

בעניין המבחן הסטטיסטי שהתפרסם במאמרו ב- *Statistical Science*

חלק שני

מאת דורון ויצטום

מבוא:

במיסמך שפירסמתי "בעניין המבחן הסטטיסטי שהתפרסם במאמרו ב- *Statistical Science*" (להלן נכנהו "החלק הראשון"), הצבעתי על טענת השקר שפירסמו מקי ושות' בעניין מקורו של מבחן זה, במאמר "Solving the Bible Code Puzzle" [1] שהופיע בגליון מאי 99 של ה- *Statistical Science*. בתגובה, פירסמו מקי ושות' מאמר באינטרנט [2] ושמו "The origin of the permutation test" ובו הם מנסים, ללא הצלחה, לחפות על כך שנתפסו בקלקלתם. כפי שנראה להלן, הדברים שכתבו רק מחזקים את טענותי. יתר על כן, מקי ושות' הסתבכו בשקרים נוספים ובהסתרת מידע.

א. על איזה ניסוי היה ההסכם:

1. לפי המכתבים שצוטטו על ידי מקי ושות' בתגובתם [2] באינטרנט, ברור לגמרי כי דיאקוניס הוא אבי ההצעה לערוך מבחן של פרמוטציות. לגבי שיטת הביצוע של המבחן, מבחינים מקי ושות' בין שתי שיטות לערוך את מבחן הפרמוטציות שהציע דיאקוניס: "type A" המתוארת במכתבי פרופ' אומן, ו- "type D" המתוארת במכתבי פרופ' דיאקוניס. במאמרם [1], כותבים מקי ושות' שהיתה הסכמה על type D:

"To correct the error in treating P_{1-4} (that is, P_1, P_2, P_3 and P_4) as probabilities, Diaconis proposed a method that involved permuting the columns of a 32×32 matrix, whose (i, j) th entry was a single value representing some sort of aggregate distance between all the appellations of rabbi i and all the dates of rabbi j . This proposal was apparently first made in a letter of May 1990 to the Academy member handling the paper, Robert Aumann, though a related proposal had been made by Diaconis in 1988. The same design was again described by Diaconis in September (Diaconis, 1990), and there appeared to be an agreement on the matter." (Section 3, emphasis mine)

זה שקר.

מקי ושות' עצמם, בנתחם את המכתבים ב[2], טוענים שהיתה אי הסכמה ואי הבנה, ושפרופ' אומן דבק כל הזמן בשיטה A, בניגוד גמור למה שכתבו ב- *Statistical Science*.

2. מקי ושות' ממשיכים וכותבים ב[1]:

"However, unnoticed by Diaconis, WRR performed the different permutation test described in Section 2."

בתיאור זה הם יוצרים את הרושם שאנחנו רימינו את דיאקוניס, וביצענו מאחורי גבו ניסוי אחר, בניגוד להסכמה שהיתה.

זה תאור מכווער ומטעה מאד. אפילו מתוך הדברים המקוטעים המצוטטים במאמרם [2] באינטרנט ברור שדבר זה אינו נכון.

הערה: עד כאן צטטנו את מקי ושות' על כך שדיאקוניס הציע שיטה D. לדעתנו, דיאקוניס שהוא אבי ההצעה לערוך מבחן של פרמוטציות, התרכז בעיקרון ולא נכנס לפרטים. ולכן, במבט שיטחי נוצר רושם של הצעת שיטה אחרת, כפי שכתב ריפס למקי:

“My guess is that Professor Diaconis probably did not look into the technical details of our work, and therefore he describes them in a general, vague and non-precise way.”

(ציטוט ממכתב מ-4 באפריל 97, אותו מכתב ש[2] מצטט ממנו).

יחד עם זאת, אינני מתכוון להיכנס לפולמוס עם מקי ושות' על פרשנותם, שהיתה שיטה D. אדרבא, כל הדיון להלן יהיה מבוסס על טענתם שדיאקוניס אכן הציע שיטה אחרת, שיטה D.

