AMERICAN UNIVERSITY OF BEIRUT
FACULTY OF HEALTH SCIENCES

Department of Epidemiology and Population Health

EPHD 313
Analysis of Categorical Data
[3 credits]

Course Syllabus
Spring Semester, Academic Year 2023-2024

Class time and Venue:

Lectures:
Date: Tuesdays
Times: 1:30 PM - 3:10 PM
Venue: Classroom 201, Van Dyck Building

Labs:
Date: Thursdays
Times: 1:30 PM - 3:30 PM
Venue: Computer Lab, Van Dyck Building

Course Instructors and Contact Details:

Name: Dr. Jaffa M.
Email: ms148@aub.edu.lb
Extension: 4603
Office Hours: Tuesdays and Thursdays after class or by appointment

Course Description:

This course aims at introducing biostatistical approaches to analyze categorical data. In particular, students will learn about (1) analysis of two-way contingency tables for independent and matched study designs, (2) analysis of three-way contingency table, (3) logistic regression for independent and for matched data, data with small sample size and rare events, (4) multicategory logit for nominal and ordinal outcomes, (5) analysis of count data through Poisson and negative binomial regression, (6) analysis using generalized linear models (GLM). The statistical package STATA will be used in this course.

Course learning Objectives

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

LO1. Analyze data summarized in contingency tables using proper tests of independence, and measures of association.

LO2. Conduct rigorous logistic regression appropriate to study design.

LO3. Conduct rigorous multicategory logit regression appropriate to the outcome.

LO4. Conduct rigorous Poisson or negative binomial regression for count outcome.

LO5. Conduct analysis using generalized models for binary and count outcomes,

LO6. Compute, and interpret estimates for population parameters and corresponding confidence intervals.

LO7. Utilize "STATA" proficiently to conduct statistical analysis.

LO8. Communicate and interpret the biostatistical results in a proper written and/or oral format.

Council on Education for Public Health (CEPH) Core/Concentration Competencies mapped to EPHD 313

- **EPHD.CC4:** Demonstrate ability to write software codes in order to manage and analyzehealth data through the use of multiple statistical software (LO7)

- **EPHD.CC5:** Apply inferential statistics and advanced statistical approaches such as regressionmodelling to analyze complex health related data (LOs 1-6)

- **EPHD.CC6:** Interpret and communicate statistical findings in oral and written format (LO8)

Essential Skills

- **Essential Skill 1: Critical Thinking/analysis**
- **Essential Skill 2: Problem Solving**
- **Essential Skill 3: Public Health Ethics**

Link to PHEO Faculty Portal

Course Learning Objectives mapped to CEPH competencies

Table 1. Mapping of course LO to CEPH competencies

	LO1	LO2	LO3	LO4	LO5	LO6	LO7	LO8
EBC4: Analyze health related data using advanced statistical techniques and software	X	X	X	X	X	X	X	
EBDC4. Demonstrate ability to write software codes in order to manage and analyze heath data through the use of multiple statistical software							X	
EBDC5. Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data		X	X	X	X	X	X	
EBDC6. Interpret and communicate statistical findings in oral and written format								X

Course requirements and Student evaluation:

Pre-requisites: EPHD310 Basic Biostatistics course, EPHD300 Principles of Epidemiology, and knowledge of different Epidemiological study designs. Otherwise permission of the course director is needed in order to register in the course.

Student Evaluation:

Table-2 Summary of students' assessments mapped to course learning objectives

	Learning Objectives							
	LO1	LO2	LO3	LO4	LO5	LO6	LO7	LO8
Test 1	X					X		
Test 2		X	X	X	X	X		
HW 1	X						X	
HW 2		X					X	X

Table -3 Description of Assessment methods, Corresponding Learning Objectives, and Tentative Due Dates that are subject to change if needed

Assessment method	Date (tentative)	Grade percentage
Test 1	*Saturday March 2 (week 7)*	*44%*
Covering lectures 1, 2a, 2b, 2c, 3		
Test 2	*TBD*	*48%*
Covering Lectures 4a, 4b, 5a, 5b, 6a, 6b, 7a, 7b, 8a, 8b, 9		
HW 1	*Thursday Feb 22 (week 6)*	*4%*
Covering lectures 1, 2a, 2b, 2c, 3		
HW 2	*Thursday April 4 (week 12)*	*4%*
Covering lectures 4a, 4b, 5a, 5b, 6a, 6b		

