

מיסמך 4: ניתוח הבחירות במחקר שלנו.

מבוא:

מאמר ביקורת על המחקר שלנו התפרסם בעיתון "גלילאו" (תשרי, התשנ"ח). מחברי הביקורת, הפסיכולוגית פרופ' בר-הלל, והמתמטיקאים ד"ר בר-נתן וד"ר מק-קי, העלו את שאלת מהימנות המחקר שלנו: האם המחקר המקורי נעשה ביושר ובהגינות. הניתוח שלהם הוא כדלקמן: הרי "כל חוקר בכל מחקר חייב לעשות בחירות מבחירות שונות בנוגע לשיטות המחקר שלו". גם במחקר המקורי היו בחירות כאלה. אם הבחירות נעשו באופן א-פריורי, כלומר מבלי לדעת כיצד ישפיעו על התוצאה, הרי יש לצפות שמספר הבחירות ה"כדאיות" (אלו שמשפרות את התוצאה) יהיה בערך כמספר הבחירות "שאינן כדאיות" (אלו המקלקלות את התוצאה). "והנה, הפלא ופלא, מתברר שכמעט תמיד, אם לא תמיד, הבחירות העיוורות כביכול, השתלמו". מסקנתם: הסיכוי שהמחקר המקורי נעשה ביושר ובהגינות "קלוש ביותר".

אך בחינה מקרוב של הניתוח שלהם מגלה כי הוא נעשה בצורה לקויה ביותר, וכי כאשר עורכים את ניתוח ה"בחירות" שהם העלו, בצורה נכונה – המסקנה בדיוק הפוכה!

ניתוח של "בחירות במחקר" מתוך כוונה לגלות אם היתה הטייה ב"בחירות", חייב לעמוד בתנאים הבאים:

1. מובן מאליו, שהעורך ניתוח מעין זה חייב להזהר בעצמו מהטייה. לכן הוא חייב לרשום לעצמו מראש, באופן א-פריורי, מהן ה"בחירות" אותן יבדוק, לפני שהוא בודק אם הבחירה "כדאית" או לא.

כדי לשכנע במסקנותיו, הוא חייב להציג את כל ה"בחירות" שבדק. אחרת, אין ערך לאמירה "שכך וכך בחירות היו 'כדאיות' וכך וכך בחירות היו לא 'כדאיות'".

2. חשבון ההטיות צריך להעשות רק עבור אותן "בחירות" שאינן נגזרות ממהות המחקר ומצרכיו. למשל, אם גלילאו החליט לברר אם יש ירחים לצדק באמצעות טלסקופ ולא באמצעות רוגטקה, הרי זו בחירה מצורכי המחקר. אם הוא החליט להשתמש לצורך זה בעדשות מזכוכית ולא מגבינה צהובה – אף זו בחירה הנובעת מצורכי המחקר. יהיה מוזר אם יאשימו בהטייה חוקר הסוקר באיזו חברה מיוצר הסטייק המבוקש ביותר, כיוון שבחר שלא לערוך את הסקר שלו בכפר טבעונים.

מתברר, שמחברי הביקורת לא עמדו אפילו באחד מן התנאים הללו. ניתוח "הניתוח" שלהם במלואו יוצג להלן. כאן רק נאמר, שבשתי הזדמנויות הציגו מחברי הביקורת רשימות של "בחירות" שבדקו. בחורף התשנ"ז הציגה פרופ' בר הלל רשימה של 13 בחירות בכנס של המרכז לחקר הרציונליות (של האוניברסיטה העברית). היא טענה כי כל אחת מ-13 הבחירות שלנו היתה "כדאית", אך המתמטיקאי פרופ' ישראל אומן הציג בתשובה ניתוח שונה לחלוטין: מתוך 13 בחירות היתה רק בחירה **אחת** "כדאית". השאר היו או בחירות שנבעו מצורכי המחקר (למשל ה"בחירה" להשתמש בתאריכים נכונים במקום בתאריכים **שגויים!**), או שהיו "לא כדאיות". **בפרק א** להלן נדון בכך בהרחבה. כאן רק נביא שתי דוגמאות לטענות שהועלו גם במאמר הנ"ל. דוגמה לבחירה **לא כדאית**: הבחירה **שלא** להשתמש בצורת התאריך "בא' בתשרי". לצערנו, יעקב אורבך ז"ל (הבלשן שבחר בצורות התאריך) אינו יכול להסביר כעת מה היו שיקוליו בבחירה זו, אך דבר אחד ברור:

אם היינו משתמשים בצורה הזאת, היתה התוצאה **משתפרת** גם עבור הרשימה הראשונה וגם עבור הרשימה השנייה, ולכן **"לא כדאי"** היה לנו להשמיטה. מה מוזר הדבר, שדוגמה זו מובאת במאמר הביקורת להדגמת בחירה **"כדאית"**!