הסכמת דיאקוניס לשיטה A:

מתוך חליפת המכתבים שמקי ושות' מצטטים ב[2] עולה בבירור, שאף אם נאמר שדיאקוניס הציע את שיטה D, הרי וודאי שאומן הציע את שיטה A, וששלב הסופי הסכים דיאקוניס לשיטה A (המכתב נמצא בחלק הראשון). בכל משא ומתן הקובע הוא ההסכם.

לנו נראה, שהסכמתו של דיאקוניס לא היתה מבלבול ולא מחוסר תשומת לב, אלא מכך שהוא לא ייחס חשיבות רבה לכך האם הניסוי יהיה בשיטה A או D. על כל פנים ברור הדבר, שדיאקוניס סיפק ראיות משכנעות לרושם זה לאורך שבע שנים, עד שאף אחד, כולל פרופ' אומן, לא העלה על דעתו את האפשרות שאולי אין זה כך. הראיות הן:

1. דיאקוניס קיבל טיוטה של המאמר (יחד עם כל שופטי המאמר שלנו), כולל תיאור הניסוי, לפני ביצוע הניסוי. מתוך הנחה פשוטה שלא מעל בתפקידו כשופט, חזקה עליו שבדק שאכן זה הניסוי שהוסכם לעשותו.

2. בשלב הבא, לאחר הניסוי, הוא קיבל את המאמר עם התוצאות. במכתב הנלווה, מיום 6 בדצמבר 91 (לא הובא על ידי מקי ושות') כותב אומן לדיאקוניס בין השאר:

“Enclosed is the paper of Witztum, Rips, and Rosenberg. The presentation was revised somewhat to make it clearer and take into account the comments of the people to whom I had previously shown it. Needless to say, the test itself was not changed in any way; it is precisely the one to which we agreed in the summer of 1990.” (ההדגשה שלי)

אם לא היתה הסכמה, ודיאקוניס חשב שמה שנעשה אינו בסדר, היה מוחה על כך. בייחוד, אם נזכור, שדיאקוניס לא רצה בפירסום התוצאות, והיה צריך למצוא תואנה משונה כדי להמליץ נגד הפירסום. לפי מקי ושות' היה יכול דיאקוניס לפסול את הניסוי בקלות בטענה שאין זה הניסוי אליו התכוון!

3. בקיץ של 92 היו שיחות ודיונים בין אומן לבין דיאקוניס בנושא המחקר שלנו לגבי עבודות נוספות. דיונים אלה הולידו מכתב הסכמה מיום 28 באוגוסט 92. מכתב זה ברור שדיאקוניס אינו חולק על כך ש: “the significance level is over 99.998%”, ולא הזכיר אפילו ברמז איזו שהיא הסתייגות משיטת הניסוי – שיטה A, אף לא לגבי עבודות עתידיות.

4. בנובמבר 97, כותבים מקי ושות' במאמרם בגלילאו מס' 25 עמ' 53:

“פרופ' פרסי דיאקוניס, מתמטיקאי וסטטיסטיקאי בעל שם עולמי... הציע להם

[WRR -] שיטה אחרת, ובה הם השתמשו במאמר שפורסם ב- Statistical

Science.” (ההדגשה שלי).

סיכום:

אפילו לטענתם ברור:

- א. שדיאקוניס הוא אבי הרעיון להשתמש במבחן של פרמוטציות.
- ב. שבשלב הראשון דיאקוניס הציע שיטה D ואומן הציע שיטה A.
- ג. שסוכם על שיטה A (אף שלטענתם סיכום זה היה בחוסר תשומת לב של דיאקוניס).

- ד. שדיאקוניס קיבל את המאמר כשופט הן לפני עריכת הניסוי והן לאחר הניסוי ואישר זאת.
- ה. שלאורך שנים בהתכתבויות והתייחסות לנושא, לא מחה מעולם על כך שהניסוי נעשה לפי שיטה A.

