

תרגיל 1

הערות כלליות:

- A. לתרגיל זה יש להשתמש בנתונים שבקובץ CRIMES.dta בתיקיית קבצי הנתונים במודל.
- B. ניתן לראות פירוט על כל משתנה בעזרת הפקודה describe לאחר פתיחת הקובץ.
- C. כדי לפתור את התרגיל יש ליצור קובץ מסוג do (סטטה) או script (R) ולשמור בו את כל הפקודות.
- D. כמו כן יש לשמור את התוצאות שמתקבלות באמצעות קובץ מסוג log (סטטה) או באמצעות הפקודה sink (R) ולהגיש את שני הקבצים כחלק מפתרון התרגיל.
- E. בתיקייה ps1 קיימים גם קבצים מסוג do או script אשר משמשים כתבנית ועשויים לסייע בהכנת התרגיל (שימו לב שיש לשנות את הפקודות בהתאם).

שאלות התרגיל:

1. צרו גרף (על ידי הפקודה scatter בסטטה; או הפקודה plot ב-R) המתאר את הקשר בין מספר השוטרים ("police") על ציר ה-X לבין מספר הפשעים ("crimes") על ציר ה-Y. מה לדעתכם ניתן ללמוד מגרף זה? האם ניתן ללמוד על קשר סיבתי בין המשתנים?
הערה לשמירת הגרף: יש לשמור את הגרף כקובץ png על מנת שניתן יהיה לפתוח אותו בקלות בתוכנות אחרות (למשל ב-word). הגרף אינו נשמר כחלק מקובץ ה-log לכן יש לצרף אותו כקובץ נפרד לפתרון. שמירת הגרף מתבצעת בסטטה: ע"י לחיצה עליו עם כפתור העכבר הימני; ב-R: על-ידי הפקודה png).
 2. צרו משתנה המתאר את מספר הפשעים לנפש (בעזרת הפקודה בסטטה: gen; בעזרת האופרטור ב-R: <-). קראו למשתנה: "crimes_per_capita".
 3. צרו משתנה המתאר את מספר השוטרים. קראו למשתנה: "police_per_capita".
 4. חזרו על הגרף מסעיף 1 עם המשתנים החדשים שיצרתם בסעיפים 2 ו-3. מה ניתן ללמוד על הקשר בין המשתנים בהשוואה לסעיף 1?
 5. חשבו את הממוצע של כל אחד מהמשתנים שהגדרתם בסעיפים 2 ו-3 לכל אחד מהאזורים הבאים: West, Northeast, Midwest and South. (בסטטה: ניתן להיעזר בפקודה sum ובשימוש ב-if עבור התנאים; ב-R: ניתן להיעזר בפקודה mean ובחלוקה לתת-מדגם על-ידי [object,columns,rows]).
- שימו לב: מכיוון שקיימים רק 3 משתנים עבור 3 מהאזורים (כל אחד מהמשתנים מקבל את הערך 1 כאשר התצפית שייכת לאזור שלו ו-0 אם לא), ניתן לחשב את הממוצע עבור האזור הרביעי עם התנאי ששלושת המשתנים של האזורים האחרים שווים ל-0.

6. חשבו את הקורלציה (מקדם המתאם - לפי הפקודה) בין מספר הפשעים לנפש לבין משתנה שיעור האבטלה ("unem"). האם התוצאה מלמדת על קשר סיבתי בין המשתנים?

7. חשבו את הממוצע של מספר הפשעים לנפש לכל שנה עבור ערים שבהן האוכלוסייה קטנה מ- 350 אלף תושבים ("pop").

8. צרו גרף פשוט של מספר הפשעים לנפש ביחס לשיעור האבטלה בשנת 87 בלבד.

9. חשבו ותנו דוגמא למצב בעולם האמיתי בו שני גורמים (או תופעות) הם בעלי קורלציה חזקה, אבל לדעתכם הקורלציה לא מייצגת קשר סיבתי. הסבירו.

10. (שאלה מתוך המבחן הסופי, 2014 מועד א') לחוקר יש קובץ נתונים של גברים בגילאים 24-34 בארה"ב שכולל את המשתנים הבאים:

$married_i = 1$ אם פרט i נשוי, $= 2$ אם פרט i לא נשוי

כתבו את הפקודות כדי ליצור משתנה $marr_i$ אשר שווה ל-1 אם פרט i נשוי ו-0 אם פרט i לא נשוי.
