

האוניברסיטה העברית בירושלים

סילבוס

סדנה בשיטות במדעי החברה: סטטיסטיקה ב - R - 56988

תאריך עדכון אחרון 08-10-2024

נקודות זכות באוניברסיטה העברית: 2

תואר: מוסמך

היחידה האקדמית שאחראית על הקורס: מדע המדינה

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר א'

שפת ההוראה: אנגלית

קמפוס: הר הצופים

מורה אחראי על הקורס (רכז): ד"ר מת'יו סיימונסון

דוא"ל של המורה האחראי על הקורס: matthew.simonson@mail.huji.ac.il

שעות קבלה של רכז הקורס:

מורי הקורס:

תאור כללי של הקורס:

NOTE: THIS COURSE HAS BEEN POSTPONED TILL THE SPRING.

This is a weeklong methodology bootcamp aimed at social science graduate students who have completed an introductory methods course and wish to extend their statistical capabilities. Third-year BA students who have taken a quantitative methods course are also welcome. Each day consists of three lectures and two practice sessions. The curriculum centers on two key topics that are widespread in contemporary political science research: generalized linear models (e.g., logit, probit, conditional logit, negative binomial) and multilevel modeling (e.g., fixed effects, random effects, mixed effects). Students will learn to clean data, run regressions, and create customized plots using the R programming language (no experience needed). Other topics include probability notation, matrix algebra, weighted least squares, bootstrapping, cross-validation, and regularization.

מטרות הקורס:

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

Upon successful completion of this course, students should be able to

1. *Understand the mathematical concepts and notation that regression is based on*
2. *Manipulate data, generate customized plots, and write functions in R*
3. *Choose the appropriate type of regression for the data they seek to model*
4. *Run regressions for a wide variety of generalized linear models and multilevel models in R*
5. *Assess model fit and use machine learning techniques to avoid overfitting*
6. *Interpret their results with the aid of model prediction, simulation, and resampling*

דרישות נוכחות (%) :

100

שיטת ההוראה בקורס: In-person, lecture and discussion

רשימת נושאים / תכנית הלימודים בקורס:

DAY 1: FOUNDATIONS

1. Introduction to R

-
2. Math Review: Exponents, Logarithms, Limits, Summation, Derivatives, Integrals, Taylor Series
3. Probability: Random Variables, Condition Probability, Moments

DAY 2: LINEAR REGRESSION AND LINEAR ALGEBRA

4. Linear Algebra: Vector and matrix multiplication, variance-covariance matrix, Eigenvalues
5. Simple Regression (Review)
6. Multiple Regression: Matrix representation of regression, multicollinearity, and generalized least squares

DAY 3: GENERALIZED LINEAR MODELS

7. Binary Models: Logit and Probit, Maximum Likelihood Estimation
8. Interpretation and Plotting
9. Categorical, Ordered, and Count Models

DAY 4: MULTILEVEL MODELS

10. Fixed Effects and Random Effects
11. Mixed Effects and Correlated Random Effects
12. Advanced Multilevel Modeling (time permitting)

DAY 5: RESAMPLING AND REGULARIZATION

13. Bootstrapping
14. Model Selection (time permitting)
15. Regularization and Dimensionality Reduction (time permitting): Ridge, LASSO, PCA, Factor Analysis

חומר חובה לקריאה:
None

חומר לקריאה נוספת:

1. R for Political Data Science: A Practical Guide (Urdinez and Cruz, 2021)
2. R for Data Science (2e) (Wickham, Çetinkaya-Rundel, and Grolemund 2023):
<https://r4ds.hadley.nz/>
3. Practical Data Analysis for Political Scientists (Kenkel 2023):
<https://bkenkel.com/pdaps/>
4. המדריך העברי למשתמש ב-R (שריד 2024): <https://il.co.hebrewr.www/>

מרכיבי הציון הסופי:

הגשת עבודה מסכמת / פרויקט גמר / מטלת סיכום / מבחן בית / רפרט 60 %
מטלות הגשה במהלך הסמסטר: תרגילים / עבודות / מבדקים / דוחות / פורום / סימולציה ואחרות
40 %

מידע נוסף / הערות:

To thrive in this course, you need to have taken a class that covers introductory statistics. Prior to the workshop, I expect you to review your notes, textbooks, or online materials to refresh your knowledge. In particular, I need you to be at least somewhat comfortable with the following topics:

- One-variable statistics: mean, variance, normal distribution, standard deviation, outliers
- Two-variable statistics: correlation coefficient
- Sampling: population versus sample, random sample, bias and non-representative samples
- Estimation: sample mean, standard error, and confidence interval
- Hypothesis testing: null hypothesis, alternative hypothesis, p-value, t-test
- Linear regression: dependent variable (outcome) versus independent variables (regressors), best-fit line and sum of least squares, intercept and coefficients, error term and residuals, simple versus multiple regression, binary (dummy) variables, interactions between variables

You also need to install R and RStudio (for free from <https://posit.co/download/rstudio-desktop/>) before class and make sure RStudio opens on your laptop.