והנה, לאחר שבע שנים, בשנת 97, מגיע מקי ומנסה לשכנע את דיאקוניס שהוא לא שם לב להבדל בין שיטה A ל-D, ושכל ההסכמה שלו נבעה רק מחוסר תשומת לב. דבר זה קשור לניתוח פסיכולוגי של הכוונות והמחשבות של דיאקוניס, אך אין זה רלוונטי כלל להסכם בין אומן לדיאקוניס ולאישורים שנתן האחרון כשופט.

אם למקי יש טענות – שיבוא בטענות לדיאקוניס. אך ליצור את הרושם כאילו פעלנו בחוסר הגינות – זו הטעיה זדונית של דעת הקהל.

ב. מדוע שיטה A:

בפרק א הבהרתי כי שיטה A היא זו שסוכם עליה. כאן המקום להבהיר גם מדוע זו השיטה הנכונה, שלא לדבר על כך ששיטה D מעולם לא הוגדרה באופן אופרטיבי: המכתב של דיאקוניס מ-5 בספטמבר 90 מגדיר כמה פרטים בניסוי בצורה שתומה מאד, ובפרט אינו מגדיר את ה"מרחק" t (לפרטים ראה בחלק הראשון).

כפי שמקי ושות' כתבו ב[1], תכלית הצעותיו של דיאקוניס היתה

"to correct the error in treating P_{1-4} (that is, P_1, P_2, P_3 and P_4) as probabilities".

טענתו של דיאקוניס היתה, שהשימוש במדדים אלה באופן ישיר למדידת המובהקות אינו נכון. שימוש כזה מבוסס על הנחה של אי-תלות ואוניפורמיות, הנחה שאינה מוצדקת לדעת דיאקוניס. מבחן הפרמוטציות נועד לנרמל את המדדים P_{1-4} , ולכן הדרך לעשות זאת היא להשוות את P_{1-4} של המידגם המקורי, לערכי P'_{1-4} של "מידגמים" המתקבלים על ידי פרמוטציות. זו היא בדיוק שיטה A.

לעומת זאת, שיטה D כלל אינה מנרמלת את ערכי המדדים P_{1-4} שהתקבלו בניסוי המקורי. במקום זה היא מנרמלת ערכים של P_{1-4} המתקבלים עבור תוצאות מדידה שונה לגמרי (ה"מרחק" t).

ג. סילוף שיטה D והעלמת נתונים על ידי מקי ושות':

1. סילוף שיטה D:

במאמר [2] כותבים מקי ושות':

"Aumann and Diaconis had agreed that a significance level of 1/1000 was a reasonable criterion for success. When WRR applied the permutation test they had designed themselves, they met that target easily. If they had performed the test that Diaconis wanted, on the same data, they would missed it. The failure of the experiment to pass the 1/1000 threshold would have greatly reduced its prospects of ever being published in a scientific journal."

נראה להלן כי טענה זו שקרית.

מקי ושות' דאגו לכתוב פסקה זו בצורה מספיק מעורפלת, ואפילו לא הסבירו איך שיטה D צריכה להתבצע. תשובה לשאלה "איך" נמצא במאמר [1], בו טוענים מקי ושות' בקשר לשיטת דיאקוניס כי השימוש בממוצע אריתמטי ("using the average distance" מהווה

"the most obvious definition of his [Diaconis'] 32x32 matrix".

ובצורה מפורשת טוענים הם כל זאת במאמר אחר [3] (בהתייחס לשיטה B שתוסבר בפיסקה הבאה):

"This is a good place to note that (B) is the most natural interpretation of the experiment which WRR were asked to perform in 1990 by Persi Diaconis (on behalf of the journal to which their paper was first

submitted). They failed to do so, but if they had the experiment would not have passed the 1/1000 milestone set for them. Whatever is the reason for it, the fact remains that [WRR] would quite likely not have been published if the Prof. Diaconis' instructions had been followed". (Emphasis mine)

את B הגדירו מקי ושות' ב[4]:

"For each pair of persons p, p' , compute one distance $t(p, p')$ by averaging the defined values $c(w, w')$ where w is in the first word-set of p and w' is in the second word-set of p' . If there are no such values defined, $t(p, p')$ is undefined. For a permutation π of the persons, define $T(\pi)$ to be the average over all p of the defined values $t[p, \pi(p)]$. If there are no such defined values, $T(\pi)$ is undefined. The result will be the rank position of $T(id)$ amongst all defined $T(\pi)$ for a large set of random permutations π ."

