

**PHC 6099 - R for Applied Biostatistical Methods**

**3 Credits**

**Course Syllabus**

|  |
| --- |
|  |
| **GENERAL INFORMATION** |

**Instructors**: Gabriel Odom

**Teaching Assistant(s)**: TBA

**Office**: AHC5-470

**Phone**: (305) 348-5486

**E-mail**: [Gabriel.odom@fiu.edu](mailto:Gabriel.odom@fiu.edu)

**Office Hours**: By appointment only

**Class Meets**: Tuesdays and Thursdays, 1330-1505, Owa Ehan, #222

|  |
| --- |
|  |
| **PREREQUISITE(S)** |

* Prior enrollment to PHC 6052 Biostatistics I or equivalent statistics course (with permission)
* Prior enrollment to PHC 6701 R for Health Data Science or equivalent data science course (with permission)
* Prior (or concurrent) enrollment to PHC 6091 Biostatistics II or equivalent statistics course (with permission)

|  |
| --- |
|  |
| **DESCRIPTION** |

This course will connect the R statistical software computing learned in an introductory data science course with statistical methods covered in Biostatistics I and II. R is an extremely versatile and powerful statistical programming language that is becoming very popular among researchers in virtually every research realm. Topics will include (but not limited) to inputting/reading data, calculation of descriptive statistics, data visualization, t-tests, confidence intervals, chi-square test, analysis of variance (ANOVA), simple and multiple regression, logistic regression, and non-parametric methods. This course is designed to enrich computing and analytical skills.

This is a course in the “flipped classroom” style. Students will read and practice code outside of class (with guidance available from instructors), and then students will present their code and reports during the scheduled class time to the instructors and their peers. We will meet twice per week for some in-class examples, but primarily to present code and discuss results. This class will be an opportunity for students to practice using nearly three dozen common statistical techniques on real-world data. This course design will prepare you for what statistical techniques you should expect to use as entry-level biostatisticians.

|  |
| --- |
|  |
| **COURSE OBJECTIVES** |

Demonstrate proficiency in the R programming skills for data processing, data management, and statistical analyses of research data relevant to public health.

Upon completion of this course, the student will be able to:

1. Read data in variety of formats and manage a dataset in R
2. Apply functions and procedures in order to calculate/produce various numerical and graphical summary statistics for univariate, bivariate and multivariate continuous and categorical data.
3. Perform hypothesis testing in R including one and two sample t-tests, confidence intervals, chi-square goodness-of-fit and testing for independence, ANOVA, correlations, and non-parametric tests.
4. Perform regression analyses using R: simple linear regression, multiple linear regression, and logistic regression.
5. Analyze and interpret real data using R.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| M | | | | |
| **MPH CORE / CONCENTRATION-SPECIFIC COMPETENCIES**  This course will assist students in developing the following MPH core competencies in biostatistics: | | | | |
|  | **COMPETENCY** | **COURSE OBJECTIVE** | **LEARNING OPPORTUNITY** | **ASSESSMENT OPPORTUNITY**  **(GRADED)** |
|  | 3. Produce advanced data science tools in the analysis of complex medical and public health data | 1-5 | All lectures. See details below. | Weekly presentations on statistical methods |
|  | 5. Effectively explain advanced biostatistics and data analytics topics to various audiences | 1-5 | All lectures. See details below. | Weekly presentations on statistical methods |

Competencies Rating Scale (See page 5 for details):

0 – Poor (does not meet expectations)

1 – Developing (partially meets expectations)

2 – Adequate (meets expectations)

3 – Excellent (exceed expectations)

|  |
| --- |
|  |
| **TEXTBOOK** |

Course materials:

1. Setting up your computer:
   1. PC: <https://derailment.netlify.app/2019-12-10-installing-r-rstudio-on-windows/>
   2. Mac: <https://derailment.netlify.app/2019-11-16-installing-r-rstudio-on-a-mac/>
2. Cetinkaya-Rundel & Hardin. Introduction to Modern Statistics, 1st Ed. <https://openintro-ims.netlify.app/>
3. Verzani, J. simpleR – Using R for Introductory Statistics. FREE Lab Manual I.
4. Verzani J. Using R for Introductory Statistics. FREE BOOK.
5. Downloaded copy of R and R Studio. FREE.
6. Personal notebook computer for use in class is highly recommended.