דוגמה נוספת: ברשימה השניה נכללו החכמים שערכם באנצ'י מרגליות משתרע על 1.5-3 טורים. הם טענו, שטעינו בחישוב הטורים ולדעתם היינו צריכים להשמיט ולהוסיף חכמים. התברר, שאם היינו פועלים **לפי הצעתם**, התוצאה היתה **משתפרת**! לפי בקשתו של פרופ' אומן חישבנו גם את התוצאה במבחן הפרמוטציות. התוצאה השתפרה מ-4 למיליון, לאחד מ-5 מיליון!

ד"ר ברנדן מק-קי פירסם באינטרנט ב-3 באפריל 97 דו"ח בדיקה של המחקר שלנו. בדו"ח זה הוא מפרט 20 בחירות אותן בדק (בעיקר בשיטת המדידה). **בפרק ב** נדון בהרחבה בדבריו. כאן רק נאמר, שאפילו אם נתעלם מכך שחלק מן ה"בחירות" היה ממש מצורכי המחקר, מתברר, שלפחות 12 בחירות מתוך ה-20 (לפי חישוביו) היו **"לא כדאיות"** עבורנו. אחת מאלה היתה "לא כדאית" באופן בולט: אם היינו בוחרים קצת אחרת, היתה התוצאה עבור הרשימה הראשונה **משתפרת פי 1,000** והתוצאה עבור הרשימה השנייה היתה **משתפרת ביותר מפי 1,000!**

בסיכום 20 הבחירות התברר כי:

א. 3 בחירות נבעו מהנחת המחקר וצרכיו: 3, 14, 19.

ב. 12 בחירות התוצאה **השתפרה**: 1, 2, 6, 7, 11, 12, 13, 15, 16, 17, 18, 20.

ג. 2 בחירות התוצאה לא השתנתה: 4, 9.

ד. רק ב-3 בחירות היתה הרעה בתוצאה: 5, 8, 10.

הרי לפנינו מאזן המצביע בצורה חותכת על כך שהבחירות שלנו היו אובייקטיביות, א-פריוריות ובלתי מוטות!

מצב מאד לא תקין הוא, שלכל הבדיקות הללו, שהם עצמם ערכו, אין כל זכר במאמר הביקורת! למעיין בפרק ג יתברר עוד, כי בחירות נוספות, עליהן הצביעו מחברי הביקורת בטענותיהם על הכללים של פרופ' הבלין ועל יישומם, נתגלו כ"לא כדאיות" עבורנו.

א.הבחירות של בר-הלל

בכנס של המרכז לחקר הרציונליות שנערך בחורף התשנ"ז, הציגה פרופ' בר-הלל רשימה של 13 בחירות שנעשו לטענתה על ידינו במחקר המקורי. לטענת פרופ' בר-הלל כל 13 הבחירות שהציגה היו לטובתנו. כל בחירה היתה בין שתי אפשרויות, ולכן הסיכוי לבחירה "טובה" הוא 1/2. הסיכוי שזה יקרה במקרה 13 פעמים הוא חצי בחזקת 13, דהיינו, סיכוי של אחד לשמונת אלפים (בערך). ולכן, לטענתה, זה הסיכוי שכל הבחירות הללו נעשו באופן א-פריורי.

כפי שכתבתי במבוא, תנאי הכרחי לניתוח של בחירות הוא, שהחוקר המנתח מציג את כל הבחירות שבדק.

איש אינו יודע כמה בחירות באמת בדקה בר-הלל: על כל פנים, מן הדיווח הלקוי שלה ב"גלילאו" אנו רואים שהיא שוכחת משום מה להזכיר אותן הבחירות המצביעות על כך שהמחקר שלנו היה א-פריורי. ואם כך, כל שכן שעליה להביא ראיה, שרשימת הבחירות שהציגה אכן נבדקה באופן א-פריורי, ואם לא – אין ערך לכל דבריה!

אך המפליא ביותר הוא, שאפילו אם נקבל את רשימת 13 הבחירות שהיא עצמה הציגה, יתברר לנו כי טענה בחושבה כי רשימה זו מוכיחה שהיתה הטייה בבחירות שלנו. הבה נבדוק את 13 הבחירות אחת לאחת.

בחירה 1: לגבי הרשימה הראשונה: כאשר צויין תאריך שאיננו נכון באנציקלופדיה של מרגליות, בחרנו להציב במקומו את התאריך הנכון, במקום להשתמש בניסוי בתאריך השגוי.

