**ניתוח נתונים**

**הרצאה**: 3 שעות

**תרגיל** : 1 שעה

**דרישות קדם:** מבוא להסתברות, שיטות מתמטיות 1.

**מטרת/תיאור הקורס:** קורס במבוא לסטטיסטיקה וניתוח נתונים. הקורס מכסה מושגי יסוד בסטטיסטיקה (כגון מובהקות, מבחני השערות, אמידה סטטיסטית), שיטות דגימה, סוגי משתנים, מידול (כגון רגרסיה, ניתוח שונות), מבחנים א-פרמטריים. הקורס מתבסס על קורס בסיסי בהסתברות.

במסגרת הקורס נשתמש ב-R ונלמד את מימוש הפקודות הרלוונטיות באמצעות R. כמו כן, נלמד ונתרגל את כלי הויז'ואליזציה של R (חבילת ggplot2), וכלים שימושיים להכנת נתונים לקראת ניתוח (כדוגמת חבילות dplyr, tidyr).

במהלך הקורס יעשה שימוש ב-Data sets "פתוחים לציבור" (לדוגמה מאתר התחרויות Kaggle, ומ-tidytuesday).

**ציון:** תרגילי בית בזוגות (20%), פרויקט מסכם (30%), ומבחן סופי (50%).

**תוכן הקורס:**

1. הקדמה: כיצד מתבצע מחקר סטטיסטי? שלבי העבודה מהתכנון, דרך ביצוע, והצגת הנתונים למקבלי החלטות.
2. אמידה סטטיסטית ושערוך פרמטרים: אמד לתוחלת, סטיית תקן, אחוזונים.
3. בדיקת השערות, חלק א': רווח בר סמך לתוחלת, מבחן לא מזווג, מבחן מזווג. מבחן student’s t, מבחן z, מבחנים א-פרמטריים.
4. בדיקת השערות, חלק ב': מבחן טיב התאמה, (Chi Square), ומבחן התפלגויות Kolmogorov Smirnov).
5. ניתוח שונות חד-כיווני
6. ניתוח שונות רב-כיווני
7. תכנון ניסויים, השוואות מתוכננות, חישוב גודל מדגם.
8. רגרסיה לינארית מרובת משתנים.
9. מתאם (קורלציה).
10. רגרסיה לוגיסטית.

**ספרות:**

Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). *OpenIntro statistics* (pp. 174-175). OpenIntro.

(זמין להורדה בקישור: <https://leanpub.com/openintro-statistics/>)

Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".

(זמין לקריאה בקישור <https://r4ds.had.co.nz/>)

ספרים נוספים:

Hines W. & D. Montgomery: Probability and Statistics in Engineering and Management Sciences. Wiley, 2008

Johnson, N.L. & Leone, F.C.: Statistics and Experimental Design Vol. 1.2, Wiley, 2nd ed., 1997.

Draper N. & H. Smith: Applied Regression Analysis, 3rd ed. Wiley, 1998.

Gibbons J.D.: Nonparametic Statistical Inference, Springer, 2011.