Holdings are inadequate

Comments:

Library holdings should be adequate to support student research for this course (see list of items held by subject heading below). Students will have access to relevant databases including MathSciNet, SpringerLink, ScienceDirect, and many others.

Table with 3 columns: LC Subject Heading, Books, Journals. Rows include Mathematical statistics, Regression Analysis, and Multivariate analysis.

Handwritten signature of Allison Bradley over a line labeled 'Evaluator's Signature'.

10/6/15

Date



J. Murrey Atkins Library

Consultation on Library Holdings

To: Shaozhong Deng  
 From: Alison Bradley  
 Date: 10/6/15  
 Subject: STAT 6115 Statistical Learning with Big Data

Syllabus

Summary of Librarian's Evaluation of Holdings:

Evaluator: Alison Bradley Date: 10/6/15

Check One:

- 1. Holdings are superior \_\_\_\_\_
- 2. Holdings are adequate   x
- 3. Holdings are adequate only if Dept. purchases additional items. \_\_\_\_\_
- 4. Holdings are inadequate \_\_\_\_\_

Comments:

Library holdings should be adequate to support student research for this course (see list of items held by subject heading below). Students will have access to relevant databases including MathSciNet, SpringerLink, ScienceDirect, and many others.

LC Subject Heading	Books	Journals
Mathematical statistics	2096	106
Data mining	1509	23

\_\_\_\_\_  
 Evaluator's Signature

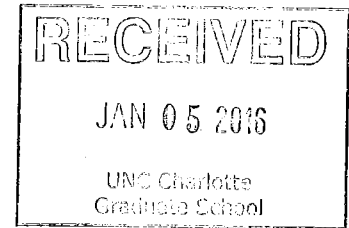
10/6/15

\_\_\_\_\_  
 Date

## Attachment II: Consultation with the PSM-DSBA program



9201 University City Boulevard, Charlotte, NC 28223-0001



To: Graduate Council Chair  
From: Mirsad Hadzikadic, Executive Director, Data Science Initiative  
RE: Cross-listing of Courses with DSBA  
Date: September 9, 2015

This memo documents our support of the proposed changes to the Professional Science Master's In Data Science and Business Analytics (DSBA). We are pleased to have the opportunity to collaborate with the Department of Mathematics and Statistics to offer and cross list two new courses for students in the DSBA program: STAT/DSBA 5110 and STAT/DSBA 6115.

The proposed changes were developed in consultation with DSBA faculty.

Sincerely,

A handwritten signature in cursive script, appearing to read "Mirsad Hadzikadic".

Mirsad Hadzikadic, Ph.D.  
Professor, Department of Software and Information Systems  
Executive Director, Data Science Initiative  
Director, Complex Systems Institute  
Director, Data Science and Business Analytics Professional Science Master's  
College of Computing and Informatics / The Graduate School  
343-A Woodward Hall  
UNC Charlotte  
Charlotte, NC 28223  
704-687-8643

## Attachment III: Letter from Graduate Curriculum Committee



UNC CHARLOTTE

College of Liberal Arts & Sciences

Mathematics and Statistics

9201 University City Blvd, Charlotte, NC 28223-0001  
t/ 704.687.0620 f/ 704.687.1392 math@uncc.edu

TO: Yuanan Diao, Chair, Dept. of Mathematics and Statistics

FROM: Shaozhong Deng, Chair, Graduate Curriculum Committee.  
Dept. of Mathematics and Statistics

RE: Establishment of New Courses STAT 5110 and STAT 6115

A handwritten signature in black ink, appearing to read "J. Deng", with a long horizontal flourish extending to the right.

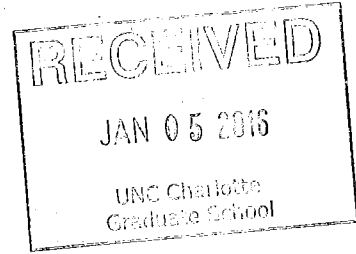
DATE: September 10, 2015

The 2015-2016 Graduate Curriculum Committee has approved by a 8-0 vote the proposal of establishing the new graduate courses STAT 5110: Applied Regression Analysis and STAT 6115: Statistical Learning with Big Data.

Please feel free to contact me at 704-687-0634 or [shaodeng@uncc.edu](mailto:shaodeng@uncc.edu) if you have any questions.

