

21 באוגוסט 2017

מיומנו של אקטואר: רגרסיה בצעדים (Stepwise Regression)

אחת הגישות האוטומטיות החזקות לניתוח רגרסיה הינה רגרסיה בצעדים. בהתבסס על שמה, תהליך הרגרסיה מתקדם במספר שלבים. ישנם מספר דרכים לקבוע את אלגוריתמי השלבים, כולל גישת הקורלציה, שיטת האקדמה, השיטה האחורנית ושיטת הקדימה ואחורה.

בשיטת הקורלציה, המשתנה התלוי (Y) מתואם עם כל המשתנים הבלתי תלויים (X) ורגרסיה מורצת, החל מהמשתנה ה- X עם ערך הקורלציה האבסולוטי הגבוה ביותר. לאחר מכן משתני ה- X הבאים מתווספים עד ש"ערכי ה- p " (ה- p -values) מצביעים על כך שמשתנה ה- X החדש איננו עוד מובהק סטטיסטית. גישה זו אמנם מהירה ופשוטה אך איננה כוללת בחובה את האינטראקציות שבין המשתנים, ומשתנה ה- X , כאשר מתווסף, יאפיל סטטיסטית על המשתנים האחרים.



בשיטת האקדמה, אנו תחילה מחושבות את הקורלציות שבין Y לבין כל משתני ה- X , מורצת הרגרסיה עבור Y על ערך הקורלציה האבסולוטי הגבוה ביותר של X ומתקבלות שגיאות ההתאמה. לאחר מכן, מחושבות הקורלציות שבין השגיאות הללו ובין יתר משתני ה- X , כאשר ערך הקורלציה האבסולוטי הגבוה ביותר מבין כל יתר משתני ה- X הוא זה שנבחר ואז מורצת רגרסיה נוספת. השיטה חוזרת על התהליך עד שערך ה- p של המקדם של משתנה ה- X האחרון איננו עוד מובהק סטטיסטית ואז התהליך נעצר.

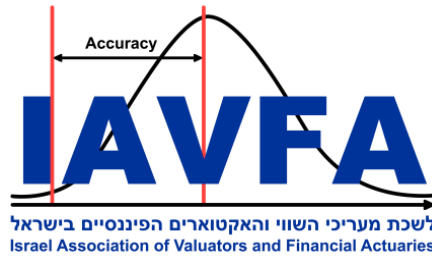
בשיטת האחורנית, מורצת רגרסיה של Y על כל משתני ה- X ואז למעשה נבדק ערך ה- p של כל אחד מהמשתנים, כאשר באופן שיטתי השיטה שוללת את המשתנה עם ערך ה- p הגבוה ביותר. לאחר מכן מורצת רגרסיה פעם נוספת וחוזר חלילה עד אשר כל ערכי ה- p הינם מובהקים סטטיסטית.

 WWW.IAVFA.ORG



ת.ד. 57334, תל אביב, 6157301

 077-5070590  153-77-5070590  IAVFA1020@GMAIL.COM



בשיטת הקדימה ואחורה, שיטת האקדמה מיושמת על מנת לקבל שלושה משתני X ולאחר מכן מיושמת גישת האחורנית על מנת לראות האם יש צורך לשלול את אחד מהמשתנים, בשל היותו מובהק סטטיסטית. יש לחזור על שיטת האקדמה ואז על שיטת האחורנית עד אשר כל יתר משתני ה- X ישקלו.

רגרסיה בצעדים הינה תהליך חיפוש אוטומטי המריץ איטרציות על כל המשתנים הבלתי תלויים והוא ממדל את המשתנים אשר נמצאו מובהקים סטטיסטית בהסבר ההשתנות במשתנה התלוי. רגרסיה בצעדים ביותר כאשר ישנם הרבה משתנים בלתי תלויים וניתן לבנות מספר רב של צירופי מודלים. לשם המחשה, נניח שאתה רוצה לקבוע האם ניתן לייחס את סך המכירות של מוצר מסוים למודעה מסוימת בעיתון מקומי. במקרה דגן שלפנינו, סך המכירות הוא המשתנה התלוי Y , בשעה שהמשתנים הבלתי תלויים X_1 עד X_5 הינם גודל המודעה, עלות המודעה, מספר הקוראים, היום בשבוע, וכמה פעמים בשבוע היא מופיעה. רגרסיה בצעדים מריצה איטרציות על משתני ה- X הללו על מנת לאתר את אלו שמובהקים סטטיסטית במודל הרגרסיה. פירוש ניתוח רגרסיה בצעדים הינו מורכב יותר יחסית לרגרסיה פשוטה והוא כולל השערות מבחני- t , מבחני- F , ANOVA, קורלציות צולבות, קורלציות עצמיות וכיוב'.

בכבוד רב,




רוני פולניצר

יו"ר ומנכ"ל לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל

 WWW.IAVFA.ORG



ת.ד. 57334, תל אביב, 6157301

 077-5070590  153-77-5070590  IAVFA1020@GMAIL.COM