

רגרסיה היא תהליך קבלת החלטות קליניות ביומיום של כולנו

מיכל כץ לוייר

ד"ר מיכל כץ-לוייר, החוג לפיזיותרפיה, אוניברסיטת תל אביב

מה פירושה של רגרסיה? הפירוש המילוני הפשוט: "נסיגה, צעד לאחור, שיבה לצורות קדומות וירודות יותר"¹. ובסטטיסטיקה? בסטטיסטיקה, אנו רגילים לפגוש את המושג "רגרסיה" במצבים מורכבים, במצבים בהם מוצגים מודלים לניבוי או לקשר, במודלים מתמטיים מרובי משתנים וסימנים. למעשה, הרגרסיה בסטטיסטיקה מצטלצלת כמו תהליך מתקדם שלא מתאים כל כך להגדרה המילונית.

לא רק מצטלצל, אכן כך. רגרסיה סטטיסטית זו משפחה של מודלים, בהם למרבית המקרים, ישנה תכונה אחת אותה רוצים לנבא באמצעות תכונות אחרות (אחת או יותר). בת המשפחה המוכרת ביותר זו הרגרסיה הליניארית (קווית). בת משפחה זו מתאימה למצבים בהם התכונה שאנו רוצים לנבא היא מדד מספרי כמותי, והנחת היסוד היא שקיים קשר קווי, בין התכונה שרוצים לנבא לבין תכונות אחרות.

'אחים' נוספים במשפחת הרגרסיה מוכרים פחות וכוללים בין השאר את "הרגרסיה הלוגיסטית" - המתאימה למצבים בהם המדד שאנו רוצים לנבא הינו בינארי (כן/לא, בריא/חולה, שמח/עצוב וכו'), "רגרסיה אורדינלית" - שמתאימה לנבא תכונות המתוארות בסולם סדר, ועוד ועוד 'אחים' ואחיות².

נעים מאד. אך האמת היא, שאין צורך בגינוני הכרות, שכן כולנו מכירים את המושג "רגרסיה" היטב. לא רק מכירים, אלא מדברים רגרסיה בסלנג המקצועי שלנו, יום יום שעה שעה. רוצים דוגמה?

בבוקר יום א' נכנס אלינו מטופל חביב לאחר שבר בצוואר הירך, ואמר "לא נעים לי להסתובב עם ההליכון בשכונה", ושאל **"מתי, לדעתכם, אוכל ללכת ללא אביזר עזר?"**

לדעתנו?? כן, יש משמעות לדעתנו שהרי אנו המומחים בתחום, קראנו לא מעט מאמרים וטיפלנו בלא מעט אנשים לאחר שבר צוואר ירך. לכן, מתאים לשאול אותנו לדעתנו, ולבקש שנבא את העתיד.

התשובה שמונחת על קצה הלשון היא הממוצע - ממוצע הזמן להליכה ללא אביזר עזר אצל אנשים לאחר שבר ירך שהגיעו לטיפול במכון.

אבל....

ברור לנו שעבור *המטופל היחיד* שניצב מולנו, האומדן שהצגנו - הממוצע, סביר להניח, איננו מדויק. שכן, הממוצע הינו תולדה של זמנים שהציגו אנשים שונים, עם יכולות שונות. אנו זוכרים כי חלקם הלכו ללא אביזר עזר בזמן קצר מהממוצע (מי יותר ופיחות) וחלקם הלכו ללא אביזר בזמן ארוך מהממוצע (מי יותר ופיחות).

אנו יודעים שאנו לא מדויקים באומדן שהצגנו. מאחר שאנחנו המומחים, אנו יכולים להאריך באופן גס את מידת חוסר הדיוק שלנו; ממצב בו "יש שונות רבה בין אנשים שונים בזמן שלוקח להגיע עד להליכה ללא אביזר עזר" עד מצב בו "אנשים שונים דיי דומים בזמן שלוקח להם להגיע להליכה ללא אביזר עזר".

במצב בו אנו מעריכים כי ישנה שונות רבה בזמן בין אנשים, אנו ממש נעים באי נוחות, מאחר וברור לנו שהאומדן שלנו לאיש העומד לפנינו איננו מדויק. לכן אנו פונים למטופל החביב ואומרים **"בא נערוך בדיקה קצרה ונוכל לתת לך תשובה יותר מדויקת"**. אנו שואלים שאלות, למשל, מה גילו של המטופל (שכן, אנו זוכרים שקיים קשר בין גיל לבין הזמן עד להליכה

1. אבן שושן, מילון חדש, תשכ"ג

2. לו היה שואל "האם אוכל ללכת ללא אביזר עזר" התשובה שלנו הייתה בינארית כן/לא והמודל המתאים היה רגרסיה לוגיסטית.

