

פינת הסטטיסטיקאית

שלום רב,

בימים אלו של חוסר ודאות, מה שמתאים לי, הוא להיצמד, ככל האפשר, לשגרה. אי לכך, הפינה של אפריל 2020, תעסוק בנושא שהתבשל אצלי כחודשיים (Before Corona) BC. לאחרונה גיליתי מספר תקלות בקבצי נתונים, שהגיעו אלי כקבצי אקסל. באחד מהמקרים, החוקר נאלץ להקליד את כל הקובץ מחדש, לאחר שבטעות ביצע מיון לא זהיר של נתוני הקובץ (מיון טור, מבלי לסמן את כל הקובץ).

אקסל הוא כלי זמין, קל יחסית לתפעול וחינמי; אך הוא בפירוש אינו מתאים לניהול מאגרי

מידע (!)

הסיבות לאי התאמה זו הן:

1. אקסל אינו כלי עקיב (traceable). בכלי עקיב, לכל נתון שמוקלד, ידוע מי הקליד אותו, מתי, ואם חל שינוי בנתון, מי שינה, מתי ומה היה קודם; כך שתמיד ניתן לשחזר אחורה.
 2. קל מאוד לגרום לנזק בלתי הפיך לקובץ, למשל באמצעות מיון לא זהיר.
 3. אם מקלידים נתונים ביותר ממחשב אחד, זה עלול לשבש את שדות התאריכים - להחליף בין יום לחודש, בחלק מהשדות; או ששדות שנראים כמו תאריך, לא מזוהים באקסל בתור תאריך.
 4. אם עושים גזור/הדבק של שדות שנוצרו מנוסחה, אפשר לקבל תוצאות מאוד לא צפויות. (חוקר אחד חווה מפח נפש, כאשר ממצא מדהים שגילה, התברר כטעות מאין זו).
 5. אם מעתיקים שדות שנוצרו מנוסחה, בשיטה של "ערכים בלבד", מאבדים את הנוסחה המקורית, כך שבעתיד לא יהיה ניתן לפענח ולהסביר איך נוצרו הנתונים.
 6. אין שיטה פשוטה למנוע הקלדת ערכים לא הגיוניים, כמו למשל משקל של 1,000 קילו.
 7. אין ביקורת על קיום שדה זיהוי חד-ערכי, מה שמאפשר הקלדת שורות נתונים שונים, עם אותו שדה מזהה. אי שמירה על שדה זיהוי תקין, מקשה על היכולת לאחד בין קבצים ממקורות שונים (נתוני אשפוז ותמותה, למשל).
 8. כאשר מקלידים הרבה שדות לנבדק, קל מאוד לדלג על עמודה ולהרוס את כל השורה. מכיוון שכך, ההמלצה החד משמעית שלי הוא להפסיק את השימוש באקסל לצורך ניהול נתונים. למזלנו, מרכז רפואי רבין תומך בכלי Redcap, שבהחלט עונה על הדרישות ממאגר מידע. ניתן לפנות לגלית בק: galitbe8@clalit.org.il בנושא זה.
- מאחלת לנו ימים רועשים, ילדים בגנים ובבתי הספר, חיילים אצל אימא שלהן וביקורים אצל הסבתא.

בברכה,

ציפי שוחט, יועצת סטטיסטית