רגרסיה ומודלים ליניאריים 52320 תשע"ה 2014-15
בוחן 1 13.05.2015

בבוחן שאלות אמריקאיות ושאלות פתוחות.

משקל כל שאלה הוא 25 נקודות כך שמספר הנקודות הכולל הוא 125. 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 במחברת הטיוטה ולהעתיק את עיקר הפתרון אל הטופס רק לאחר בדיקה.

חומר עזר מותר: מחשבון.

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 $x^T y = \sum_{i=1}^n x_i y_i$.

תזכורת: נגדיר מודל לרגרסיה פשוטה עם חותך: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ עבור נתונים $(x_1, y_1), ..., (x_n, y_n)$. עבור מודל זה שגיאת הרבועים הפחותים $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$ ניתנת לכתיבה בצורות הבאות:
$SSE = \left(y - \hat{\beta}_0 - \hat{\beta}_1 x\right)^T \left(y - \hat{\beta}_0 - \hat{\beta}_1 x\right) = (y - \bar{y})^T (y - \bar{y}) - \hat{\beta}_1^2 (x - \bar{x})^T (x - \bar{x})$ כאשר האומדים למקדמים הם: $\hat{\beta}_1 = \frac{(x-\bar{x})^T (y-\bar{y})}{(x-\bar{x})^T (x-\bar{x})}, \ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

1. עבור מודל הרגרסיה הפשוטה עם חותך לעיל, יהיה $\hat{y}$ וקטור התחזיות ויהיה $e = y - \hat{y}$ וקטור השגיאות. אילו מהגדלים הבאים **תמיד** שווים לאפס? $\sum_{i=1}^n e_i, \ \sum_{i=1}^n x_i e_i, \ \sum_{i=1}^n y_i e_i$

   (א) כולם

   (ב) אף אחד

   (ג) רק $\sum_{i=1}^n e_i, \sum_{i=1}^n x_i e_i$

   (ד) רק $\sum_{i=1}^n e_i$

   (ה) רק $\sum_{i=1}^n y_i e_i, \sum_{i=1}^n x_i e_i$

   (ו) אף אחת מהתשובות לעיל אינה נכונה

2. נניח כעת כי נתונים כל הסכומים הבאים עבור נתונים $(x_1, y_1), ..., (x_n, y_n)$:

$$S_x = \sum_{i=1}^n x_i = 6.1, \ S_y = \sum_{i=1}^n y_i = 42.6, \ S_{xx} = x^T x = 16.45, \ \ S_{yy} = y^T y = 99.02, \ S_{xy} = x^T y = 21.01, \ n = 20$$

חשבו את $\hat{\beta}_0, \hat{\beta}_1$ ואת ה־SSE עבור מודל הרגרסיה עם חותך והנתונים שלעיל.

3. נגדיר כעת מודל רגרסיה **ללא** חותך: $y_i = \gamma_1 x_i + \epsilon_i$ עבור נתונים $(x_1, y_1), ..., (x_n, y_n)$. הוכיחו שבמודל זה אומד הריבועים הפחותים $\hat{\gamma}_1$ עבור γ_1 שווה ל-$\frac{x^T y}{x^T x}$.

4. חשבו את ערכו של $\hat{\gamma}_1$ ואת השגיאה הריבועית $SSE_1 \equiv \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ כאשר $\hat{y}_i = \hat{\gamma}_1 x_i$ עבור המודל **בלי** החותך עם הנתונים בשאלה 2 לעיל

5. באופן כללי, עבור רגרסיה לינארית פשוטה, נסמן ב-SSE את השגיאה הריבועית במודל עם חותך ואת SSE_1 את השגיאה הריבועית במודל ללא חותך עבור **אותם הנתונים**. סמנו את התשובה הנכונה

(א) תמיד $SSE_1 \leq SSE$ ויש מקרים בהם אי השוויון חזק

(ב) תמיד $SSE_1 \geq SSE$ ויש מקרים בהם אי השוויון חזק

(ג) תמיד $SSE_1 = SSE$

(ד) אף אחת מהתשובות לעיל איננה נכונה