Web Resources:

[**https://www.r-project.org/**](https://www.r-project.org/)

[**https://www.rstudio.com/**](https://www.rstudio.com/)

[**https://www.statmethods.net/**](https://www.statmethods.net/)

[**https://stats.idre.ucla.edu/r/**](https://stats.idre.ucla.edu/r/)

[**https://paulvanderlaken.com/2017/08/10/r-resources-cheatsheets-tutorials-books/**](https://paulvanderlaken.com/2017/08/10/r-resources-cheatsheets-tutorials-books/)

|  |
| --- |
|  |
| TOPICS COVERED |

**See class schedule below.**

|  |
| --- |
|  |
| **CLASS POLICIES AND PROCEDURES** |

The class will meet twice per week for 1 hour and 35 minutes each day.

**Required Presentations**

Each class period, you (the students) will teach us and your classmates how to perform statistical techniques in R. Your presentation will be in the form of a Quarto document with the necessary code shown, and this presentation should be between 15-25 minutes plus time for questions (some methods will need more time than others). Your rendered Quarto document should be published to your website (quartopub being the easiest option); you will email me the final quarto document, its rendered output, and all figures/data files necessary to build it. Show the data cleaning and pre-processing steps in a “hidden” (collapsed) chunk, but you should show:

* + - 1. Introduction to the method
      2. Mathematical definition of your method
      3. the source of the raw data file
      4. the cleaned columns from the data which are necessary to execute your method
      5. assumptions of the method
      6. plots necessary to check the assumptions of the method
      7. the code to run the method
      8. the output of the code
      9. brief interpretation of the output

**Further Details**

R is a free program. It is available at the computer lab, and you can also download and install it on your personal computer or laptop (which is highly recommended). There will be weekly reports to present in class and a final analytical assignment accompanied by the report and presentation. See evaluation for details.

If safety becomes a concern due to severe weather, we will follow FIU policies and procedures and the course schedule will be adjusted by the instructor. If safety is a concern due to COVID-19, we will work out some scenario where affected students can join the class via Zoom.

It is the student’s responsibility to initiate the paperwork required to drop or withdraw from courses. Failure to attend classes does not constitute proper procedure for dropping or withdrawal and may result in an F.

As a courtesy to your fellow students and the instructor please put your mobile phones on silent, and if you have to answer a phone call as an emergency please step out of the classroom or a lab.

**Disability**

The University provides reasonable accommodations and services to all students on a nondiscriminatory basis consistent with legal requirements as outlined in the Americans with Disabilities Act as Amended (ADAAA) of 2008 and the Rehabilitation Act as Amended. If you have disability and/or need special assistance during the course or exams, please make arrangements through the Office of Disability Services (305-348-4131). More information about the accommodation process is available at <http://drc.fiu.edu/index.php>

|  |
| --- |
|  |
| **EVALUATION AND GRADING** |

Methods of evaluation:

1. 3 presentations worth 300 points each, for a total of 900 points.
2. 1 final presentation worth 300 points

All presentations will be assigned during the first week of class. Students should start working on the presentation for their assigned method immediately. Students will present a rendered Quarto code document during the class for which that method belongs. Students are encouraged to save all their programs for future use.

