

בחירת המאפיינים הסטטיסטיים של ויישומו לניתוח רב-משתני של נתוני בריאות

הרולד שיפ

בהנחיית: פרופ' בוריס פורטנוב, ד"ר איתי דטנר, פרופ' אמ' בנימין רייזר

אוניברסיטת חיפה, הפקולטה לניהול, המחלקה לניהול מידע וידע

באפדימיולוגיה, מחלת הסרטן ומחלות כרוניות אחרות, תוצאות הניתוחים הסטטיסטיים מצטברות לעיתים קרובות בתוך אזורים גיאוגרפיים. לעתים קרובות התוצאות הבריאותיות מצטברות בתוך אזורים גיאוגרפיים. עם זאת, רמת הצבירה הנדרשת יכולה לגרום לאובדן של מידע פרטני כולל מידע במרחב.

ישנה שיטה חלופית, אמד גרעין כפול (DKD), שיכולה לעזור בכמה חסרונות של צבירה. גישה זו הינה א-פרמטרית המשמשת להחלקת נתוני תצפיות באזור העניין ללא תלות בגבולות קבועים. בפרט, DKD הינה שיטת אומדני פונקציות שהיא היחס בין שתי פונקציות אחרות, שכל אחד מהם כשלעצמו נאמד על ידי החלקת סט של תצפיות לתוך פונקציה רציפה וחלקה.

לשיטת DKD יש יתרון חשוב נוסף הקשור לפרטיות. במחקר אפדימיולוגי, אסור לפרסם מידע אישי הקשור בבריאותו של אדם על ידי רשויות בריאות לחוקרים. שיטת החלקה של DKD מסתירה את מיקום התצפיות ובכך מגנה על הפרטיות של החולים ובכך מאפשרת מתן נתוני מקרי מחלה.

למרות השימוש ב-DKD במחקרים שונים, התכונות הסטטיסטיות של שיטה זו עדיין לא ידועות באופן מלא. הביצועים של החלקת הקרנל רגיש בפרמטר כוונון, את רוחב הפס, אשר שולט בכמות ההחלקה. אמנם אין שיטה מקובלת אחת כדי לבחור בפרמטר זה, קיים מספר טכניקות טובות בשימוש בפועל.

מטרת מחקר זה הינה לבחון את טכניקת אמד גרעין כפול מנקודת מבט סטטיסטי. בפרט, אנו משתמשים בסימולציות מונטה קרלו כדי ללמוד על המאפיינים האמפיריים של DKD, עם הדגש בשני תחומים: הדיוק הכולל של DKD והיכולת של DKD להעריך ערכי שיא.

התוצאות הראשוניות נראות תומכות אחד עם התנהגות אסימפטוטית התיאורטית של החלקת הקרנל. ההתפלגויות של טעות אמידה נראות בהתאם עם התאוריה של ההטיה והשונות באומדני חלקה הקרנל. הממצאים הראשוניים תומכים באומדן של ערכי השיא, הן הגודל והן המיקום. עם זאת, ניסיונות לחשב רווחי סמך עד כה לא היו מוצלחים.