

### תיאור המשימה:

המעבדה עוסקת בקשר בין דירוג חברתי כלכלי של יישובים לבין תוצאות הבחירות בנוסף לקובץ תוצאות הבחירות יש להוריד ולהשתמש במשימה זו בקובץ דירוג חברתי כלכלי לפי יישובים: HevratCalcaliYeshuvim.csv או HevratCalcaliYeshuvim.txt

1. הצליבו את נתוני הבחירות על פי **יישובים** עם הנתונים בקובץ הדירוג החברתי הכלכלי. מצאו והציגו את רשימת היישובים שמופיעים בשני הקבצים. כמה יישובים קיבלתם? בשאלה זו ובשאלות הבאות השתמשו רק ביישובים אלו לניתוח הנתונים. הציגו את רשימת היישובים שנותרו - כמה יישובים קיבלתם? חשבו את תוצאות הבחירות (שכיחות הקולות לכל מפלגה עבור 10 המפלגות הגדולות) רק ביישובים אלו והציגו אותן מול התוצאות הכלל ארציות בבר עמודות כפול. תארו האם יש הבדלים משמעותיים בין התוצאות הארציות לתוצאות ביישובים שקיבלתם.  
(**הערה:** אין חובה להוסיף ידנית יישובים המופיעים באיות שונה וכן יישובים המופיעים במועצות אזוריות בקובץ הדירוג החברתי הכלכלי.)
2. הציגו את תוצאות הבחירות שהיו מתקבלות בכל אחד מעשרת הדירוגים החברתיים כלכליים מ-1 עד 10. כלומר, יש להראות גרף עם 10 subplots מסודרים בתבנית של 2X5 כאשר בכל subplot יש גרף עמודות המתאר את שכיחות הקולות של 10 המפלגות הגדולות ביותר רק ביישובים עם הדירוג חברתי כלכלי המתאים.
3. עבור כל מפלגה, צרו גרף עמודות בו מתואר שכיחות הקולות של המפלגה ביישובים ב-10 האשכולות, מסודרים. כלומר, יש לעשות גרף עם 10 subplots מסודרים בתבנית של 2X5 כאשר בכל subplot יש את שכיחות הקולות של מפלגה מסוימת ב-10 האשכולות מסודרים לפי הסדר.  
(**הערה:** בשאלה זו אנו מציגים למעשה את אותה אינפורמציה אשר בשאלה 2 אבל בדרך אחרת).
4. חשבו בעזרת **קובץ הקלפיות** עבור כל עיר מדד להטרוגניות ההצבעה באופן הבא:  
א. נגדיר את  $p_i$  להיות וקטור שכיחויות ההצבעה **לכל המפלגות** ביישוב מסויים  $i$ , ואת  $p_i^b$  להיות וקטור השכיחויות על פני כל המפלגות בקלפי ה- $b$  **בישוב זה**. (שימו לב שאלו סימונים מעט שונים ממה שהשתמשנו עד כה).  
ב. יהי  $d_i^b$  המרחק הריבועי בין וקטור שכיחויות ההצבעה בקלפי  $p_i^b$  לבין השכיחות בכל העיר  $p_i$  (על פי ההגדרה שבשאלה 3 במעבדה 1):  
$$d_i^b = (p_{i1} - p_{i1}^b)^2 + \dots + (p_{ik} - p_{ik}^b)^2$$
  
ג. נגדיר את ההטרוגניות בעיר  $i$  להיות המרחק הריבועי הממוצע של שכיחויות הקלפיות מכלל השכיחות בעיר:  
$$h_i = (\sum_b d_i^b) / B_i$$
  
כאשר  $B_i$  הוא מספר הקלפיות הכולל בעיר  $i$ .  
(כלומר ההטרוגניות מתקבלת ע"י חישוב שונות שכיחות ההצבעה ביישוב של כל מפלגה וסכימת השונות על פני כל המפלגות)

כעת הציגו scatter-plot של ההטרוגניות ההצבעה בערים  $h_i$  מול מדד ג'יני  $g_i$  עבור הערים הניתן בקובץ הדירוג החברתי כלכלי (זהו מדד המודד אי שיוון כלכלי), כאשר גודל כל עיגול פרופורציוני למספר הקולות ביישוב. חשבו גם את קורלצית spearman בין ההטרוגניות ההצבעה למדד ג'יני.  
מה מסקנתכם? האם יש קשר סיגניפיקנטי בין אי השיוון הכלכלי להטרוגניות ההצבעה?

#### הערות:

- חשבו על עיצוב הגרפים. תנו כותרת לצירים, שימו לב לאורך הצירים.
- השתמשו בצבעים, עובי נקודה, וכו' כדי להדגיש נקודות חשובות.
- מותר לכם להיות יצירתיים; נסו לכלול יותר ישובים בניתוח המשולב של הבחירות והנתונים הסוציאקונומיים. בשאלות 2,3 אפשר גם לחשב ולהציג על ה-bars עבור כל מפלגה את סטיית התקן של שכיחות ההצבעה למפלגה על פני ישובים באותו אשכול