For the final analytical assignment report and presentationstudents will find a publicly available data set (many are included in R packages) and analyze it using several of the methods learned in class (descriptive statistics, graphical presentation of data, modeling). If student has a difficulty finding a data set instructor will assign one. The report and presentation should include statement of the problem(s), brief description of the variables in the data set including dependent and independent variables, analytical method(s) of choice and justification, results and conclusions.

|  |  |
| --- | --- |
| Homework assignments and competency assessments\* | 75% |
| Final Analytical Assignment and Presentation | 25% |
| \*Competency Assessments: Presentation assignments will require that you 1) prepare the data for analysis, check the assumptions of the specified model, 3) execute the model and include the model output, and 4) write a short paragraph describing the results and conclusions based on the analytics that were performed. The competencies will be assessed using rubrics above. These homework and competency assessments will constitute 70% of the total grade. | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Grading Scale** | | | |
| A | 92.6 – 100 | C+ | 76.6-79.5% |
| A- | 89.6 – 92.5 | C | 69.6-76.5 |
| B+ | 86.9 – 89.5 | D | 59.6-69.5 |
| B | 82.6 – 86.5 | F | < 59.5 |
| B- | 79.6 – 82.5 |  |  |

|  |
| --- |
|  |
| **FIU HONOR CODE** |

Florida International University is a community dedicated to generating and imparting knowledge through excellent teaching and research, the rigorous and respectful exchange of ideas and community service. All students should respect the right of others to have an equitable opportunity to learn and honestly to demonstrate the quality of their learning. Therefore, all students are expected to adhere to a standard of academic conduct, which demonstrates respect for themselves, their fellow students, and the educational mission of the University. All students are deemed by the University to understand that if they are found responsible for academic misconduct, they will be subject to the Academic Misconduct procedures and sanctions, as outlined in the Student Handbook.

Misconduct includes: Cheating – The unauthorized use of books, notes, aids, electronic sources; or assistance from another person with respect to examinations, course assignments, field service reports, class recitations; or the unauthorized possession of examination papers or course materials, whether originally authorized or not. Plagiarism – The use and appropriation of another’s work without any indication of the source and the representation of such work as the student’s own. Any student who fails to give credit for ideas, expressions or materials taken from another source, including internet sources, is responsible for plagiarism. All students are expected to abide by the Florida International University Honor Code. Any violation will be reported.

In addition, students are required to sign an honour code at the end of all assignments and exams. The honour code should state “*I have neither given nor received any unauthorized aid on this [assignment, exam]*”. Please sign after the statement.

**Student conduct**

**Students are responsible for knowing and complying with all FIU Policies and Regulations which are listed in the Student Handbook and also at the following link: http://policies.fiu.edu/files/740.pdf. The following are excerpts:**

**Reason for the policy:**

**“**Graduate students at Florida International University are expected to adhere to the highest standards of integrity in every aspect of their lives. Honesty in academic matters is part of this obligation. Academic integrity is the adherence to those special values regarding life and work in an academic community. Any act or omission by a graduate student which violates this concept of academic integrity and undermines the academic mission of the University shall be defined as academic misconduct and shall be subject to the procedures and penalties that follow.”

**Definition of academic misconduct:**

Academic misconduct is defined as the following intentional acts or omissions committed by any FIU graduate student:

***“Cheating***: The unauthorized use of books, notes, aids, electronic sources; or unauthorized use of on-line exams, library materials or assistance from another person with respect to examinations, course assignments, field service reports, class recitations; or the unauthorized possession of examination papers (or on-line examinations) or course materials, whether originally authorized or not. Any student helping another cheat may be found guilty of academic misconduct”

***Plagiarism***: The deliberate use and appropriation of another's work without any indication of the source and the representation of such work as the student's own. Any student, who fails to give credit for ideas, expressions or materials taken from another source, including internet sources, is guilty of plagiarism. Any student helping another to plagiarize may be found guilty of academic misconduct.

***Misrepresentation***: Intentionally lying to a member of the faculty, staff, administration, or an outside agency to gain academic advantage for oneself or another, or to misrepresent or in other ways interfere with the investigation of a charge of academic misconduct.

***Misuse of Computer Services***: The unauthorized use of any computer, computer resource or computer project number, or the alteration or destruction of computerized information or files or unauthorized appropriation of another's program(s).

***Bribery***: The offering of money or any item or service to a member of the faculty, staff, administration or any other person in order to commit academic misconduct.

Conspiracy and Collusion: The planning or acting with one or more fellow students, any member of the faculty, staff or administration, or any other person to commit any form of academic misconduct together.

