**תרגיל להגשה - סטטיסטיקה תיאורית בעזרת R**

מטרת התרגיל היא יישום הדברים שנלמדו בשלושת השבועות הראשונים על קובץ נתונים אמיתי. מצאו באמצעות האינטרנט או כל מקור אחר קובץ נתונים לפרויקט. קובץ ראוי הינו קובץ הכולל לפחות כמה עשרות תצפיות על שני משתנים כמותיים (משתנים רציפים) אשר יש טעם לבדוק ביניהם קשר. הקובץ ישמש גם לתרגיל ההגשה השני ולכן יש להמנע מבחירת קובץ בו המשתנה המסביר הוא זמן קלנדרי.

שפע נתונים נמצאים ברשת. תחפשו ב [google](http://www.google.com/) תחת מילות המפתח interesting data sets  או ביטויים דומים ותגיעו לעשרות אתרים עם נתונים מתחומי דעת שונים. באתר של kaggle יש קבצים רבים זמינים. חיפוש ומציאת קובץ מתאים הם חלק חשוב של התרגיל ויש להקדיש לו זמן ומחשבה. במהלך התרגיל תצטרכו להסביר את התוצאות של הרגרסיה ולכן חשוב שהן יענו על שאלת מחקר שמעניינת אתכם.

1. ציינו מהו מקור הנתונים, כולל קישור רלבנטי. ציינו בקצרה את המשמעות של משתנה X ואת המשמעות של משתנה Y כולל יחידות המדידה של המשתנים.
2. תארו את הנתונים בקובץ שבחרתם באמצעות הכלים שלמדנו תוך שימוש ב R. התיאור יהיה הן גרפי (היסטוגרמות, דיאגרמת פיזור דו-משתנית) והן בעזרת מדדי מיקום ופיזור (ממוצע, חציון, אחוזונים, סטית תקן). יש לבאר את המדדים שחושבו ולציין את יחידות המדידה של המדדים. חשוב לבחור את מדדי המיקום הרלוונטים לבעיה (ממוצע, אחוזונים, מינימום ומקסימום – תלוי בבעיה הספציפית).
3. חקרו את הקשר הלינארי בין המשתנים תוך שימוש ב R. חשבו את מקדם המתאם, מיצאו את קו הרגרסיה, פרשו את הפרמטרים וכתבו את מסקנותיכם בדבר טיב הקשר שמצאתם, תוך שימוש במדדי טיב קשר רלוונטים והתייחסות לתצפיות חריגות אם יש כאלו.
4. בחרו 5 תצפיות עבורן תעשו חיזוי ורשמו את ערך המשתנה המסביר, הערך החזוי על פי קו הרגרסיה, הערך האמיתי ואת השארית.

**העבודה צריכה להיות קצרה ומתומצתת – שני עמודים המסכמים את הנתונים וניתוחם כפי שמוסבר למעלה. העבודה תוגש במודל כקובץ PDF. חשוב שבדוח אותו אתם כותבים תתייחסו לנתונים עצמם ולמשמעות הניתוח ולא תדברו במושגים של משתנה מסביר ומשתנה מוסבר או במושגים של X ו Y. תקפידו לכתוב את הדוח על הבעיה הספציפית אותה אתם חוקרים.**

התרגיל מהווה 10% מהציון הסופי (במידה והציון במבחן הסופי הוא לפחות 50).

יש להגיש את העבודה במודל עד לתאריך 3.2.2024.

**ניתן ורצוי לבצע ולהגיש את העבודה בזוגות.**