אבל שיטה B המתוארת כאן אינה תואמת כלל את שיטה D במכתבו של דיאקוניס מ-5 בספטמבר 90: כאן הסטטיסטיקה $T(\pi)$ שהוא ממוצע, באה במקום שתי הסטטיסטיקות P1 ו-P2 שבשיטה D! P3 ו-P4 הן חזרה על P1 ו-P2 בלי "רבי", כך שיש סך הכל ארבע סטטיסטיקות).

נפרט: בשיטה B שלשה שלבים. למשל, לגבי המידגם השני:

בשלב א, עושים ממוצע $t(p, p')$ של כל ערכי המפגשים, של הזוגות (שם, תאריך), השייכים לאדם מסויים, ומקבלים מספר אחד. אם קבוצת ערכי המפגשים ריקה, הממוצע לא מוגדר. כך עושים לגבי כל האישים במידגם ומקבלים N מספרים, כמספר הממוצעים המוגדרים.

בשלב ב, עושים ממוצע של N המספרים הנ"ל ומקבלים מספר יחיד: $T(id)$. בשלב ג, עושים כך גם לגבי המידגמים המתקבלים מן המידגם השני על ידי פרמוטציות המתאימות לכל אישיות תאריך הקשור לאישיות אחרת. מדרגים את $T(id)$ בין ה- T' המתקבלים על ידי הפרמוטציות. דירוג זה הוא המובהקות.

ברור שבשיטה זו אין אפשרות לבדיקה על ידי P1 ו-P2 (ולא על ידי P3 ו-P4). וזה בניגוד משווע למכתב של דיאקוניס מ-5 בספטמבר 90, בו כתוב שישנה הסכמה על ארבע הסטטיסטיקות! ומכתב זה של דיאקוניס היה מענה למכתב של אומן מיום 19 ביוני 90 בו הגדיר את ארבע הסטטיסטיקות.

ואילו מקי ושות' טוענים בזה שדיאקוניס התכוון לסטטיסטיקה אחת ויחידה T, אחרת לגמרי מארבע הסטטיסטיקות שהוזכרו!

לכן, ברור ששיטה זו המוצעת על ידי מקי ושות' היא המצאה שלהם, ומעולם לא הוצעה על ידי דיאקוניס לאומן. האבסורד בולט שבעתיים על רקע דברי מקי ושות' ב[2] לגבי המכתב הנ"ל של דיאקוניס: "very little in his description is incorrect or vague". אני רוצה לראות מי עלי אדמות יכול להבין מתוך משא ומתן מסוכם על P4, P3, P2, P1, שהכוונה בעצם ל-T, הסותר לחלוטין את P4, P3, P2, P1.

אז איך מעיזים מקי ושות' לטעון:

"(B) is the most natural interpretation of the experiment which WRR were asked to perform in 1990 by Persi Diaconis".

הגדרת t:

עד כאן דברנו על הגדרת T בשלב ב דלעיל. הגדרת t לעומת זאת, שייכת לשלב א. מבדיקת המכתבים המשקפים את המשא ומתן בין אומן לדיאקוניס, מתברר שלטענה שדיאקוניס התכוון כאן לממוצע, אין שום כיסוי. המקום היחיד שבו דן דיאקוניס בהגדרה

אפשרית ל- t (כלומר, בערך מספרי מייצג עבור קבוצת ערכי המפגשים השייכת לאישיות אחת) הוא במכתבו מיום 3 באוגוסט 88, לצורך מבחן פרלמינארי שהוא עצמו ביצע. ובמקום יחיד זה הוא כותב "smallest" ולא "average". בפרק ד נסביר גם למה אי אפשר להגדיר את t כממוצע אם רוצים להשתמש בשיטה D. אני מזכיר לקורא, שנוקקתי לכל זה רק כדי להראות את חוסר היושר בטענות מקי ושות', והסילוף המכוון והשיטתי שלו. עד כמה שהדבר נוגע לנו, ידענו רק מה שפרופ' אומן אמר לנו: שהוסכם על שיטה A.