(כמונו) רגרסיה מסכמת את כל המידע, אך בשונה מאתנו מציגה בסימנים את מה שאנו אומרים במילים. למשל: " ככל שאדם מבוגר יותר משך הזמן ארוך יותר " (לכן סימן + במשוואה). או, "טווח תנועה רחב מלמד על משך זמן קצר יותר" (לכן סימן - במשוואה).

(כמונו), רגרסיה לא סבלנית עם תכונות שתרומתן לניבוי קטנה. לא סבלנית עד כדי כך שאינן נכללות בסיכום הכמותי הכולל (במספרים, ניתן לומר בזהירות שה-b של תכונות אלו לא שונה במובהק מ-0).

(כמונו), רגרסיה לא סבלנית עם משתנים מנבאים שיש קשר חזק ביניהם.⁴ הסטטיסטיקה מציגה באופן כמותי וברור ש"חבל על הזמן" לבצע שתיים או יותר בדיקות דומות (ושוב, במספרים, ניתן לומר בזהירות שה-b של אחד משני משתנים אלו לא שונה במובהק מ-0).

(כמונו) מטרת שילוב המידע היא לשפר את האומדן הראשוני שלנו, הממוצע. ברגרסיה ההצלחה של המודל להתקרב לערכי האמת מבוטאת באחוזים. באופן מקורב ניתן לומר שהאחוזים מבטאים את התועלת בשימוש במודל החדש מול מודל הניבוי הראשוני הפשוט בו השתמשנו - הממוצע.

האם כולנו מסכימים שהרגרסיה היא פשוט הגרסה הכמותית - מספרית של היומיום שלנו?

אני מקווה שחדי העין שמו לב שהתחמקנו מהתשובה, מדוע קוראים להליך המורכב הזה רגרסיה.

ללא אביזר עזר, במטופלים לאחר שבר צוואר ירך; ככל שאדם מבוגר יותר משך הזמן ארוך יותר). לאחר מכן, מבצעים הערכות שאנו מכירים שהן קשורות/ מנבאות את משך הזמן עד להליכה ללא אביזר עזר אצל מטופלים לאחר שבר בצוואר ירך.³ למשל, בודקים טווח תנועה אקטיבי בירך (אנו יודעים שטווח מוגבל מלמד על משך זמן ארוך יותר), ומודדים את משך הזמן שעורך לעלות ולרדת שמונה מדרגות (זמן ממושך קשור בזמן ממושך יותר עד להליכה ללא אביזר עזר). אנו רוצים להיות יעילים מחד ומדויקים מאידך ולכן בוחרים את השאלות וההערכות המתאימות ביותר ונותנים אומדן חדש, אחר.

אם שאלנו ובדקנו את המאפיינים המתאימים (המנבאים), אז סביר להניח שהאומדן החדש שלנו המבוסס על המידע שאספנו, קרוב יותר למשך הזמן שבאמת ייקח למטופל להגיע להליכה ללא אביזר עזר.

מסכימים שאלו הם חיינו? נכון שהסיפור נשמע ממש מוכר?

בסטטיסטיקה, פשוט נתנו שם להליך הלוגי הזה - "רגרסיה". התוספת בהליך הסטטיסטי הוא ב"כימות" המספרי של תהליך החשיבה הזה. דמיינו שכל המידע שיש לנו בראש ממטופלים קודמים, כל ההערכות, כל התוצאות של הליך השיקום של אנשים לאחר שבר בצוואר ירך נכנסים לתהליך הניבוי בתוספת מספרים וסימנים שיוצרים נוסחה.

המשתנה (התכונה) שרוצים לנבא הוא תוצאת הנוסחה. כל משתנה (תכונה) מנבא/קשור מקבל משקל כמותי (מסומן באות b) בהתאם לאופיו ולחשיבותו בניבוי. המבנה של הנוסחה בסיפור שלנו יראה כמו:

$$\text{זמן מדרגות} + b3 + (\text{טווח תנועה אקטיבי בירך}) - b2 - (\text{גיל}) + a + b1 = \text{משך זמן}$$

3. כדאי לנהוג לפי "חוק החסכנות" (Occam's razor) - התרגום של החוק להערכות בפיזיותרפיה הוא שיש לבדוק אך ורק את התכונות עם תוקף ניבוי גבוה ליכולת הליכה ללא אביזר בסיום תקופת הטיפול במכון. וכי אין טעם לבדוק תכונות שלא קשורות או לבדוק פעמיים אותו דבר, או מאד (מאד) דומה.

4. בסטטיסטיקה משתנים בלתי תלויים שמספרים אותו דבר, נמצאים במצב של "מולטיקולינאריות" במילים סטטיסטיות, המשתנים קשורים חזק אחד לשני ולפיכך, לצורך המודל, אנו לא זקוקים לשניהם.