

תיאור המשימה:

המעבדה עוסקת בתיקונים לאחוז הצבעה וכן בקשר בין דירוג חברתי כלכלי של יישובים לבין תוצאות הבחירות.

1. כתבו פונקציה שעושה תיקון לתוצאות הבחירות ואמידת השכיחויות q_j באוכלוסיה של כל מפלגה על ידי פתרון בעיית הרגרסיה הלינארית הבאה על פי שיטת הריבועים הפחותים. תוכלו להשתמש בפונקציה OLS מתוך מודול הפיתון statsmodel (זו פונקציה מקבילה מהרבה בחינות לפונקציה lm ב-R).
- ראשית, אומדים את ערכי u_j^{-1} ע"י מיזעור הביטוי: (כאן i סוכם על הקלפיות ו- j על המפלגות)
$$\sum_i (\sum_j n_{ij} u_j^{-1} - \hat{\eta}_i)^2$$

- שנית, מחשבים את האומדים ל- $\hat{\eta}_{ij}$ ע"י הכפלת n_{ij} באומדים ל- u_j^{-1} ומכאן מחשבים את האומדים \hat{q}_j לשכיחויות ההצבעה המתוקנות כפי שהודגם בכיתה.
2. כעת חזרו על הסימולציות בשאלה 2 ממעבדה 3. חשבו את האומדנים \hat{q}_j לכל אחת מ-3 הדרכים לבחירת v_{ij} על פי התיקון בשאלה 1 (כלומר בדומה לשאלה 2.2 במעבדה הקודמת אך בעזרת התיקון החדש). עבור כל אחת מ-3 דרכי הסימולציה ועבור כל מפלגה הוסיפו ל-bar-plot משאלה 2.4 במעבדה הקודמת עמודה חדשה עבור התיקון החדש (מוצג +/- סטיית תקן) והציגו את ה-bar-plots החדשים (כלומר לכל מפלגה יהיו 4 עמודות). מה מסקנותיכם? באיזה תיקון כדאי להשתמש ומתי?
3. קראו את הנתונים מקובץ הדירוג החברתי הכלכלי eshkol_hevrati_calcali.csv או eshkol_hevrati_calcali.xlsx והצליבו אותם עם נתוני הבחירות על פי יישובים ליצירת data-frame משותף. השתמשו בקוד הישוב (set_code בקובץ זה) כדי להתאים בין הקבצים. (ניתן להשתמש בפקודת join של pandas). הציגו את רשימת היישובים שמופיעים בשני הקבצים. כמה יישובים קיבלתם? בשאלה זו ובבאה השתמשו רק ביישובים אלו לניתוח הנתונים. חשבו את תוצאות הבחירות (שכיחות הקולות q_j לכל מפלגה מתוך כלל הקולות הכשרים עבור 8 המפלגות הגדולות) רק ביישובים אלו והציגו אותן מול התוצאות הכלל ארציות בבר עמודות כפול. האם יש הבדלים משמעותיים בין התוצאות הארציות לתוצאות ביישובים שקיבלתם?
4. א. 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 8 המפלגות הגדולות ביותר רק ביישובים עם הדירוג חברתי כלכלי המתאים.
ב. עבור כל מפלגה, צרו גרף עמודות בו מתואר שכיחות הקולות של המפלגה ביישובים ב-10 האשכולות, מסודרים. כלומר, יש לעשות גרף עם 8 subplots מסודרים בתבנית של 2X8 כאשר בכל subplot יש את שכיחות הקולות של מפלגה מסוימת ב-10 האשכולות מסודרים לפי הסדר (הערה: בסעיף זה אנו מציגים למעשה את אותה אינפורמציה אשר בסעיף א. אבל בדרך אחרת).

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 יישובים בניתוח המשולב של הבחירות והנתונים הסוציאקונומיים. בשאלה 4 אפשר גם לחשב ולהציג על ה-bars עבור כל מפלגה את הממוצע +/- סטיית התקן של שכיחות ההצבעה למפלגה על פני יישובים באותו אשכול.