Course Approval Form

For instructions:
http://registrar.gmu.edu/facultystaff/catalog-revisions/course/

Action Requested: (definitions available at website above)
[ ] Create NEW [ ] Inactivate
[x] Modify (check all that apply below)

Course Level:
[x] Undergraduate [ ] Graduate

[x] Title (must be 75% similar to original) [ ] Repeat Status [x] Prereq/coreq [ ] Grade Mode
[ ] Credits [ ] Schedule Type [ ] Restrictions [ ] Other: _____

College/School: College of Science **Department:** Biology
Submitted by: Larry Rockwood **Ext:** 3-1031 **Email:** lrockwoo@gmu.edu

Subject Code: BIOL **Number:** 312
(Do not list multiple codes or numbers. Each course proposal must have a separate form.)

Effective Term: [ ] Fall [x] Spring [ ] Summer *Year* 2017

Title:
Current: Biostatistics
Banner (30 characters max w/ spaces):
New: Biostatistics for Bioinformatics

Fulfills Mason Core Req? (undergrad only)
[ ] Currently fulfills requirement
[ ] Submission in progress

Credits: (check one)
[x] Fixed → 4
[ ] Variable → to
[ ] Lec + Lab/Rct→ 0 or

Repeat Status: (check one)
[x] Not Repeatable (NR)
[ ] Repeatable within degree (RD) →
[ ] Repeatable within term (RT) →
Max credits allowed: (required for RT/RD status only)

Grade Mode: (check one)
[x] Regular (A, B, C, etc.)
[ ] Satisfactory/No Credit
[ ] Special (A, B C, etc. +IP)

Schedule Type: (check one)
LEC can include LAB or RCT if linked sections will be offered
[x] Lecture (LEC)
[ ] Lab (LAB)
[x] Recitation (RCT)
[ ] Internship (INT)
[ ] Independent Study (IND)
[ ] Seminar (SEM)
[ ] Studio (STU)

Prerequisite(s) (NOTE: hard-coding requires separate Prereq Checking form; see above website):
BIOL 214 and CDS 130 with a grade of C or better, or its equivalent, or permission of instructor.

Corequisite(s):

Restrictions Enforced by System: Major, College, Degree, Program, etc. Include Code(s).

Equivalencies (check only as applicable):
[ ] YES, course is 100% equivalent to
[ ] YES, course renumbered to or replaces

Catalog Copy (Consult University Catalog for models)

Description (No more than 60 words, use verb phrases and present tense)	**Notes** (List additional information for the course)
This undergraduate course is designed for students in the Bioinformatics concentration within the BS in Biology major and for students interested in advanced multivariate statistical methods.	

Indicate number of contact hours: Hours of Lecture or Seminar per week: 3 Hours of Lab or Studio: 1
When Offered: (check all that apply) [ ] Fall [ ] Summer [x] Spring

Approval Signatures

Department Approval Date

College/School Approval Date

If this course includes subject matter currently dealt with by any other units, the originating department must circulate this proposal for review by those units and obtain the necessary signatures prior to submission. Failure to do so will delay action on this proposal.

Unit Name	Unit Approval Name	Unit Approver's Signature	Date

Undergraduate or Graduate Council Approval

UGC or GC Council Member

Provost's Office

UGC or GC Approval Date

Form revised 9/2/2016

Course Proposal Submitted to the College of Science Curriculum Committee (COSCC)

The form above is processed by the Office of the University Registrar. This second page is for the COSCC's reference. Please complete the applicable portions of this page to clearly communicate what the form above is requesting.

FOR ALL COURSES (required)
Course Number and Title: BIOL 312

Date of Departmental Approval: September 23rd, 2016

FOR INACTIVATED/REINSTATED COURSES (required if inactivating/reinstating a course)

- Reason for Inactivating/Reinstating:

FOR MODIFIED COURSES (required if modifying a course)

- Summary of the Modification:
  The syllabus has been revised to reflect the new title and the role of this course in the Bioinformatics curriculum within the BS Degree in Biology
- Text before Modification (title, repeat status, catalog description, etc.):
  BIOL 312, 4 credits
  Repeat status: No
  Catalog Description: Use of probability and descriptive and multivariate statistical techniques in interpreting biological data.

