American University of Beirut

Faculty of Health Sciences

**EPHD 312**

**Analysis of Continuous data**

**[2 credits]**

**Course Syllabus**

**Fall Semester, Academic Year 2023-2024**

**Class Time and Venue:**

Wednesday 4:00 pm – 5:30 pm (Vandyck 201)

Friday 4:00 pm – 5:30 pm (Vandyck 201)

**Course Instructors and Contact Details:**

Khalil El Asmar (PhD)

Assistant Professor

Office: VanDyck Hall 215

Email: ke05@aub.edu.lb

Office hours: By appointment

# **Course Description**

In this course students will learn to use regression analysis to address a research question. It covers basic exploratory data analysis for univariate (outcome) continuous observations with single or multiple covariates, followed by regression methods and diagnostics with a main focus on multiple regression. The emphasis of the course is on the application of statistical techniques that are carried out using the statistical platform Stata. Lectures include lab sessions, article reading and appraisal as well as group discussions.

Total Credits: 2

* Lecture hrs: Friday (16:00 to 17:30)
* Lab/Lecture hrs: Wednesday (16:00 to 17:30)

**Course Learning Objectives:**

**By the end of the course, students will be able to:**

1. Write programing code in Stata to analyze epidemiological and clinical data
2. Adapt Stata codes to manage cross sectional and longitudinal epidemiological and clinical data.
3. Explain and implement various statistical tests and regression techniques.
4. Diagnose violation of regression assumptions in a regression framework.
5. Prepare detailed and objective statistical reports to communicate results to scientific and lay audience
6. Create custom data visualizations.
7. Synthesize statistical findings in short oral presentations.

**MPH-EPBS Concentration Competencies mapped to EPHD 312**

* EPHD.CC4: Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software
* EPHD.CC5: Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data
* EPHD.CC6: Interpret and communicate statistical findings in oral and written format

**Assigned Text Book and Readings**

Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al.

**Course requirements and Student evaluation:**

**Pre-requisites: EPHD 310**

**Attendance**:

" Students who miss more than one-fifth of the sessions of any course in the first ten weeks of the semester (five weeks in the case of the summer term) are dropped from the course if the faculty member has stated in the syllabus that attendance will be taken."

**Student Evaluation:**

Students will be evaluated on the following:

- 2 data analysis Projects (50%: 25% P1, 25% P2)

- 2 assignments (20%) meet objectives 1-4

- knowledge check (I) (15%)

- knowledge check (II) (15%)

Table-1 Summary of students' assessments mapped to course learning objectives

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | |
|  | LO1 | LO2 | LO3 | LO4 | LO5 | LO6 | LO7 |
| Assessment Method #1: First data analysis project (20%). | X | X |  |  | X | X |  |
| Assessment Method # 2: Second data analysis project (25%) | X | X |  | X | X | X | X |
| Assessment Method #3: Assignment 1 (10%) | X |  |  |  |  |  |  |
| Assessment Method #4: Assignment 2 (10%) |  |  | X | X |  |  |  |
| Assessment method #5; knowledge check 1(15%) |  |  | X |  |  |  |  |
| Assessment method #6; knowledge check 2(15%) |  |  | X | X |  |  |  |

Table -2 Description of Assessment methods, Due Dates and Corresponding Learning Objectives

| **Assessment method** | | **Date (tentative)** | **LOs covered** | | **Grade percentage** | |
| --- | --- | --- | --- | --- | --- | --- |
| *Project 1* | | *October 6* | *LO 1, LO 2, LO 5, LO 6* | | *20%* | |
| In this first individual project, you will analyze a clinical data to answer a pre-specified research question. The aim of this project is to focus on your Stata coding and focus on writing a correct code. You will also be doing inferential statistics at the bivariate (t-tests, ANOVA and pearson correlations) as well as well the multivariate (single and multiple regression analysis) level. The focus of this project will be on running proper regression analysis using Stata and on data visualization. | | | | | | |
| *Project #2:* | | *December 1* | *LO 1, LO 2, LO 4, LO 5, LO6, LO7* | | *25%* | |
| In the second individual project you will be provided with a data from a prospective cohort, with its entire metadata. You will develop your research question and analysis plan. After discussion and approval from the course instructor, you will analyze your data using advanced regression techniques covered in the course. A special attention will be given to regression diagnostics. You will write a detailed statistical report and prepare a 15 minutes power point presentation to communicate your findings in lay terms. | | | | | | |
| *Assignment #1* | *September 22* | | | *L01* | | *10%* |
| In this first assignment you will use simple regression analysis, to test associations between an outcome of interest and a single exposure. You will also test the validity assumption of linear regression. You will use Stata to perform bivariate and basic regression analysis in addition to basic data management. | | | | | | |
| Assignment #2 | Nov 1 | | | LO3 LO4 | | 10% |
| In this second assignment you will use regression analysis, to test associations between an outcome of interest and different exposures. You are expected to manually calculate Partial F tests in this assignment. The assignment will also cover regression diagnosis and interpretation of regression models. | | | | | | |
| Knowledge check #1 | Oct 18 | | | LO3 | | 15% |
| This first exam will assess your understanding of various statistical tests and regression techniques, using a set of problem solving exercises. You will be expected to derive regression equations, interpret statistical outputs. You are expected to use Stata as part of this exam. | | | | | | |
| Knowledge check #2 | TBA | | | LO3 LO4 | | 15% |
| The second exam will assess your understanding of advanced regression techniques such as interactions, model buildings and regression diagnostics. The exam is comprised entirely on problem solving and interpretation of results using existing regression outputs from Stata. | | | | | | |