2. העלמת נתונים:

כל הדברים שהעלנו עד כה חמורים דיים. אך מיד נראה שאין זה הכל. בעקבות הצגת שיטה B על ידי מקי ושות' בפרוטוקול מיום 17 באפריל 97 (פרוטוקול שקדם ל[מק4]), שלח עמיתי פרופ' ריפס מכתב למקי ביום אחד במאי 97' ובו הסביר מדוע הוא דוחה את שיטה B, ובמיוחד מדוע השימוש בממוצע אריתמטי אינו נכון. הוא חזר על דבריו אלה בביקורתו על [3]. ראה [5] (דבריו יובאו בפרק ד). ביום 9 באוגוסט 98, פירסם מקי מיסמך בשם: "Revisiting The Permutation Test". במיסמך זה הוא מתאר ניסוי בו הוא עושה ווריאנט על שיטה B (ראה הגדרה בסעיף הקודם), כאשר בשלב א הוא מגדיר את t להיות:

"the average of the logarithm of the defined $c(w, w')$ values."

בשלב ב הוא מגדיר את T כסכום על הממוצעים הלוגריתמיים שהתקבלו בשלב א. ניסוי זה בא לענות על השגותיו הנ"ל של ריפס, וכך כותב מקי עצמו:

"Use of the logarithm gives much stronger prominence to the word pairs with small distances' and in my opinion meets the objection of Rips that ordinary mean 'averages out' the alleged ELS phenomenon."

מקי מציג את התוצאות שקיבל:

"For the WRR data, this method gives very respectable scores of 125/million for the first list and 8/million for the second. The combined list surely gives very much better than 1/million, but I have not computed it."

כלומר, מקי מדווח כאן על הצלחה של $8/1,000,000$ עבור המידגם השני! מיסמך זה נשלח על ידו בתאריך הנזכר בדואר אלקטרוני לשאר מחברי [1]. לניסוי זה אין זכר ב[1].

מסקנות:

- מקי ושות' יודעים היטב שהצלחת הניסוי לא נבעה מבחירה מכוונת של שיטת הרנדומיזציה, שהרי גם שיטת מקי הצליחה.
- מקי ושות' העלימו ממערכת ה- St. Sc. את קיומו של הניסוי הזה.

המשמעות של העלמה זו היא קיצונית: כאשר מעלימים תוצאות סטטיסטיות רלוונטיות, אין ערך ומשמעות לשאר הדברים שהוצגו. אין לנו ספק, שאם מערכת ה- St. Sc. היתה יודעת שמקי ושות' מעלימים מהם נתונים, לא היו מפרסמים את המאמר שלהם.

ועוד הם מעיזים לכתוב במאמרם [1]:

"Nothing we have chosen to omit tells a story contrary to the story here."

3. האומנם לא היינו עוברים את הסף של 1/1000?

כאן נציג תוצאות של ניסויים נוספים, שהם על בסיס שיטה D. אם t אמור להיות גודל מספרי המייצג או מסכם את כל ערכי מפגשי הזוגות (שם, תאריך) השייכים לאישיות מסוימת, הרי בעבודה המקורית שלנו היו רק שני מדדים כאלה לסיכום התוצאות: P_1 ו- P_2 .

אומנם, השימוש ב- $t=P_1$ הוא בעייתי כאשר גודל קבוצת הערכים קטן מאד, כפי שצפוי לקרות כאן עבור חלק מן האישים (ישנן אפילו קבוצות בנות 1-2 ערכים), אבל לשם השלמת התמונה נשתמש גם באפשרות זו.