- Text after Modification (title, repeat status, catalog description, etc.):
  BIOL 312, 4 credits
  Repeat status: No
  Catalog Description: This undergraduate course is designed for students in the Bioinformatics concentration within the BS in Biology major and for students interested in multivariate statistical methods.
- Reason for the Modification:
  As stated above, the syllabus has been revised to reflect the role of this course in the Bioinformatics Concentration within the BS Degree in Biology.

FOR NEW COURSES (required if creating a new course)

- Reason for the New Course:

- Relationship to Existing Programs:

- Relationship to Existing Courses:

- Semester of Initial Offering:

- Proposed Instructors:

- Insert Tentative Syllabus Below

BIOL 312: Biostatistics for Bioinformatics

Spring 2017
4 Credits

Instructor: Dr. Robert Brown E-Mail: rbrownf@gmu.edu

Office: EXPL suite 1200 Phone: 3-1050
Office hours: TBD

Required Resources

Textbook
TBD
Online resource for multivariate statistics : http://www.statsoft.com/Textbook/

Blackboard
We will use the learning management system Blackboard in this class. Lecture notes, discussion boards, exam preparation materials and more resources will all be posted to this site. Log in at http://mymason.gmu.edu.

Laptop
You should own a laptop that meets or exceeds the specifications set by the biology department. Please bring your laptop to lecture since we will be learning statistical techniques largely with the help of computer software

Software
This course will introduce Python, Perl, and R programming for multivariate statistics. Python and PERL are native installs on much unix platforms including Macintosh.
Python resources can be found at https://www.python.org
Perl resources can be found at http://www.cpan.org
R statistical software is open source and free! It is available for download at http://www.r-project.org/. Please also download the free graphical user interface called RStudio available at http://www.rstudio.com/.

- Instructions for programming will be available throughout the course:
  - In lectures and lecture notes
  - On the problem set assignment pages
  - During office hours (professor and GTA)
  - Internet resources: StackExchange, YouTube, etc.

- There is a steep learning curve to becoming comfortable with scripting languages. Nevertheless, it is required…and a HIGHLY marketable skill!!!

Learning more advanced R: THIS IS OPTIONAL. For students who love R and programming there is a wealth of information out there. If you are one of us, please check out the excellent tutorials by Swirl
- http://swirlstats.com/students.html
- http://www.coursera.org

Course Description

This class will be designated as a Students as Scholars Research Inquiry (RI) Course, which means that you will learn about the recursive process of scholarly inquiry through studying previous scholarship. This will prepare you to conduct original research in future courses. Because this is a methods based course, we expect that you will learn the content and skills enabling you to evaluate scholarly work in the field of bioinformatics. To learn

more about Students as Scholars, visit oscar.gmu.edu. This undergraduate course is designed for students in the Bioinformatics concentration of the Biology Major and students interested in advanced statistical methods in biology. Students will learn various methods in biostatistics that are used in current bioinformatics applications and will read and evaluate previous scholarly works that use these techniques.

Student Outcomes

The student outcomes for this course are those of a research inquiry (RI) course as designated by Students as Scholars.

- Students will articulate and refine a scholarly question
- Students will follow ethical principles
- By learning when to apply various statistical methods, students will learn to choose an appropriate discovery process for scholarly inquiry
- By practicing writing the results of statistical analyses, students will apply appropriate scholarly conventions when reporting or performing
- Through learning the applications of various biostatistics techniques, students will situate the scholarly inquiry within a broader context in their field

Course Structure

Students will attend a series of lectures throughout the semester where they will learn about current biostatistics techniques and their applications in bioinformatics. Students will be assigned weekly problem sets to apply statistical techniques to novel situations. Once monthly, students will read an original scientific article using the technique discussed in lecture and will participate in an online discussion pertaining to the article. In addition, students will take 3 exams covering the mathematics of the statistical techniques and their applications.