**Attachment IV:**

**Course Descriptions and Syllabi**



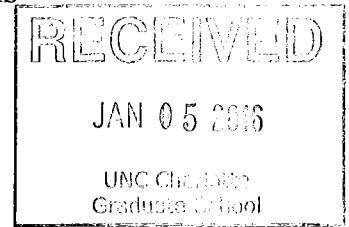
**STAT 5110/DSBA 5110 Catalog Copy**

**STAT 5110. Applied Regression Analysis (3).** Prerequisites: MATH 2164 and STAT 2122 or STAT3128 or equivalent, or permission of the department. This course provides theoretical and practical training in statistical modeling with particular emphasis on the application of linear regression and multivariate statistical analysis. The basic fundamentals and statistical inference techniques associated with regression models will be introduced. Students will also learn how to apply the statistical techniques to extract information from data generated in various application areas using statistical software. Topics generally covered include but not limited to linear regression, model adequacy checking and diagnostics, generalized linear regression, and multivariate statistical analysis. (*Fall*)

**DSBA 5110. Applied Regression Analysis (3).** Prerequisites: MATH 2164 and STAT 2122 or STAT3128 or equivalent, or permission of the department. This course provides theoretical and practical training in statistical modeling with particular emphasis on the application of linear regression and multivariate statistical analysis. The basic fundamentals and statistical inference techniques associated with regression models will be introduced. Students will also learn how to apply the statistical techniques to extract information from data generated in various application areas using statistical software. Topics generally covered include but not limited to linear regression, model adequacy checking and diagnostics, generalized linear regression, and multivariate statistical analysis. (*Fall*)



## STAT 5110 Applied Regression Analysis Syllabus



### Instructor information:

Name: Dr. Shaoyu Li  
 Office: Fretwell 380D  
 Office Number: 704-687-1374  
 Email: [sli23@uncc.edu](mailto:sli23@uncc.edu)  
 Office Hours: TBD

### Course information:

Course Name: Applied Regression Analysis

Prerequisites: MATH 2164 and STAT 2122 or STAT 3128 or equivalent

Course Materials: There is no required textbook for this course. All materials (lecture notes, homework assignments, practice worksheets, etc.) will be posted on Moodle.

Reference Books: *Introduction to Linear Regression Analysis* (5th Edition) by Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining  
*Applied Multivariate Statistical Analysis* (6th Edition) by Richard A. Johnson and Dean W. Wichern

Course Description: This course provides theoretical and practical training in statistical modeling with particular emphasis on the application of linear regression and multivariate statistical analysis. The basic fundamentals and statistical inference techniques associated with regression models will be introduced. Students will also learn how to apply the statistical techniques to extract information from data generated in various application areas using statistical software. Topics generally covered include but not limited to linear regression, model adequacy checking and diagnostics, generalized linear regression, and multivariate statistical analysis.

Software: Software R ([www.r-project.org](http://www.r-project.org)) will be introduced and used throughout the course.

### Course Objectives:

The objectives of this course are (1) to provide students basic knowledge of how to build up an appropriate regression model, (2) to teach students how to conduct statistical inferences and multivariate statistical analysis, (3) to introduce students how to assess the advantages and disadvantages of different models and techniques, and (4) to train

students how to perform the analysis using a statistical software.

**Instructional Method:**

The primary format for instruction will be lectures.

**Means of Student Evaluation:**

**Homework assignments (25%):** Homework will be assigned after every lecture and will be required to be turned in the following week.

**Course project (25%):** Students will make groups to work on a course project at the end of the semester.

**Exams:** There will be one midterm (25%) and one comprehensive final exam (25%) throughout the semester. The midterm exam will be given during a regularly scheduled lecture. The final exam will be from 2:00-4:30pm, Wednesday, December 15, 2016.

The exams will be closed book and closed notes. However, one handwritten cheat sheet of size A4 (8.3 in x 11.7 in) will be permitted.

Course letter grades will follow the scale listed below (not curved):

90-100%	A
80-89%	B
70-79%	C
0-69%	U

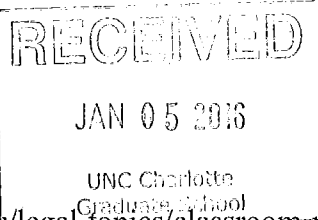
**Class Policies:**

**Late homework policy:** Homework is due at the beginning of class on the due date. No late homework will be accepted.

**Make-up policy:** *NO make-up* for the midterm exam. The missed midterm exam will receive a score of 0 unless a valid reason is presented before the time of the exam or after it in the case of an emergency. Only extreme situations with an official documented excuse will allow a student to be “excused” from the exam. The score for the excused exam will then be determined from the corresponding exam questions on the final exam.

**Attendance policy:** Students are expected to be present and ON TIME for ALL class sessions, to keep a copy of the course syllabus, and to take class notes. You are responsible for the material covered if you are absent from class. Any changes in this syllabus or in the scheduling of exams will be announced during class.

