



סילבוס

## עיבוד נתונים בסטטיסטיקה מודרנית - 52311

תאריך עדכון אחרון 24-01-2019

נקודות זכות באוניברסיטה העברית: 4

היחידה האקדמית שאחראית על הקורס: סטטיסטיקה

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר ב'

שפת ההוראה: עברית

קמפוס: קרית א"י ספרא

מורה אחראי על הקורס (רכז): אור צוק

דוא"ל של המורה האחראי על הקורס: [or.zuk@mail.huji.ac.il](mailto:or.zuk@mail.huji.ac.il)

שעות קבלה של רכז הקורס: בתאום מראש

מורי הקורס:

ד"ר אור צוק

מר עומר רונן

תאור כללי של הקורס:

---

הקורס יעסוק בשיטות ניתוח נתונים המשתמשות במחשב בסטטיסטיקה, וייתרכז בניתוח נתונים בהיקף גדול וממימד גבוה. נדון באתגרים חישוביים וסטטיסטיים העולים מניתוח נתונים אלו, בדגש על שיטות פרקטיות ויעילות חישובית. במהלך הקורס נלמד ונממש שיטות סטטיסטיות שונות, ונשתמש בהן לניתוח נתונים הן נתונים מסומלצים והן נתונים אמיתיים מתחומים שונים.

#### מטרות הקורס:

ממטרת הקורס היא להציג לסטודנטים שיטות מודרניות בסטטיסטיקה

#### תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

להבין מספר שיטות סטטיסטיות מודרניות, לממש אותם בשפת תכנות סטנדרטית באופן יעיל, ולהפעיל אותן על נתונים אמפיריים על מנת לפתור בעיה מדעית קונקרטית.

#### דרישות נוכחות (%) :

אין

שיטת ההוראה בקורס: הרצאה ותרגול.

#### רשימת נושאים / תכנית הלימודים בקורס:

רשימה התכנים יכולה להשתנות:

0. עיבוד מקדים של נתונים: נורמליזציה, השלמת מידע חסר, ויזואליזציה
1. בדיקת השערות: חישובי עוצמה, חישוב  $p$ -value באמצעות סימולציות, שיטות לבדיקת השערות מרובות
2. שיטות רגרסיה: רגרסיה לינארית מרובת משתנים, שיטות לבחירת משתנים וייצוג דליל
3. שיטות קלסיפיקציה: רגרסיה לוגיסטית, CART, רשתות נוירונים
6. שיטות לבחירת מודל ומיצוע על מודלים שונים
5. שיטות לינאריות ולא לינאריות להורדת מימד:  
PCA, manifold learning
6. אלגוריתמים למציאת צבירים

#### חומר חובה לקריאה:

אין

---

חומר לקריאה נוספת:

*The Elements of Statistical Learning* □ *Data mining, inference and prediction*  
(Tibshirani, Hastie and Friedman)  
<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

*Large Scale Inference*, Bradley Efron  
[http://statweb.stanford.edu/~ckirby/brad/LSI/monograph\\_CUP.pdf](http://statweb.stanford.edu/~ckirby/brad/LSI/monograph_CUP.pdf)

*Advanced Data Analysis from an Elementary Point of View*, Cosma Rohilla Shalizi  
<http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/>

הערכת הקורס - הרכב הציון הסופי :  
מבחן מסכם בכתב/בחינה בעל פה 0 %  
הרצאה 0 %  
השתתפות 0 %  
הגשת עבודה 40 %  
הגשת תרגילים 60 %  
הגשת דו"חות 0 %  
פרויקט מחקר 0 %  
בחנים 0 %  
אחר 0 %

מידע נוסף / הערות:  
(יעודכן בהמשך)