
האוניברסיטה העברית בירושלים

סילבוס

נושאים מתקדמים בניתוח נתונים ברפואה ובריאות הציבור -
98449

תאריך עדכון אחרון 14-02-2019

נקודות זכות באוניברסיטה העברית: 2

תואר: מוסמך

היחידה האקדמית שאחראית על הקורס: בריאות הציבור ורפואה קהילתית

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר ב'

שפת ההוראה: עברית

קמפוס: עין כרם

מורה אחראי על הקורס (רכז): ד"ר שי כרמי

דוא"ל של המורה האחראי על הקורס: shai.carmi@huji.ac.il

שעות קבלה של רכז הקורס: בתאום מראש באימייל

מורי הקורס:
ד"ר שי כרמי
גב שמאם לוי

תאור כללי של הקורס:

הקורס יכסה נושאים מודרניים בסטטיסטיקה הרלוונטיים למחקרים ביו-רפואיים ואפידמיולוגיים. החומר יכלול נושאים מתקדמים בבדיקת השערות, אמידה, וסיווג. בפרט, יבחנו מצבים בהם קיים מידע רב-ממדי ובהיקף גדול ("data big"), כמו גם מצבים בהם מספר התצפיות קטן יחסית. הקורס יעמוד על ההזדמנויות, המחקריות והיישומיות, שגלומות בנתונים מסוגים שונים, ועל היתרונות והחסרונות של השיטות השונות. תרגול החומר יהיה ברובו ע"י תרגילי תכנות בשפת R.

מטרות הקורס:

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

בוגרי הקורס יוכלו:

- * להפעיל שיקול דעת בבחירת מבחן סטטיסטי תקף ובעל עוצמה מירבית עבור הנתונים העומדים בפניהם ולנתח את הנתונים והתוצאות.
- * לזהות את בעיית ההשערות המרובות ולשלוט על קצב התגליות השגויות.
- * להבין תחת אילו תנאים נחוצות שיטות דגימה (תמורות, bootstrap) לבדיקת השערות ואמידה ולדעת להפעיל את השיטות.
- * להבין את מטרות האלגוריתמים ללמידת מכונה ואת הבסיס התיאורטי ללמידה.
- * להפעיל שיטות נפוצות לסיווג נתונים בהיקף גדול ואת יתרונותיהן וחסרונותיהן.
- * להפעיל שיטות נפוצות למציאת צבירים והורדת מימד.
- * ליישם את השיטות שנלמדו בשפת התכנות R.

דרישות נוכחות (%):

שיטת ההוראה בקורס: הרצאות

רשימת נושאים / תכנית הלימודים בקורס:

- 1) מבוא להסקה סטטיסטית: בדיקת השערות ואמידה, התפלגות המדגם, רווח הסמך, שגיאות ועוצמה, הטיה ושונות.
- 2) עמידות של מבחנים סטטיסטיים: ההנחות מאחורי מבחנים נפוצים וכיצד בודקים האם מתקיימות.
- 3) מה עושים כאשר ההנחות אינן מתקיימות: טרנספורמציות, מבחנים אי-פרמטריים (מבחן הסימן,

- מבחן וילקוקוסון, רווח סמך לחציון).
- (4 שיטות המבוססות על נראות: אומד נראות מירבית, מבחן יחס הנראות, תכונות.
- (5 בדיקת השערות מרובות: הצגת הבעיה, תיקון בונפרוני ותכונותיו, שיעור התגליות השגויות ושיטת בנימיני-הוכברג להגבלתו.
- (6 שיטות דגימה מחדש לבדיקת השערות ואמידה: הגרלת מספרים אקראיים מתוך התפלגות, שיטת תמורות מבחני bootstrap.
- (7 מבוא ללמידת מכונה: מטרות כלליות ודוגמאות, סוגי אלגוריתמים, הטיה ושונות, גורמים ל-cross-validation, התופעה וחשיבות overfitting.
- (8 סיווג: שיטת kNN, רגרסיה לוגיסטית, הפרספטרון ורשתות נוירונים. (אם יהיה זמן: עצי החלטה), יתרונות וחסרונות של השיטות הנפוצות, מדדי דיוק.
- (9 למידה לא מפוקחת: מציאת צבירים: k-means ו-UPGMA. שיטת PCA להורדת מימד.
- (10 תכנות בסיסי בשפת R עם יישומים לנושאים שלנמדו.

חומר חובה לקריאה:
אין

חומר לקריאה נוספת:

- An Introduction to Statistical Learning: with Applications in R*, Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Springer, 2013.
- Learning From Data*, Yaser S. Abu-Mostafa, Malik Magdon-Ismael, and Hsuan-Tien Lin. AMLBook, 2012.
- Pattern Recognition and Machine Learning*, Christopher M. Bishop, Springer, 2007.
- All of Statistics: A Concise Course in Statistical Inference*, Larry Wasserman, Springer, 2004.

הערכת הקורס - הרכב הציון הסופי :

- מבחן מסכם בכתב/בחינה בעל פה 70 %
הרצאה 0 %
השתתפות 0 %
הגשת עבודה 0 %
הגשת תרגילים 30 %
הגשת דו"חות 0 %
פרויקט מחקר 0 %
בחנים 0 %
אחר 0 %

מידע נוסף / הערות:

- הקורס מתבסס על ידע בקורס סטטיסטיקה בסיסי, כפי שניתן לתלמידי רפואה/מדעים ביו רפואיים/בריאות הציבור. יש לחזור על החומר של הקורס הבסיסי לפני תחילת קורס זה.

הקורס ישים דגש על הצד האינטואיטיבי והמעשי של השיטות, ללא ירידה לפרטים המתמטיים מעבר

לנדרש.

ידע בשפת תכנות אחת לפחות היא דרישת חובה לקורס. השפה המומלצת לקורס היא R, אך ניתן יהיה להשתמש בכלים דומים כגון Python או Matlab.

תרגילי הבית בקורס יהיו בעיקר (אך לא רק) תרגילי תכנות. התרגילים עשויים להיות מאתגרים ולדרוש השקעה מרובה מסטודנטים ללא נסיון בתכנות.