**Grade-change policy:** if you have any questions regarding the grading of any of your graded material, you should return the graded material to your instructor and attach a short note with your questions *before leaving* the classroom. This must be done on the day that the instructor hands back the material in class. Once the graded material has left the classroom, no grade changes will be made.



**Other University Policies:**

Classroom Policies: <http://legal.uncc.edu/legal-topics/classroom-policies-and-practices>

Academic Integrity: <http://legal.uncc.edu/policies/up-407>

**Tentative Schedule/Topical Outline:**

Components	Session No.	Lectures/Discussion/Exams
<i>Linear Regression Models</i>	<u>Week 1</u>	Review of Basic Statistical Concepts
	<u>Week 2</u>	Matrix Notation and Matrix Algebra
	<u>Week 3</u>	Introduction Parameter Estimation
	<u>Week 4</u>	Hypothesis testing
<i>Generalized Linear Regression Models</i>	<u>Week 5</u>	Analysis of Variance
	<u>Week 6</u>	Confidence Intervals in Multiple Linear Regression
	<u>Week 7</u>	Prediction of New Observations
	<u>Week 8</u>	<b>Mid-term Exam</b>
	<u>Week 9</u>	Model Adequacy Checking
	<u>Week 10</u>	Generalized Linear Models—Logistic Regression Model
	<u>Week 11</u>	Generalized Linear Models—Logistic Regression Model
	<u>Week 12</u>	Aspects of Multivariate Analysis
<i>Multivariate Analysis</i>	<u>Week 13</u>	Inference About A Mean Vector
	<u>Week 14</u>	Multivariate Linear Regression Models
	<u>Week 15</u>	Principle Components and Factor Analysis
	<u>Week 16</u>	<b>Final Exam</b>

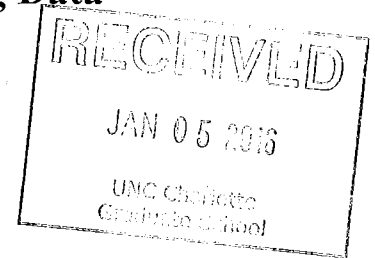
*Please note that this schedule is tentative and may change during the semester.*

**STAT 6115/DSBA 6115 Catalog Copy**

**STAT 6115. Statistical Learning with Big Data (3).** Prerequisite: STAT 5110 or STAT 5123 or permission of the department. This course provides students a survey of major statistical learning concepts and methods for big data analysis, including both supervised and unsupervised learning such as resampling methods, support vector machines, model selection and regularization, tree-based methods and ensembles, statistical graphics. Students learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms in case studies in financial investment, gene identification, and feature selection in high-dimensional spaces. (*Spring*)

**DSBA 6115. Statistical Learning with Big Data (3).** Prerequisite: STAT 5110 or STAT 5123 or permission of the department. This course provides students a survey of major statistical learning concepts and methods for big data analysis, including both supervised and unsupervised learning such as resampling methods, support vector machines, model selection and regularization, tree-based methods and ensembles, statistical graphics. Students learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms in case studies in financial investment, gene identification, and feature selection in high-dimensional spaces. (*Spring*)

# STAT 6115 Statistical Learning with Big Data Syllabus



## Instructor Information:

Name: Dr. Jiancheng Jiang  
 Office: Fretwell 350A  
 Office Number: 704-687-0631  
 Email: [jjiang1@uncc.edu](mailto:jjiang1@uncc.edu)  
 Office Hours: TBD

## Course Information:

Course Name: Statistical Learning with Big Data  
 Prerequisite: STAT 5110 or STAT 5123, or permission of the department  
 Textbook: There is no required textbook for this course. The course material will be a selective combination of a number of excellent books in statistical learning, data mining, and machine learning, such as  
*An Introduction to Statistical Learning with Applications in R*, by G. James, D. Witten, T. Hastie and R. Tibshirani, Springer, 2013.

(The university has full access to the eBook of the textbook.)

Course Description: This course provides students a survey of major statistical learning concepts and methods for big data analysis, including both supervised and unsupervised learning such as resampling methods, support vector machines, model selection and regularization, tree-based methods and ensembles, statistical graphics. Students learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms in case studies in financial investment, gene identification, and feature selection in high-dimensional spaces. Software R will be used for simulations and for real data analysis.

## Course Objectives:

By providing students a survey of major statistical learning concepts and methods for big data analysis, the course is designed for students to learn how and when to apply statistical learning techniques, how to critically evaluate the performance of learning algorithms, and how to build predictive models. Students completing this course should be able to apply basic statistical learning methods to build predictive models, to properly tune and select statistical learning models, to correctly assess model fit and error, and to

build an ensemble of learning algorithms for big data.

### **Instructional Method:**

The primary format for instruction will be lectures.