***Falsification of Records***: The tampering with or altering in any way of any academic record used or maintained by the University.

**Ethical Guidelines for Statistical Practice:** “Statistics play a vital role in many aspects of science, economy, governance, and even entertainment. It is important that all statistical practitioners recognize their potential impact on the broader society and the attendant ethical obligations to perform their work responsibly. Furthermore, practitioners are encouraged to exercise "good professional citizenship" in order to improve the public climate for, understanding of, and respect for the use of statistics throughout its range of applications” (American Statistical Association)

**AI Policy**

Within this class, you are welcome to use Large Language AI models (ChatGPT, GPT, DALL-E, Stable Diffusion, Midjourney, GitHub Copilot, and anything after) in a totally unrestricted fashion, for any purpose, at no penalty. However, you should note that all large language models still tend to make up incorrect facts and fake citations, code generation models have a tendency to produce inaccurate outputs, and image generation models can occasionally come up with highly offensive products. You will be responsible for any inaccurate, biased, offensive, or otherwise unethical content you submit regardless of whether it originally comes from you or an AI/large language model. If you use an AI model, its contribution must be acknowledged in the submitted work; you will be penalized for using an AI model without acknowledgement. Proper acknowledgement includes: the name and hyperlink to access the AI, the version and license of the AI model, and (when reasonable) the prompt used to generate the results. The university's policy on plagiarism still applies to any uncited or improperly cited work, whether created by other human beings or synthesized by an AI solution. (*Adapted from Ryan S. Baker, Creative Commons License 4.0.*)

|  |  |  |
| --- | --- | --- |
|  | | |
| **WEEK** | **TOPIC** | **REPORTS** |
| 1 | Review of R, tidyverse, Tidy Data principles, and Quarto. | No reports, pick teams instead; ensure that you have R/RStudio updated. |
| 2 | EDA and Graphics:   * skimr * table1 * gtsummary * ggplot2 | 1. Mosaic plots and box/violin plots 2. Scatterplots (with facets) 3. 3-D densities/Hex plots/rayshader 4. How to explore data with skimr 5. Demographics tables with table1 6. Analysis tables with gtsummary |
| 3 | One- and Two-Feature Tests:   * One-sample continuous * One-sample categorical * Transformations * Two-sample categorical | 1. Z-test 2. One-sample/paired t-test 3. One-sample/paired Wilcoxon 4. Transformations to Normality 5. McNemar’s Test 6. Fisher’s Exact Test 7. Chi-Square Goodness of Fit 8. Permutation tests |
| 4 | Two- and Three-Feature Tests:   * Two-sample continuous * Three-feature categorical * Mixed type tests | 1. Two-sample t-test 2. Two-sample Welch’s test 3. Mann-Whitney U test 4. Cochran’s Q test 5. Chi-Square Independence 6. One-Way ANOVA |
| 5 | Three-Feature Tests and Linear Models   * Two-Way Methods * Intro to Linear Models | 1. Two-way ANOVA 2. Welch’s ANOVA 3. Kruskal-Wallace Test 4. Durbin’s / Friedman’s Test 5. Repeated Measures ANOVA 6. Random Intercept (Effect) Models |
| 6 | Regression and Advanced Models   * Previous Methods as Regression * OLS and Correlation * Linear Models | * + - 1. Update your previous reports to show how linear regression could be applied to answer the same question [ALL student teams]       2. Correlation Matrices and Covariances       3. Multiple Regression       4. Polynomial regression       5. Generalized Linear Models: Binary       6. Generalized Linear Models: Ordered       7. Generalized Linear Models: Count |
| 7 | Power and Sample Size; Introduction to the Tidymodels Syntax | 1. FINAL PRESENTATION: Assignments on Power and Sample Size from Zoran Bursac |
| 8-11 | Special Topics | TBD |
| 12 | Final presentation / exam | Presentation on Power and Sample Size assignments from Zoran Bursac |

**NOTE**

The schedule and assignments are subject to change.