אם כך, עלינו למדוד ארבע סטטיסטיקות עבור הבחירה $t=P_1$, וארבע עבור $t=P_2$. סך הכל 8 מדידות. התוצאה החזקה ביותר מתקבלת עבור הסטטיסטיקה P_4 במקרה $r=0.000013$: $t=P_2$.

המובהקות הכוללת היא, אם כן, $r=8 \times 0.000013=0.0001$, זו תוצאה הטובה פי 10 מן הסף של $1/1000$.

כדי להשלים את התמונה, נערוך את הניסוי עם $t=\text{smallest}$, שהוא הערך בו בחר דיאקוניס לצורך ניסוי פרלמינארי שהוא ביצע. זה הדבר היחיד הנזכר במכתבי דיאקוניס. עבור $t=\text{smallest}$ מקבלים: $\min r_{1-4}=0.000153$. זו תוצאה הטובה פי 1.5 מן הסף של $1/1000$ (כל הניסויים נעשו עם 10,000,000 פרמוטציות ועם אותו זרע סטטיסטי ששימש בניסוי המקורי).

לסיכום: השתמשנו כאן להגדרת t בכל הדרכים לסכם קבוצת ערכי $c(w, w')$ שנוכרו קודם הניסוי, ולא מצאנו כיסוי לטענת מקי ושות' שלא היינו עוברים את הסף של $1/1000$.

ד. שיטה D והממוצע:

1. על הממוצע כשיטת סיכום במחקר שלנו:

(א) אנחנו משוכנעים כי השימוש בממוצע אריתמטי כדי למצע את כלל התוצאות בניסויים מסוג שאנו עורכים, הוא דבר שגוי באופן עקרוני. באחד במאי 97, כשבועיים לאחר שמקי ושות' העלו את רעיון שיטה B, כתב ריפס אל מקי:

“Experiment B is absolutely unacceptable for me, and let me explain why. This research is oriented towards checking the claim that there is a hidden text in Genesis which is based on ELS’s. We do not know what should be contained in this hidden text, so we make guesses. The input of each guess is a pair of words (w, w') . For each such pair of words we compute some functional $c(w, w')$. The functional $c(w, w')$ was designed as to reflect some intuitive idea (“close meeting between ELS’s”, where “close” is understood for some cylindrical metrics on the text). To have a “small” value of $c(w, w')$ means “success” (a close meeting between ELS’s detected), otherwise “failure”. Now we have to count the number of successes per number of guesses in order to decide whether we encounter a “remarkable” deviation from randomness. Both statistics P1 and P2 do it...

Now what does the procedure of the experiment B? It AVERAGES the values of $c(w, w')$, in other words it punishes the successes for the failures. (For example, I would be very happy to have SYSTEMATICALLY a 1/100 per every 10 guesses; even such an impressive result would be AVERAGED OUT!)”

ב- 16 ביולי 97 ענו מקי ושות' [3], וכתבו בתוך דבריהם:

“... while we acknowledge that (B) does not test for precisely the same phenomenon, it does test for something related...”

וכפי שראינו בפרק הקודם, בעקבות דיון זה ראה מקי צורך לערוך ניסוי אחר שהוא “meets the objection of Rips that ordinary mean ‘averages out’ the alleged ELS phenomenon.”

(ב) נוסף ונדגים מדוע אין זה נכון להשתמש במוצע לגבי המחקר שלנו. ניח לשם הפשטות שלכל אדם ברשימת האישים ישנו כינוי אחד ושלוש צורות תאריך, כלומר שלושה זוגות (שם, תאריך). לפי מקי ושות', צריך לקחת ממוצע של שלושה מספרים עבור כל אישיות. אבל אפשר לעשות זאת אחרת: לערוך את הניסוי 3 פעמים – פעם אחת עבור כל צורת תאריך. מקי ושות' עשו זאת, והרי התוצאות המופיעות ב[1]:

Date Form	List 1	List 2
D M	0.165751	0.000017
bD M	0.000008	0.008844
D bM	0.008488	0.008804

את המובהקות הנובעת מנתונים אלה מחשבים כך: מכפילים את הערך הטוב ביותר פי שלוש. עבור הרשימה הראשונה: $r=0.000024$. עבור הרשימה השנייה: $r=0.000051$.