Policies and other General Notes:

Academic integrity:

Education is demanding and you need to properly manage your time. Do not hesitate to use the resources around you but do not cut corners. Cheating and plagiarism will not be tolerated. Review the Student Code of Conduct and familiarize yourself with definitions and penalties. Cheating might earn you a failing mark on the assignment, at the very least. You might fail the course in which you cheated, be warned, suspended or expelled from University and a permanent mention of the disciplinary action might be made in your student records. If you're in doubt about what constitutes plagiarism, ask your instructor because it is *your* responsibility to know. Remember that the American University of Beirut has a strict anti-cheating and anti-plagiarism policy. Do not become a lesson to others. For further information, kindly visit AUB's Policies and Procedures or http://pnp.aub.edu.lb/general/conductcode.

Students with Disabilities:

If you have a disability, for which you may request accommodation in AUB classes, consult the website for more information and make arrangements with the Coordinator (http://www.aub.edu.lb/sao/Pages/Students_20with_20Special_20Needs.aspx). Also, please see the instructor of this course privately in regard to possible support services that can be provided to you.

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. To report an incident, contact the University's Title IX Coordinator Trudi Hodges at 01-350000 ext. 2514, or titleix@aub.edu.lb. An anonymous report may be submitted online via EthicsPoint at www.aub.ethicspoint.com.
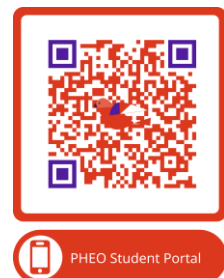
Accessible Education Office (AEO):

The Accessible Education Office (AEO) coordinates academic accommodations and services for all eligible AUB students with disabilities (such as ADHD, learning difficulties, mental health conditions, chronic or temporary medical conditions, and others). If you have a disability for which you wish to request accommodations at the department, faculty or university level, please contact AEO as soon as possible. Once you register with our office, we will assist you in receiving appropriate accommodations and will liaise with your instructors and any related entity to best support your needs. AEO is located in West Hall room 314, and can be reached by phone at 1-350000 ext. 3246 or by email: accessibility@aub.edu.lb. Information about our services can be found at: https://www.aub.edu.lb/SAO/Pages/Accessible-Education.aspx

Writing:
Written communication is essential for communication, health education and behavioral science. You are expected to proofread and spell-check any written documents before submission. Points will be deducted from the grades for low quality writings. You are encouraged to contact AUB's Writing Center, located in Ada Dodge Hall, 2nd floor or West Hall, 3rd floor. Appointments can be booked online: aub.mywconline.com, over the phone (Ext. 4077) or by walking in.

Public Health Education Office
Please refer to the Public Health Education Office Student Portal:
https://sites.aub.edu.lb/fhspheostudent/

Detailed course outline:

Week	Dates	Topic	Relevant Assignment (where learning on this topic will be assessed)	Course learning objective e LO
Week 1- Week 5	Jan 18-Feb 15	**Introduction to the course and to STATA** Discussion of LOs and course syllabus. Introductory STATA lab **Introductory lecture** Categorical variables, probability concepts and notations **Contingency tables** Introduction to contingency tables. Probability structure of contingency tables: Joint probability, Marginal probability, Conditional probability. Independence tests for nominal variables: Pearson's Chi-Square test (Used for large sample size), Likelihood Ratio test (Used for large sample size), Fisher's exact test (Used for small sample size). Measures of association for nominal variables: Difference in proportions, Odds Ratio, Relative Risk, Cramer's V. Measures of Association for ordinal variables: Goodman and Kruskal Gamma "Gamma" kendal's Tau_b. Hypothesis Testing for ordinal association (linear trend) using Goodman and Kruskal Gamma "Gamma" kendal's Tau_b. Measures of Association for dependent samples: Percent agreement Kappa statistic Hypothesis testing for dependent samples McNemar's test of association. Measures of Association between ordinal predictor X and binary outcome Y using Somers'D measure of Association Hypothesis Testing for trend between ordinal predictor and binary outcome using Cochrane-Armitage trend test. Three-Way Contingency Table: testing for interaction, confounding and M-H OR estimates	Midterm 1	LOs 1, 6