### **Means of Student Evaluation:**

**Attendance (5%):** Attendance is required. For each un-excused absence, you will lose 10% of the attendance points. If you miss 10 or more than 10 classes, you will receive a U for the final grade without exception. If you miss a class, it will be your responsibility to obtain the assignment and find out what material is covered.

**Pop quizzes (5%):** Several pop quizzes will be given either online or at any time in lectures. All pop quizzes will be open book and open notes. No make up for missed quizzes if there is no legitimate excuse.

**Homework assignments (20%):** Reading relevant sections in reference books after classes is always expected. Homework will be assigned in class or through Moodle. Homework will be due at the beginning of class on the due date. No late homework will be accepted.

**Course project (15%):** Students will make groups to work on a course project at the end of the semester.

**Midterm exam (25%):** You will be informed in advance about the date and the material coverage of the midterm exam. The midterm exam may be open or closed book. If it is closed book, you may choose to prepare a formula sheet for reference for the exam.

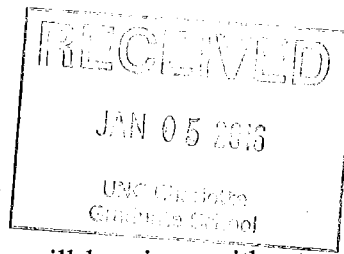
**Final exam (30%):** The final exam may be open or closed book. A formula sheet and some selected tables will be provided for the final exam if appropriate. The final exam will contain 20% multiple choice, true/false, or blank-filling questions and 80% free response questions. If a student misses the final exam, his/her final grade for the course would be either an I or a U, depending on his/her performance on previous work.

Course letter grades will follow the scale listed below (not curved):

90-100%	A
80-89%	B
70-79%	C
0-69%	U

### **Software and Computing:**

Students are encouraged to bring a laptop to each class. The course will make extensive use of the free R software environment for statistical computing and graphics. Introductory tutorials/labs will be provided for students unfamiliar with R. Students are encouraged to install R on their laptops before the semester begins. Copies of R may be downloaded from <http://www.rproject.org>.



**Class Policies:**

**Exam make-up policy:** No make-up exam will be given without prior notification and documented acceptable excuse. If you miss an exam due to a legitimate excuse, you need to call or email me either before the exam or within 24 hours after the exam in order to be allowed a make-up exam.

**Cheating:** Cheating in any form will result in a “U” for the course grade and may be reported to the University.

**Classroom policy:** Coming to class late or leaving class early is NOT allowed. Any distraction behavior such as TALKING or LAUGHING or YELLING in class is prohibited. Electronic devices other than a laptop also cause distraction during classes and are thus not permitted either. More classroom policies can be found at

<http://legal.uncc.edu/legal-topics/classroom-policies-and-practices>

**Academic integrity:** The UNC Charlotte is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behavior at <http://legal.uncc.edu/policies/up-407>

**Disability services:** Students who require accommodations in this course due to a disability affecting mobility, vision, hearing, learning, or mental or physical health should contact the Office of Disability Services at Fretwell 230, 704-687-4355 (phone).

**Grief absence policy for students:** UNC Charlotte recognizes that a time of bereavement is very difficult for a student. The University therefore provides the following rights to students facing the loss of a family member through the Grief Absence Policy for Students (GAPS): Students will be excused for funeral leave and given the opportunity to earn equivalent credit and to demonstrate evidence of meeting the learning outcomes for missed assignments or assessments in the event of the death of a member of the student’s family.

**Others:** The instructor reserves the right to revise this syllabus if it is necessary.

**Tentative Schedule:**

- Week 1. Introduction to statistical learning with big data and examples
- Week 2. Overview of supervised learning methods: motivating examples, trade-off between simple and complex models, classification problems, lab
- Week 3. Statistical learning with linear models – part I: examples, simple linear regression
- Week 4. Statistical learning with linear models – part II: multiple linear regression with examples, lab
- Week 5. Statistical learning with classification: logistic regression, linear discriminant analysis and it’s extension, lab

- Week 6. Resampling methods and variable selection: cross-validation, bootstrap, lab
- Week 7.
  1. Model selection: subset selection and shrinkage
  2. Midterm exam
- Week 8. Regularization, dimension reduction, and lab for big data analysis
- Week 9. Nonparametric smoothing: spline methods with big data
- Week 10. Kernel smoothing: local polynomial smoothing with big data
- Week 11. Decision trees: regression trees and classification trees
- Week 12. Regression trees: bagging and boosting, lab
- Week 13. Support vector machines: one-versus-one and one-versus-all classifications, lab
- Week 14. Unsupervised learning methods: principle component analysis with applications to big data
- Week 15.
  1. Unsupervised learning: Clustering, lab
  2. Final exam