Attendance and Late Work Policy

Attendance and performance are highly correlated in any class. Therefore, regular attendance and participation in class are expected. Aside from personal medical emergencies, absences will not be excused. Please let me know about any planned excused absences to make arrangements for making up any missed work in advance. Also note that work-related absences, work in other classes, oversleeping, or meetings with other professors are not considered personal medical emergencies.

All problem sets are due at the beginning of class on the day that they are due. You should NOT expect to be allowed to turn in assignments later in the day for full credit. Late work will not be accepted except in the case of a documented excused absence. You will not be allowed to make up in-class assignments or exams unless you have a documented excused absence. It is your responsibility to provide written documentation from a third party of your emergency or university-excused absence.

Academic Integrity

The integrity of the University community is affected by the individual choices made by each of us. GMU has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow at all times are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students, give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment, ask for clarification. No grade is important enough to justify academic misconduct. Plagiarism means using the exact words, opinions, or factual information from another person without giving the person credit. Writers give credit through accepted documentation styles, such as parenthetical citation, footnotes, or endnotes. Paraphrased material must also be cited, using MLA or APA format. A simple listing of books or articles is not sufficient. Plagiarism is the equivalent of intellectual robbery

and cannot be tolerated in the academic setting. If you have any doubts about what constitutes plagiarism, please see me. For more information about the honor code, please see http://oai.gmu.edu/.

Disability Accommodations

If you have a learning or physical difference that may affect your academic work, you will need to furnish appropriate documentation to the Office of Disability Services. If you qualify for accommodation, the ODS staff will give you a form detailing appropriate accommodations for your instructor. In addition to providing your professors with the appropriate form, please take the initiative to discuss accommodation with them at the beginning of the semester and as needed during the term. Because of the range of learning differences, faculty members need to learn from you the most effective ways to assist you. If you have contacted the Office of Disability Services and are waiting to hear from a counselor, please let your instructor know.

Diversity

George Mason University promotes a living and learning environment for outstanding growth and productivity among its students, faculty and staff. Through its curriculum, programs, policies, procedures, services and resources, Mason strives to maintain a quality environment for work, study and personal growth.

An emphasis upon diversity and inclusion throughout the campus community is essential to achieve these goals. Diversity is broadly defined to include such characteristics as, but not limited to, race, ethnicity, gender, religion, age, disability, and sexual orientation. Diversity also entails different viewpoints, philosophies, and perspectives. Attention to these aspects of diversity will help promote a culture of inclusion and belonging, and an environment where diverse opinions, backgrounds and practices have the opportunity to be voiced, heard and respected.

The reflection of Mason's commitment to diversity and inclusion goes beyond policies and procedures to focus on behavior at the individual, group and organizational level. The implementation of this commitment to diversity and inclusion is found in all settings, including individual work units and groups, student organizations and groups, and classroom settings; it is also found with the delivery of services and activities, including, but not limited to, curriculum, teaching, events, advising, research, service, and community outreach.

Acknowledging that the attainment of diversity and inclusion are dynamic and continuous processes, and that the larger societal setting has an evolving socio-cultural understanding of diversity and inclusion, Mason seeks to continuously improve its environment. To this end, the University promotes continuous monitoring and self-assessment regarding diversity. The aim is to incorporate diversity and inclusion within the philosophies and actions of the individual, group and organization, and to make improvements as needed.

Privacy

Student privacy is governed by the Family Educational Rights and Privacy Act (FERPA) and is an essential aspect of any course. Your instructor cannot discuss your educational record with your parents, your friends, or anyone except for you. Because your GMU email is the only one that we can explicitly identify as belonging to you, your instructor will only communicate with you via email using your GMU address, and your instructor will not discuss grades via email. Students must use their MasonLive email account to receive important University information, including messages related to this class. See http://masonlive.gmu.edu for more information.