אבל אם היינו לוקחים ממוצע משלושה מספרים עבור כל רשימה, היינו מקבלים שינוי דרמטי: עבור הרשימה הראשונה: $r=0.058$. עבור הרשימה השנייה: $r=0.0059$.

כמובן, זה אינו דומה למיצוע על כל אישיות בפני עצמה, אבל זה מספיק להדגים את האבסורד שבגישה זאת.

2. מדוע אי אפשר להגדיר הממוצע t :

בפרק הקודם הראינו ששיטה B כלל אינה תואמת את שיטה D. בפיסקה דלעיל הסברנו בצורה עקרונית למה השימוש במוצע אריתמטי אינו מתאים לניסויים שלנו. עכשיו נסביר למה אי אפשר להגדיר $t=average$ לצורך יישום שיטה D. נזכיר: בשיטה D מפעילים את ארבע הסטטיסטיקות P_{1-4} לאחר קביעת הטבלה על ידי t . שתי הסטטיסטיקות P_1 ו- P_3 מודדות כמה ערכי $c(w, w')$ נמצאים בקטע $(0, 0.2)$, ונותנות את ההסתברות לכך. נראה מיד, כי אם הגדרנו $t=average$ איננו יכולים להשתמש ב- P_1 ו- P_3 כדי ליישם את שיטה D. נביא דוגמא:

- יש לנו 10 ערכים כדלקמן: $1/5, 1/5, 1/5, 3/5, 1/5, 4/5, 1/5, 5/5, 1/5, 2/5$. הערך של P_1 (או P_3) הוא הסיכוי לכך שששה ערכים נפלו בקטע $(0, 0.2)$, והוא 0.00636. אבל הממוצע של עשרת המספרים הוא 0.4, והוא נמצא מחוץ לקטע $(0, 0.2)$.

דוגמא מן המדגם השני:

- לגבי אישיות מס' 23 ברשימה ישנם 12 ערכים: $1/125, 4/125, 10/115, 22/125, 19/125, 102/125, 78/115, 124/125, 56/125, 9/125, 5/115, 99/125$. הסיכוי לכך ששבעה ערכים נפלו בקטע $(0, 0.2)$ הוא 0.0039. אבל הממוצע של תריסר המספרים הוא 0.358, והוא נמצא מחוץ לקטע $(0, 0.2)$.

כלומר: אם השערת המחקר שלנו אומרת שיש לצפות להצטברות ערכי $c(w, w')$ בקטע $(0, 0.2)$, והרי זו הגדרת P_1 , אנו מצפים מראש, שהמעבר מערכים בדידים לממוצעים אריתמטיים יגרום לסטיית התוצאות מחוץ לקטע $(0, 0.2)$ לכיוון ה-1. חישוב נכון דורש מציאת הטרנספורם של החסם 0.2, במעבר לממוצע אריתמטי. אבל במקרה שלנו החישוב מסובך עוד יותר, שכן הזזת החסם ימינה משפרת את התוצאה גם עבור הערכים הבדידים. לשיטתנו, אנו משוכנעים שדיאקוניס כלל לא התעניין בפרטי הניסוי אלא רק בעיקרון. אולם מי שיטען שדיאקוניס הציע לקחת $t=average$ במסגרת שיטה D, מטיל על