Week 7	Saturday March 2	**Midterm 1 Lect 1, 2a-c, 3**	Midterm 1	LO 1, 6
Week 6	Due Thursday Feb 22	**HW 1 Lect 1, 2a-c, 3**		LO 1, 7
Week 6- Week 9	Feb 20– March 12	**Simple Logistic Regression** Simple Logistic regression modelIf x is dichotomous If x is continuousIf x is categorical Inference for logistic regression Wald Test Likelihood Ratio Test Model checking (Goodness of Fit) Logistic regression for small sample size (exact logistic regression) Logistic regression for rare events **Multiple Logistic Regression and Model Selection** Multiple Logistic Regression and interpretation. Testing the Significance of the model using Likelihood Ratio Test for the overall model. Individual Wald Tests. Test for Comparing Nested Models. Interaction. Interpretation and Significance. Model building	Midterm 2	LO 2
Week10- Week11	March 19- March 26	**Multicategory Logit Models** Logit Models for Nominal Responses Baseline-Category Logits Estimating Response Probabilities Logit Models for Ordinal Responses Cumulative Logit Models for Ordinal Responses Cumulative Logit Models with Proportional Odds Property Test of proportionality of odds ratios	Midterm 2	LO 3
Week12	Due Thursday April 4	**HW2 lect 4a-b, 5a-b, 6a-b**		LOs 2, 7, 8
Week12- week 13	April 2 – April 9	**Poisson Regression** Poisson Random Variable and Poisson Distribution. Properties of a Poisson Distribution. Poisson Regression Interpretation and Testing. Overdispersion and Testing for Overdispersion using Negative Binomial Regression. Regression Methods for Count Data with Overdispersion:	Midterm 2	LO 4

		1) Negative Binomial Regression. 2) Poisson Regression with Robust Standard Error. Poisson Regression for Rate Data. Poisson Regression for Small Sample Size.		
Week14-Week15	April 16 – April 23	**Generalized Linear Models (GLM)** Introduction to GLM. Components of GLM. Random Component, Systematic Component, Link Function. Generalized Linear Models for Binary Outcome with: a) Continuous Explanatory Variable: Linear Probability Model Logistic Regression Model Poisson Regression Model Generalized Linear Models for Binary Outcome with b) Categorical Explanatory Variable: Linear Probability Model Logistic Regression Model Poisson Regression Model Generalized Linear Models for Count Outcome: Poisson Regression Model. Model Inference in GLMs	Midterm 2	LO 5
Week16	April 30	**Logistic Regression for Matched Case-Control Studies** Definition of matching Data structure for matched case-control Logistic Regression for matchedcase-control study.	Midterm 2	LO2
Weeks of May 7-May 18	Date will be decided on by the Registrar's Office	**Midterm 2: Lectures 4 to 9** (till matched case-control logistic regression inclusive)		LOs 2-6

Appendix I. Reinforced – Introduced CEPH competencies

Introduced competencies: the competency is introduced at a basic level. Instruction and learning activities focus on basic knowledge, skills and entry-level complexity. The competency is **not assessed.**

Reinforced competency: The competency is reinforced with feedback; students demonstrate the outcome at an increasing level of proficiency (above the introductory stage). Instruction and learning activities concentrate on enhancing and strengthening existing knowledge and skills, as well as expanding complexity. The competency is **not assessed.**

Kindly put an X where applicable. Please remove the tables/rows that are not applicable to your course

Core Competencies	Introduced	Reinforced
CC3. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate		X
CC4. Interpret results of data analysis for public health research, policy or practice		X

EPHD competencies	Introduced	Reinforced
EBCC4: Demonstrate ability to write software codes in order to manage and analyze health data through the use of multiple statistical software		**X**
EBCC5: Apply inferential statistics and advanced statistical approaches such as regression modelling to analyze complex health related data		**X**
EBCC6: Interpret and communicate statistical findings in oral and written format		**X**
EBCC9: Analyze Health related data using advanced statistical techniques and software packages		**X**