

## פרק עשרים וחמישה

### רעיון פשוט

תרתי אחר דרך לעשות רפליקציה פשוטה, הדומה למדגמים המקוריים, והמבוססת על החלפת השמות והכינויים שבהם בכינויים חדשים, בעוד התאריכים נותרים על כנם. שאפתי לערוך החלפה זו בדרך קלה ומכנית עד כמה שאפשר. כל מה שהיה חסר לי היה הרעיון עצמו... והנה, אחד ממבקריו גלגל את הרעיון לפתחי. הבה נתחיל מן הסוף.

בשנת התשנ"ט (1999 למנינם) פרסמתי באינטרנט מאמר המתאר עבודה מן הסוג המבוקש. עבודה זו הוצגה לראשונה בהרצאה שנשאתי ב"ז בטבת באותה שנה, ב"מרכז לחקר הרציונליות וההחלטות האינטראקטיביות" באוניברסיטה העברית בירושלים. אצטט כאן את התקציר למאמר:

### רפליקציה

#### למדגם השני של גדולי חכמי התורה

#### תקציר

הניסוי המרכזי המתואר במאמרנו *Equidistant Letter Sequences in the Book of Genesis*, נעשה לגבי הרשימה *השניה* של גדולי חכמי התורה. לצורך רפליקציה של ניסוי זה, הוכנו שתי רשימות נוספות של שמות וכינויים חדשים לגבי אותם האישים – זאת לפי רעיונו של אחד המבקרים. מדידת המפגשים בין השמות לתאריכים נעשתה בדומה לניסוי המתואר במאמרנו המקורי. המובהקות הסטטיסטית נמדדה ונמצאה גבוהה לגבי אחת מן הרשימות:  $p = 0.000042$ .

בעקבות מדידה זו, הוכנו באותה השיטה שתי רשימות חדשות לגבי האישים ברשימה *הראשונה* [שהתפרסמה אף היא במאמרנו המקורי]. המובהקות הסטטיסטית נמדדה, ולגבי אחת מן הרשימות היה ערכה:  $p = 0.0344$ .

במבוא למאמר, הזכרתי את עבודתנו המקורית, אשר לפי מאפייניה תוכנן הניסוי החדש:

מאמרנו *Equidistant Letter Sequences in the Book of Genesis* שהתפרסם ב-*Statistical Science* עסק במתאם שבין שמותיהם של גדולי ישראל לבין תאריכי הלידה והפטירה שלהם, המופיעים בדילוגים שווים של אותיות בספר בראשית. הניסוי המרכזי במאמר, מבחן הראנדומיזציה, נערך לגבי הרשימה *השניה* של גדולי ישראל, והצביע על מובהקות גבוהה של המפגשים בין השמות לתאריכים:  $p = 0.000016$ .

על רקע זה, הסברתי מהו הרעיון החדש לרפליקציה, וכיצד הוא נולד:

בעבודה הנוכחית נערך ניסוי על רשימה של כינויים חדשים לאותם האישים, רשימה המבוססת אך ורק על הכללים הקיימים כבר, ללא הזדקקות להחלטות חדשות של מומחים. ברשימה השניה של השמות והכינויים<sup>1</sup> (שהוכנה על ידי המומחה לביבליוגרפיה, פרופסור ש"ז הבלין), היו שמות וכינויים מן הסוגים הבאים:

1. "רבי פלוני".
2. שם משפחה.
3. שם מלא.
4. כינוי מקובל.
5. כינוי של חכם על שם יצירתו.

המתמטיקאי פרופסור אלכס לובוצקי העיר במאמר ביקורתי<sup>2</sup>, שאפשר להשתמש גם בכינויים מסוג אחר, כינויים בהם אדם נקרא על שם אביו. לפי לובוצקי,<sup>3</sup> אדם מתייחס לשם אביו על ידי הביטוי "בן" או על פי הקיצור "ב"ר".

מילון אבן שושן<sup>4</sup> מבאר, כי "ב"ר" הוא קיצור ל"בן רבי". לפי רעיון זה, אם שם האב "פלוני", הבן יכונה על שם האב בביטוי "בן פלוני" או בביטוי "בן רבי פלוני".

הכנתי שתי רשימות חדשות של כינויים של אישי הרשימה השניה: רשימה אחת ובה כינויים מן הצורה "בן פלוני", ורשימה נוספת ובה כינויים מן הצורה "בן רבי פלוני". רשימות אלו שימשו בסיס למדידת המפגשים בין הכינויים לתאריכים, בדיוק כפי שנעשה בעבודתנו המקורית.

להשלמת החקירה ערכתי ניסוי נוסף: הכנתי שתי רשימות כאלו גם עבור אישי הרשימה הראשונה, ומדדתי את המפגשים באותה צורה. בגוף המאמר מתוארת דרך איסוף הנתונים עבור הרשימות החדשות והם מוצגים במלואם. דרך החישוב מוצגת בצד התוצאות במלואן. הדברים נמצאים בנספח א14.

- - -

בעבודה המתוארת לעיל נוצרו שני מדגמים עם הכינויים מן הצורה "בן פלוני" (להלן מדגמי "בן"), ושני מדגמים עם הכינויים מן הצורה "בן רבי פלוני" (להלן מדגמי "בן רבי"). מדידת המובהקות של מדגמי "בן רבי" נעשתה באותו אופן שבו נערכה מדידת מדגמי "בן": התייחסתי לשני סוגי המדגמים באופן שווה, והפעלתי עליהם אותו אלגוריתם. הבה נתבונן בתוצאות מקרוב ונבדוק מה משמעותן.