Course Assignments and Grades
Your final grade will be a percentage out of 100

Midterm Exam:	25%
Final Exam:	25%
Problem sets:	30%
Discussion Posts	20%

Grades will be distributed in the following manner:

A+	97-100
A	93-96
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	70-76
D	60-69
F	59 or less

Tentative Schedule

Week	**Topics**	**Assignments Due**
1	Review & Baseline Probability Distributions: i) Normal Distribution ii) Binomial Distribution ii) Poisson Distribution Parametric - Non-parametric statistics	
2-3	Regression and Correlation	Problem set 1 – Analysis of 2 parameter systems and ways to relate the direction of the relationship and the impact of one variable on the other versus other confounding variables. Assignments will relate to non-parametric data.
4-6	Analysis of Variance (ANOVA) i) F distribution ii) Model I ANOVA iii) Model II ANOVA iv) Two-way Analysis of Variance	Problem set 2 – Analysis of randomly selected groups (subsets of total) based on their means. Testing to see if they are statistically different from one another. Expanding beyond 2 groups and the associated issues. Use of the null hypothesis to prove or disprove hypothesis within chosen statistical significance.
7	Applications of ANOVA i) MANCOVA ii Maximum Likelihood	**Exam 1 DUE**
8-9	Dealing with Large Data Sets: Data Manipulation and Visualization i) Open Source Tools ii) Large Data Visualization Methods iv) Team Project Development	Problem set 3 – Address issues of dealing with 100s or features for each sample and showing significance between groups per feature. Open source tools and methods to visualize the resulting datasets. Chose teams and data topics for class project and presentation
10-13	Multivariate Data i) Coefficients of Association ii) Cluster Analysis iii) Analysis of Similarities, Mantel's Test iv) Principal Component Analysis (PCA) v) Multiple Discriminant Analysis vi) Krustal Wallis	Problem set 4 Use of dimension reduction and other discrimination techniques to allow for analysis of large multi-dimensional datasets. Apply methods to visualize and determine statistical significance.
14	Applied Multivariate Statistics Team Projects	Problem set 5 Presentation and discussion of team projects and the proper use of the bioinformatic statistics presented in class
	FINAL EXAM	

Description of Assignments
The course assignments are to allow the student to apply the techniques in class with a goal of their recognition of the proper methods to apply to a specific dataset-- and recognize the methods limitations. To develop and innate understanding of why a method is appropriate, or not, and is not to insist on wrote memorization of formulas. Enough work must be presented to ascertain the student understood the assignment.
Today knowledge of the underlying reasoning is more important than the how, which can be found online.
Teamwork will be applied to projects, to instill to sample research or work environment where each team member brings unique skills to the task.

Scientific Literature Discussion Posts
You will have the opportunity to see the statistical techniques in action while reading 4 scientific articles over the semester course. Each will be based on the lecture topic from the previous week. Discussion questions will be posted along with the article so that you may think critically as you read the scholarly article. A 2-3 paragraph post detailing your answers to the discussion questions is due at midnight the day before the following class. Please work to analyze the article and synthesize the findings with other knowledge. Do not simply summarize the article in your discussion post.

Problem Sets

Throughout the semester, you will complete five problem sets each worth 6% of the final course grade. These problem sets will enable you to apply what you learned about a statistical technique in lecture to novel situations at home to help solidify your knowledge of the test and its applications. These problem sets will also drastically improve your R skills. Datasets used for homework problem sets will be heavily based on real life data. Therefore, there will be steps needed to clean and massage the data prior to analysis. These skills are critical to biostatistics because real data are messy! Scripting code will be presented along with examples of each statistical technique during lecture.

Exams
This course has one take home midterm exam and one final exam. Questions will cover the theoretical basis of each statistical technique, the mathematics involved in the technique (where applicable), and will mainly focus on the correct application of the technique. For the take home midterm, you will submit your annotated scripting code and resulting output. Be sure to correctly interpret the output in your annotations for each analysis. The two midterms will not be cumulative but the final exam will be.