

## מודלים סטטיסטיים ויישומיהם 52518 תשע"ח – תרגיל 10

להגשה עד 8.1.18 בשעה 23:55

1. מצורף קובץ בשם credit.csv עם נתונים בנוגע ל-1000 הלוואות של לקוחות בנק גרמני. המשתנים המסבירים בקובץ (רשימה חלקית מתוך הנתונים המלאים) הם:

א. good\_credit: משתנה בינארי המציין האם רמת הסיכון של הלקוח טובה (1) או לא (0).

ב. term: משך ההלוואה בחודשים.

ג. amount: סכום ההלוואה המבוקש (במארק גרמני).

ד. age: גיל הלקוח.

ה. land: משתנה בינארי המציין האם הלקוח בעל אדמה (1) או לא (0).

המשתנה המוסבר יהיה good\_credit. בצעו ראשית ניתוח תיאורי של הנתונים והצביעו על מאפייני הנתונים השונים (התפלגות, קשרים אפשריים בין משתנים, תצפיות חריגות וכן הלאה). הריצו מודל רגרסיה לוגיסטית לבדיקת הקשר בין המשתנים המסבירים למשתנה המוסבר. יישמו את הכלים בקובץ logcheck1.pdf לבדיקת טיב התאמת המודל ובמידת הצורך הריצו מודל חלופי. דונו בתוצאות.

הערה: אין צורך למצוא מודל אופטימלי, מעבר למודל הבסיסי התאימו רק מודל חלופי אחד.

2. נניח כי  $Q_m \sim \text{Poi}(e^{\alpha + \theta_m})$  עבור  $m = 1, \dots, M$  כאשר  $\sum_{m=1}^M e^{\theta_m} = 1$  ונסמן  $N = \sum_{m=1}^M Q_m$ . הראו כי ההתפלגות המותנית של הוקטור  $Q = (Q_1, \dots, Q_M)$  בהנתן  $N = n$  היא מולטינומית:

$$(Q|N = n) \sim \text{Multinomial}(n; e^{\theta_1}, \dots, e^{\theta_M})$$

3. הריצו את מודל הרגרסיה הפואסונית שפורסם באתר בקובץ Poisson Regression Example. חשבו אומד עבור  $E[Y|X = x^*]$  ורווח ברמת סמך של 95% עבור הוקטור  $x^*: (Sex = 1, Age = 0.6, Income = 0.9, HScore = 11)$

4. שאלה זו עוסקת במודלים לוג לינאריים

א. כתבו פונקציה ב-R המקבלת כקלט מערך תלת-ממדי של ערכי  $\pi$  ומחשבת את  $\bar{\theta}_{\dots}$  ואת ערכי  $\lambda$  השונים.

ב. מערך תלת-ממדי מוגדר ב-R באופן הבא:

```
g <- array(data, c(A,B,C))
```

כאשר data הוא וקטור המכיל את הנתונים ו-A,B,C הם מספרי הרמות התואמים. עבור נתוני הקובץ

ex10q4.csv צרו מערך תלת-ממדי תואם והריצו את הפונקציה אותה כתבתם על g.

ג. ודאו כי מתקיים  $P(A = i, B = j, C = k) = P(A = i, B = j)P(C = k)$ , כלומר כי זהו המודל  $(AB, C)$ .