

ניתוח נתונים

הרצאה: 3 שעות

תרגיל: 1 שעה

דרישות קדם: מבוא להסתברות, שיטות מתמטיות 1 (חדו"א)

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). הקורס מתבסס על קורס בסיסי בהסתברות. במסגרת הקורס נשתמש ב-R ונלמד את מימוש הפקודות הרלוונטיות באמצעות R. כמו כן, נלמד ונתרגל את כלי היוזואליזציה 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-tidyuesday).

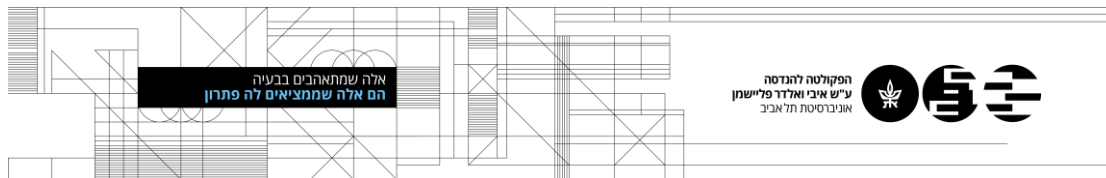
ציון: מבחן 70%, פרויקט מסכם 30%

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.
4. בדיקת השערות, חלק ב': מבחן טיב התאמה, (Chi Square)
5. רגרסיה לינארית מרובת משתנים.
6. מתאם (קורלציה).
7. ניתוח שונות חד-כיווני
8. ניתוח שונות רב-כיווני
9. תכנון ניסויים, השוואות מתוכננות, חישוב גודל מדגם.

ספרות:

OpenIntro statistics is an introduction to statistics with R, it doesn't contain everything we will learn, but provides a good intro to some topics. Downloadable for free [here](<https://leanpub.com/openintro-statistics/>) (click on the "download sample" and the entire book downloads as a pdf file).



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

R4DS (R for Data Science) is a highly recommended book for learning R, and specifically *tidyverse* which is a collection of useful packages for data science. The book is mostly "technical", i.e., it does not provide much theoretical details. This book is also available in an online format [here](<https://r4ds.had.co.nz/>).

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

Most of the theory I present during the course comes from these two books:

* Walpole R.E., Myers R. H, Myers S. L., and Ye K.: Probability & Statistics for Engineers & Scientists. Prentice Hall, 9th ed., 2011. Available [online](https://fac.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineers_and_scientist.pdf)

* Runger G. & D. Montgomery: Applied Statistics and Probability for Engineers. Wiley, 7th ed., 2018. An old edition is available [online](<http://www.um.edu.ar/math/montgomery.pdf>)

Additional books:

* 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.: Nonparametric Statistical Inference, Springer, 2011.

Additional Sources

You can find various online videos teaching statistics theory along with R coding examples. One such place is the Statistics of DOOM channel on youtube:

<https://www.youtube.com/channel/UCMdihazndR0f9XBoSXWqnYg>.