דיאקוניס חשד מכוער שהוא המציא שיטה מתוך ידיעה מראש שתוליך לכשלון. כי לפי הנתונים ששלחנו לו ניתן היה לראות שבדיקת הצטברות הממוצעים האריתמטיים בקטע $(0,0.2]$ תוליך להעלמות התופעה. אנו לא חושדים בדיאקוניס שהיה מציע ניסוי שמתוך הנתונים **שכבר נמצאים בידו**, היה יכול לדעת שהוא עתיד להיכשל. הנתונים אלה שנשלחו לדיאקוניס לפני שהציע את מבחן הפרמוטציות, היו ערכי $c(w,w')$ עבור כל האישים במידגם השני. מנתונים אלה לבדם אפשר לחשב את הממוצע עבור כל איש ואיש, וכך לדעת כמה ערכי ממוצע יפלו בקטע $(0,0.2]$, ומהו הסיכוי לכך. עבור P_1 הסיכוי לכך הוא: 0.92, ועבור P_3 הסיכוי לכך הוא: 0.82. אם כן, מנתונים אלה ידוע גם בלי מבחן הפרמוטציות שהניסוי ייכשל.

ה. על הטענה שכל ההצלחה שלנו נבעה משיטת הרנדומיזציה:

במאמר [6] שפרסמתי בעיתון מדעי, ובמאמר נילווה באינטרנט [7], הדגמתי כיצד מופעלת שיטת רנדומיזציה אחרת מזו של דיאקוניס על המידגם השני. הכוונה לשיטת RPWL (= Randomization by Permutations of Words' Letters), שהוגדרה לראשונה ב[8] עבור טיפול במידגמי "כותרת", עבורם אי אפשר ליישם מבחן פרמוטציות מן הטיפוס שהציע דיאקוניס.

כל הטענות שהעלה מקי לגבי שיטת הפרמוטציות של דיאקוניס אינן רלוונטיות לגבי RPWL, והמובהקות המתקבלת היא הרבה יותר טובה: $r=0.00000188$. עד היום הזה, מקי ושות' לא התייחסו לתוצאות אלה, ולא ציינו שום חיסרון בשיטת מדידה זו. במאמר שלהם [1] ב-St. Sc., המוגדר כמאמר review, הם מביאים ביקורת על שיטת המדידה המקורית, מביאים רפליקציות שלהם, אבל מתעלמים מרפליקציה חשובה זו וגם מהרפליקציה של מקי עצמו (ראה לעיל פרק ג סעיף 2) הרלוונטיות לדיון בשאלה האם התוצאה המקורית שקיבלנו נובעת אך ורק משיטת המדידה.

ועוד הם מעיזים לכתוב שם:

"Nothing we have chosen to omit tells a story contrary to the story here."

ביבליוגרפיה:

1. המאמר:

B. McKay, D. Bar-Natan, M. Bar-Hillel & G. Kalai, *Solving the Bible Code Puzzle*, Statistical Science, Vol. 14, No. 2, 150-173.

2. המאמר:

B. McKay & G. Kalai, *The Origin of the Permutation Test*, on web site: <http://cs.anu.edu.au/~bdm/dilugim>.

3. התגובה:

D. Bar-Natan, A. Gindis, A. levitan, B. McKay, *The New ELS Tests – A rejoinder*, July 16, 97, at the above web site.

4. הדיווח:

D. Bar-Natan, A. Gindis, A. levitan, B. McKay, *Report on New ELS Tests of Torah*, 29 May 97, at the same web site.

5. התגובה:

E. Rips, *Preliminary Analysis and Comments on the Report of New ELS Tests*, June 19, 97.

6. ד. ויצטום, על הרמז בדילוג השווה: מדידה חדשה של מדגם גדולי התורה, בדד, כתב-עת לענייני תורה ומדע, חוב' 7 – קיץ התשנ"ח, הוצאת אוניברסיטת בר-אילן, רמת-גן. (נמצא באתר במדור פירסומים מדעיים).

7. ד. ויצטום וי. ברמז, על מדידה חדשה של מדגם גדולי חכמי התורה, (כ"ה אלול התשנ"ח, גירסה קודמת – י"ד איר), כאן באתר.

8. ד. ויצטום, א. ריפס וי.רוזנברג, צופן חבוי בדילוג שווה בספר בראשית: מובהקות סטטיסטית של התופעה, פרה-פרינט, אביב התשנ"ו.