א. נזכיר כי המטרה היתה להחליף את "כינוי הבלין" בכינויים חדשים ולמדוד מחדש את המדגמים המתקבלים. האומנם הצלחנו ליצור רפליקציות לניסויים המקוריים?

<sup>1</sup> הרשימה השניה נמצאת בטבלה 2 בנספח א6.

<sup>2</sup> א' לובוצקי, צפנת פענח, "הארץ" א' אלול התשנ"ז (3 בספטמבר 1997 למנינם).

<sup>3</sup> דוגמתו של לובוצקי לשתי הצורות היא לגבי "הגאון מוילנא" (מס' 5 ברשימה הראשונה), ששם אביו היה "שלמה זלמן": הוא נקב בצורה "בן שלמה זלמן", ובצורה "ב"ר שלמה זלמן".

<sup>4</sup> א' אבן שושן, המלון החדש, קרית ספר, ירושלים התשמ"ט, ערך "ב"ר" עמ' 139.

מתברר, כי יש הבדל מהותי בין מדגמי "בן רבי" למדגמי "בן" מבחינת כמות הזוגות (כינוי, תאריך) המופיעים בדילוג שווה בספר בראשית.

1. במדגמי "בן" יש הרבה זוגות כאלה (89 עבור אישי הרשימה הראשונה ו-52 עבור אישי הרשימה השנייה) ומיוצגים בהם רבים מאישי כל מדגם (פרטים בנספח א14).
2. לעומת זאת במדגם "בן רבי" עבור אישי הרשימה הראשונה ישנו כינוי אחד ("בן רבי משה") בלבד המופיע בדילוג שווה בספר בראשית, והוא נוגע לשני אישים בלבד (מתוך 34). במדגם "בן רבי" עבור אישי הרשימה השנייה ישנם שני כינויים ("בן רבי משה" ו"בן רבי בָּעֵר") כאלה, והם נוגעים ל-4 אישים בלבד (מתוך 32).

מסקנות:

- (א) יש במדגמי "בן" דגימה של רוב האישים וניתן לראות בהם רפליקציות.
- (ב) אין במדגמי "בן רבי" משום רפליקציות בגלל שהם דוגמים אישים ספורים בלבד מתוך הקבוצות העומדות לבדיקה.

## ב. תוצאות מדגמי "בן"

1. תוצאות מדגם "בן" עבור הרשימה השנייה של גדולי חכמי התורה: הרצת מבחן הפרמוטציות על מדגם "בן" גילתה, כי הוא מצליח במובהקות חזקה. הסטטיסטי "המצליח" ביותר במבחן זה היה "המידה הכוללת לקרבה"  $P_1$ , וההסתברות לדירוג הגבוה שלו במרוץ עם המדגמים המשובשים היה:  $p = 0.000021$  (בערך אחד לחמישים אלף).
- כדי לקבל את המובהקות הסטטיסטית הכוללת עבור שני הסטטיסטים שנמדדו,  $P_1$  ו- $P_2$ , עלינו להכפיל תוצאה זו פי 2 (בונופֶרוֹני). כך נקבל:  $p = 0.000042$  (בערך אחד לעשרים וחמישה אלף).

2. תוצאות מדגם "בן" עבור הרשימה הראשונה של גדולי חכמי התורה:

הרצת מבחן הפרמוטציות על מדגם "בן" העלתה כי אין מובהקות (פרטים בנספח א14).

## ג. צירוף כינויי "בן" למדגם הגדול השני

עד כאן מדדנו את מדגמי "בן" בפני עצמם. אולם, יש עניין לראות מהי תוצאת הוספת כינויי "בן" למדגם הגדול השני (כל מה שטען פרופסור לובוצקי הוא, שניתן להשתמש גם בכינויים מסוג זה). מתבררות העובדות הבאות:

1. נזכיר, כי אמות המידה המקוריות שלנו להצלחת המדגם הגדול השני היו ערכי הסטטיסטים  $P_1$  ו- $P_2$ . צירוף כינויי "בן" למדגם הגדול השני נותן את התוצאות הבאות:

(א) ערכו של  $P_1$  משתפר (קטן) פי 16,900.

(ב) ערכו של  $P_2$  משתפר (קטן) פי 465.

2. המובהקות הסטטיסטית החזקה ביותר עבור המדגם הגדול השני התקבלה במבחן הרנדומיזציה עבור הסטטיסטי  $P_4$ . צירוף כינויי "בן" למדגם הגדול השני, מחזק את מובהקות  $P_4$  פי 450, וערכה הוא

$$.r_4 = 1.5 \times 10^{-9}$$

ד. מדגמי "בן רבי"

כאמור לעיל (בסעיף א) מדגמים אלה אינם יכולים לשמש כרפליקציה, בגלל מיעוט האישים הכלול בהם. אולם, ראויים הם להצטרף למדגמים המקוריים. נזכיר, כי אמות המידה המקוריות שלנו להצלחת המדגמים המקוריים היו ערכי הסטטיסטים  $P_1$  ו- $P_2$ .

1. צירוף כינויי "בן רבי" למדגם הגדול השני נותן את התוצאות הבאות:
  - (א) ערכי הסטטיסטים נחלשו (גדלו) פי 5-7.
  - (ב) במבחן המובהקות, נחלשה המובהקות הכוללת פי 4.
2. צירוף כינויי "בן רבי" למדגם הראשון נותן את התוצאות הבאות:
  - (א) ערכי הסטטיסטים השתפרו (קטנו) פי 3-5.
  - (ב) במבחן המובהקות, התחזקה המובהקות הכוללת פי 4.  
(ראו עוד פרטים בנספח א14).