רגרסיה ומודלים לינאריים 52307 תשע"ח 2017-18
בוחן 1 26.12.2016

בבוחן שאלות אמריקאיות ושאלות פתוחות.

ליד כל **סעיף** מצויין מספר הנקודות המקסימלי עבור **פתרון מלא** . מספר הנקודות הכולל הוא 110. בכל מקרה, ציון הבוחן הוא 100 לכל היותר.

שימו לב שהשאלות הן בדרגת קושי שונה כך שמומלץ לא להתעכב יתר על המידה על שאלה מסויימת.

אנא הקפידו על ההנחיות הבאות:

- כתבו את ת.ז. **(לא את השם!)** בראש כל עמוד של טופס הבחינה.
- אין לצרף לטופס דפים נוספים.
- אין לתלוש דפים מטופס הבחינה.

לתשומת לבכם לגבי השאלות הפתוחות:

- **תשובה סופית ללא דרך לא תזכה בניקוד כלשהו (ציון 0).**
- בשאלות הפתוחות יש לכתוב את הפתרון רק במקום המוקצה לכך, מעל לכל קו כתבו שורה אחת בלבד בכתב יד קריא. (השאלות נכתבו כך שניתן לכתוב פתרון תמציתי לכל סעיף).
- מגבלת המקום תאכף באופן קפדני. פתרונות אשר יחרגו מהמקום המותר, יהיו בכתב קטן מכדי שיהיה קריא, ו/או יכללו יותר משורת כתב אחת לכל קו לא ייבדקו.
- מומלץ מאוד לפתור תחילה את השאלה בדפי טיוטה ולהעתיק את עיקר הפתרון אל הטופס רק לאחר בדיקה.

חומר עזר מותר: מחשבון. דף-נוסחאות אחד דו-צדדי.

משך הבוחן: שעה

בהצלחה!

סימונים: נכתוב משתנים בכתיב וקטורי, כאשר .. $x, y,$ הם וקטורי עמודה. x_i מסמן את האיבר ה־i של וקטור x ו־$\bar{x}$ מסמן את הממוצע של וקטור x. עבור שני וקטורים x, y באורך n המכפלה הסקלרית שלהם היא $x^T y = \sum_{i=1}^{n} x_i y_i$. עבור וקטור x הנורמה בריבוע היא: $||x||^2 = x^T x = \sum_{i=1}^{n} x_i^2$.

1. עבור מדגם $\mathcal{D} = \{(x_1, y_1), ..., (x_n, y_n)\}$ מבצעים רגרסיה לינארית פשוטה (עם חותך). כעת הניחו שמבצעים טרנספורמציות לינאריות על x ועל y : $x' = ax,\ y' = cy$ עבור סקלרים $a, c \neq 0$ לקבלת מדגם חדש $\mathcal{D}' = \{(ax_1, cy_1), ..., (ax_n, cy_n)\}$ ועושים רגרסיה לינארית פשוטה של y' מול x'. יהיו $\hat{\beta}_0, \hat{\beta}_1$ אומדי הרבועים הפחותים של המדגם הראשון ו־$\hat{\beta}'_0, \hat{\beta}'_1$ של המדגם החדש D'.

(א) [20 נק'] $\hat{\beta}'_1$ שווה ל:

i. $\hat{\beta}_1$
ii. $\frac{a}{c}\hat{\beta}_1$
iii. $a\hat{\beta}_1$
iv. $\frac{c}{a}\hat{\beta}_1$
v. $c\hat{\beta}_1$

(ב) [20 נק'] $\hat{\beta}'_0$ שווה ל:

i. $\hat{\beta}_0$
ii. $\frac{a}{c}\hat{\beta}_0$
iii. $a\hat{\beta}_0$
iv. $\frac{c}{a}\hat{\beta}_0$
v. $c\hat{\beta}_0$

(ג) [15 נק'] ה־R'^2 של המדגם החדש D' במונחים של R^2 עבור המדגם המקורי D שווה ל:

i. R^2
ii. $\frac{a}{c}R^2$
iii. $\frac{a^2}{c^2}R^2$
iv. $\frac{c}{a}R^2$
v. $\frac{c^2}{a^2}R^2$

2. חוקר פוקימונים אסף נתונים על השפעת תכונות שונות של פוקימונים על מהירותם וקיבל מטריצה $\boldsymbol{X}$ של משתנים מסבירים ווקטור y של המשתנה המוסבר עם n תצפיות ו־p משתנים ($n > p$). החוקר מתאים מודל רגרסיה מרובה רגיל לנתונים עם אומד הרבועים הפחותים $\hat{\beta}$ והאומד הרגיל לשונות S^2. סטטיסטיקאי חדשני מציע לחוקר שיטה חדשה לניתוח הנתונים: שכפול כל הנתונים כך שכל תצפית תופיע פעמיים, ובסה"כ יהיו $2n$ תצפיות, ואז חישוב אומד הרבועים הפחותים והאומד לשונות הרגילים על הנתונים המשוכפלים. (מסומנים ב־$\hat{\beta}^{(d)}$ ו־$S^{(d)^2}$)

(א) [20 נקודות] הסטטיסטיקאי טוען שבשיטתו החדשה מספר התצפיות הוכפל ולכן השונות של אומד הריבועים הפחותים החדש עבור כל משתנה $\hat{\beta}_j^{(d)}$ קטנה יותר מאשר באומד הרגיל $\hat{\beta}_j$. האם הוא צודק? הוכיחו את תשובתכם.

(ב) [20 נקודות] הוכיחו **שהאומדן** לשונות $S^{(d)^2}$ עבור הנתונים המשוכפלים קטן יותר מאשר S^2 עבור הנתונים המקוריים .

(ג) [15 נקודות] בודקים בעזרת מבחן t את השערת האפס $H_0 : \beta_j = 0$ עם האלטרנטיבה $H_1 : \beta_j \neq 0$ עבור משתנה j כלשהו ועבור רמת מובהקות $\alpha > 0$ נתונה, פעמיים: פעם אחת מבצעים את המבחן עבור הנתונים המקוריים ופעם נוספת מבצעים את המבחן עבור הנתונים המשוכפלים (מבלי לדעת שהם משוכפלים). סמנו את התשובה הנכונה:

i. אם נדחה את H_0 במבחן עם הנתונים המשוכפלים , אזי נדחה את H_0 גם עם המבחן עם הנתונים המקוריים, אך לא בהכרח להיפך

ii. אם נדחה את H_0 במבחן עם הנתונים המקוריים , אזי נדחה את H_0 גם עם המבחן עם הנתונים המשוכפלים, אך לא בהכרח להיפך

iii. נדחה את H_0 במבחן עם הנתונים המקוריים אם ורק אם נדחה את H_0 במבחן עם הנתונים המשוכפלים

iv. אף אחת מהתשובות לעיל אינה נכונה