

תיאור המשימה:

המעבדה עוסקת באמידת וחיזוי שינויים בין מערכות הבחירות. יש להשתמש בקבצי תוצאות הבחירות על פי קלפיות בבחירות מועד א ומועד ב ב-2019, וכן בקובץ דירוג חברתי כלכלי לפי ישובים:
HevratCalcaliYeshuvim.csv או HevratCalcaliYeshuvim.txt

1. חשבו את סטיית תקן ו-p-value לאברי המטריצה M שנאמדו בשיטת רגרסיה לינארית עם רבועים פחותים בשאלה 1 סעיף א. במעבדה 7, עבור השערת האפס $m_{jk}=0$, ומבחן חד צדדי עם אלטרנטיבה חיובית עבור כל איבר. דווחו על כל ערכי ה-M הסיגניפיקנטיים ברמת מובהקות של 0.001.
2. חשבו סטיית תקן בשיטת ה-Bootstrap לכל איבר במטריצה M שנאמדה בשיטת הריבועים הפחותים בשאלה 1 סעיף א. במעבדה 7. השתמשו ב-100 איטרציות bootstrap. כעת, השתמשו בקירוב הנורמלי עבור ההתפלגות של כל איבר במטריצה הנאמדת M כאשר הניחו שכל איבר הוא אומד בלתי מוטה של ערך M האמיתי. על סמך קירוב זה חשבו p-value עבור בדיקת ההשערות משאלה 1 ודווחו על כל ערכי ה-M הסיגניפיקנטיים ברמת מובהקות של 0.001. השוו את תוצאותיכם לתוצאות שאלה 1.
3. חזרו על שאלה 2 אך הפעם עבור אומדי M עם אילוצי החיוביות וכן הנרמול לאחד בכל שורה משאלה 3 במעבדה 7 - כלומר אמידה באמצעות nlms, איפוס ערכים קטנים מ-0.005 ונרמול כל שורה ל-1. כמו כן, עבור המפלגות שלא השתנו בדקו את השערת האפס: $m_{jj}=1$ והאלטרנטיבה $m_{jj}<1$. הסבירו למה לא ניתן להשתמש בשיטה של שאלה 1 עבור שאלה זו, ומה היתרון של שיטת ה-Bootstrap כאן.
4. **הערה:** כשהאומד עבור פרמטר מסוים זהה עבור כל מדגמי ה-bootstrap, סטיית התקן הנאמדת על פי ה-bootstrap היא 0. במקרה זה כדי להמנע מבעיות נומריות נגדיר את ה-p-value להיות 0.5.
4. חלקו את הקלפיות המשותפות לשתי מערכות הבחירות פעם אחת באופן אקראי לשתי קבוצות זרות: train ו-test בחלוקה של 80% מהקלפיות ב-train ו-20% ב-test. אמדו את שלושת המודלים משאלות 1 א, 2 ו-3 במעבדה 7 על נתוני ה-train וחשבו את השגיאה הריבועית הממוצעת בחיזוי תוצאות בחירות ב על פני כל קלפיות ה-test עבור כל מודל. את השגיאה הריבועית הממוצעת יש לחשב רק עבור התצפיות המתאימות ל-10 המפלגות הגדולות בכל קלפי ב-test במערכת הבחירות של ספטמבר - כלומר אין להשתמש בתצפיות המתאימות ל"לא הצביעו" - גם אם השתמשם בתצפיות אלו לצורך אמידת המודל. איזו שיטה נותנת שגיאה ריבועית ממוצעת מינימלית?
4. **הערה:** מספר התצפיות ב-test יהיה מספר קלפיות ה-test כפול 10. עבור כל תצפית כזו מקבלים שגיאת אמידה ריבועית ויש למצע שגיאות אלו
5. אמדו מודל נוסף החוזה את תוצאות בחירות ספטמבר באמצעות בחירות אפריל וכן מדד חברתי כלכלי באופן הבא: חלקו את הקלפיות ל-3 קבוצות על פי אשכולות דמוגרפיים של ישוביהן המופיעים בקובץ HevratCalcaliYeshuvim.txt ממעבדה 4. קבוצה אחת תהיה קלפיות השייכות לישובים עם מדד חברתי-כלכלי 1-5, אחרת ישובים עם מדד חברתי כלכלי 6-10, וקבוצה שלישית קלפיות השייכות לישובים עבורם לא מיפנו את הישוב למדד חברתי כלכלי (אלו יהיו בעיקר ישובים קטנים במועצות אזוריות). בכל קבוצה אמדו את M כל קבוצת ה-train של שאלה 4, לפי אומד ה-nlms של שאלה 3 מעבדה 7 כך שתקבלו 3 מטריצות M שונות.
- בשלב החיזוי השתמשו עבור כל קלפי בקבוצת ה-test במודל המתאים (כלומר במטריצה הנאמדת M המתאימה לישוב של קלפי זו). חשבו את השגיאה הריבועית הממוצעת על קלפיות ה-test והשוו אותה לשגיאות בשאלה 4. בנוסף, הציגו את 3 מטריצות ה-M שאמדתם ותארו את ההבדלים הבולטים ביניהן.
6. **בנוסף:** נסו להתאים מודלים נוספים החוזים את תוצאות בחירות ספטמבר מתוך בחירות אפריל ולשפר את השגיאה הרבועית הממוצעת ביחס לשאלות 4,5. מותר לכם להשתמש במידע דמוגרפי/כלכלי/וכו' נוסף על כל ישוב.

הערות:

- חשבו על עיצוב הגרפים. תנו כותרת לצירים, שימו לב לאורך הצירים. השתמשו בפונטים גדולים.
- השתמשו בצבעים, עובי נקודה, וכו' כדי להדגיש נקודות חשובות.