**Policies and other General Notes:**

|  |
| --- |
| **Moodle Surveys**  This course uses Moodle as the Learning Management system. Your AUBnet account will allow you access to the course content and activities, by logging into: http://moodle.aub.edu.lb |
| **Academic Integrity**  Education is demanding and time management is essential. Do not hesitate to use the resources around you but do not cut corners. Cheating and plagiarism will not be tolerated. Please review the Student Code of Conduct in your handbook and familiarize yourself with definitions and penalties. If you’re in doubt about what constitutes plagiarism, ask your instructor because it is your responsibility to know. The American University of Beirut has a strict anti-cheating and anti-plagiarism policy. Penalties include failing marks on the assignment in question, suspension or expulsion from University and a permanent mention of the disciplinary action in the student’s records  **Students with Disabilities**  AUB strives to make learning experiences accessible for all. If you anticipate or experience academic barriers due to a disability (including learning difficulties, mental health, chronic or temporary medical conditions), please inform me immediately so that we can privately discuss options.  In order to ensure that you receive the support you need and to facilitate a smooth accommodation process, you are encouraged to register with the Accessible Education Office (AEO): [accessibility@aub.edu.lb](mailto:accessibility@aub.edu.lb); +961-1-350000, x3246; West Hall, 314’. |
| **Non-Discrimination – Title IX – AUB**  AUB is committed to facilitating a campus free of all forms of discrimination including sex/gender-based harassment prohibited by Title IX. The University’s non-discrimination policy applies to, and protects, all students, faculty, and staff. If you think you have experienced discrimination or harassment, including sexual misconduct, we encourage you to tell someone promptly. If you speak to a faculty or staff member about an issue such as harassment, sexual violence, or discrimination, the information will be kept as private as possible, however, faculty and designated staff are required to bring it to the attention of the University’s Title IX Coordinator. Faculty can refer you to fully confidential resources, and you can find information and contacts at[www.aub.edu.lb/titleix](https://www.aub.edu.lb/titleix). **To report an incident**, contact the University's Title IX Coordinator Trudi Hodges at 01-350000 ext. 2514, or [titleix@aub.edu.lb](mailto:titleix@aub.edu.lb). An anonymousreportmay be submittedonline via EthicsPoint at [www.aub.ethicspoint.com](https://www.aub.ethicspoint.com/). |

**Detailed course outline:**

Schedule of Lectures, Practical Sessions, Readings and Assessments

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Detailed course schedule** | **Course learning objective** | **CEPH or distinct competency** |
| Week 1  (Aug 28 – Sep 1) | Syllabus+ Introduction to simple Linear regression | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Inferences for slope and intercept in a simple linear regression + intro to project 1 | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Week 2  (4-8 Sep) | No Lab: Online activity work on Project 1 | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Application session simple linear regression, ANOVA, hypothesis testing | LO1, LO2, LO6 | Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Week 3  (11-15 Sep) | Confidence intervals and prediction intervals | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Lab session: application session | LO1, LO2, LO6 | Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Week 4  (16-22 Sep) | No Lab session: take home assignment 1 | LO1, LO2, LO6 | Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Extrapolation + correlation coefficients | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Week 5  (25-29 Sep) | No Class: Prophet’s birthday |  |  |
| Simple linear regression assumption checking | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Week 6  (2 – 6 Oct) | Project#1 (due date Oct 6) | LO1, LO2, LO6 | Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Week 7  (9-13 Oct) | Multiple linear regression | LO1, LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Application on multiple linear regression + intro to project #2 | LO1, LO2, LO6 | Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Week 8  (16-20 Oct) | Exam 1 (October 18) | LO1, LO2, LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Partial F-test | LO1, LO2, LO6 |  |
| Week 9  (23-27 Oct) | Application session | LO1, LO2, LO3, LO4 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Regression diagnostics (1) | LO4 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Week 10  (Oct 30 – Nov 3) | Take home: Assignment 2 | LO1, LO2, LO3, LO4 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Regression diagnostics (2) | LO1, LO2, LO6 | Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Week 11  (6 – 10 Nov) | No Classes. Work on Project #2 | LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data  Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software |
| Week 12  (13-17 Nov) | Confounding and interaction | LO1, LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Dummy variables in regression | LO1, LO3 | Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |
| Week 13 (Nov 20- 24 Nov) | Application session | LO1,LO2, LO3 | Synthesize statistical findings in short oral presentations. |
| Week 14 (Nov 25- 1 Dec) | Presentation of Project 2 | LO7 | Synthesize statistical findings in short oral presentations. |

\*knowledge check 2 to be decided by the registrar.

**Full reference for readings – organized by sequential topics**

* Topic 1: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 5: Straight line regression analysis
* Topic 2: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 6: The correlation coefficient and straight line regression analysis
* Topic 3: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 7: The analysis of variance table
* Topic 4: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 8: Multiple regression analysis
* Topic 5: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 9: Statistical inference in multiple regression
* Topic 6: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 11: Confounding and interaction in regression
* Topic 7: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 12: Dummy variables in regression
* Topic 8: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 14: Regression diagnostics
* Topic 9: Applied Regression Analysis and Other Multivariable Methods, 5th Edition, Kleinbaum et al. Chapter 16: Selecting the best regression equation

**Appendix –I Mapping of Course Learning Objectives to MPH-EPBS Concentration Competencies**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | LO1 | LO2 | LO3 | LO4 | LO5 | LO6 | LO7 | Other CORE and/or Concentration courses that also address competency) | Assignment where Competency primarily assessed |
| Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software | A | A |  |  |  |  |  |  | Project 1 & 2 (individual) assignment 2 |
| Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data |  |  | A | A |  |  |  |  | Assgt 1& 2 /Midterm and final exam |
| Interpret and communicate statistical findings in oral and written format |  |  |  |  | A | A | A |  | Individual presentations and report based on projects 1 and 2 |

A=Assessed in EPHD 312