תשובה: לפי הנחת המחקר, ביטויים שיש קשר מושגי ביניהם נוטים להיפגש בדילוגים שווים בבראשית. הניסוי נועד לבחון הנחה זו, על ידי בדיקת מפגשים בין שמות אישים לתאריכי הלידה והפטירה שלהם. כמובן שאין קשר בין שם אדם לתאריך שגוי של לידתו או פטירתו!...

זוהי דוגמה מובהקת לבחירה הנובעת מצורכי המחקר (ראה תנאי 2 במבוא). לפי הנחת המחקר אנו מצפים מראש, שהתאריכים הנכונים יצליחו יותר.

בחירה 2: כמו 1, אך לגבי הרשימה השניה.

תשובה: כמו התשובה הקודמת. אך כאן הפלא כפול: סוף סוף בחירה זו כבר נקבעה עבור הרשימה הראשונה, וכפי שכותבת בר-הלל ב"גלילאו" כל הבחירות נשמרו ללא שינוי גם לקביעת הרשימה השניה. כלומר, אין זו כלל בחירה!

בחירה 3: לגבי הרשימה הראשונה: כאשר צויין תאריך שאיננו נכון באנציקלופדיה של מרגליות, בחרנו להציב במקומו את התאריך הנכון, במקום להשמיט את התאריך.

תשובה: גם זו דוגמה מובהקת לבחירה הנובעת מצורכי המחקר: סוף סוף אנו זקוקים לתאריכים כדי ליצור זוגות שמות-תאריכים, ולבחון את הנחת המחקר שלנו!.

בחירה 4: כמו 3, אך לגבי הרשימה השניה.

תשובה: כמו התשובה ל"בחירה" 2.

בחירה 5: לגבי הרשימה הראשונה: אנו בחרנו להשתמש גם בתאריכי לידה וגם בתאריכי פטירה. לדעתה, יכולנו להשתמש רק בתאריכי פטירה.

תשובה: אנו סבורים כי תאריכי לידה ופטירה רלוונטיים באותה מידה, וכי אין מקום להשמיט לא את אלה ולא את אלה. לפי הנחת המחקר, ככל שיהיו יותר זוגות, הסיכוי יגדל שהתוצאות יהיו יותר מוצלחות. אף זו דוגמה מובהקת לבחירה שנבעה מצורכי המחקר.

בחירה 6: כנ"ל לגבי הרשימה השניה.

תשובה: ראה תשובה לבחירה 2.

בחירה 7: לגבי הרשימה הראשונה: אנו בחרנו לציין את היום ה-15 בחודש, ואת היום ה-16 בו בשני אופנים: כ"ט"ו ו"ה", וכ"ט"ז ו"ו". לעומת זאת יכולנו להסתפק ב"ט"ו ו"ט"ז.

תשובה: לפי דרך המניין המקובלת הרי אחרי י"ג, י"ד, היה צריך לבוא י"ה, י"ו. אלא, ששתי הצורות האחרונות בנויות מאותיות השם הקדוש, ולכן נמנעו להשתמש בהן. במקומן השתמשו בט"ו לציין 15 ובט"ז לציין 16.

אך בתורה עצמה נזכר השם הקדוש עצמו פעמים רבות, ואם מחפשים תאריך בדילוג אותיות **בתורה**, אין שום סיבה לשנות את הצורה י"ה לט"ו! לפי הנחת המחקר צריך לבדוק צורה זו, וזו הסיבה שפרופ' ריפס הציע זאת.

בחירה 8: כנ"ל לגבי הרשימה השניה.

תשובה: ראה תשובה לבחירה 2.

סכום ביניים: כל שמונה הבחירות הללו מצטמצמות בעצם לארבע בחירות לגבי הרשימה הראשונה: לגבי הרשימה השניה הן לא היו בגדר בחירות. וארבע הבחירות נובעות **כולן** מהנחת המחקר, ויש רק להתפלא מדוע פרופ' בר הלל התעלמה מכך. נעבור לבחירות הבאות.

בחירה 9: הבחירה שלא להשתמש בצורת התאריך "בא' בתשרי" בצד שלוש הצורות: "א' תשרי" "בא' תשרי" ו"א' בתשרי".

תשובה: בט"ו טבת התשנ"ז, יותר משבועיים לפי הכנס בו העלתה פרופ' בר הלל את טענותיה, היא קבלה מאתנו תשובה בכתב לטענה זו. (ראה מיסמך 2: "בר הלל ובר נתן שואלים – ויצטום וריפס משיבים"). אני מצטט כאן תשובה זו:

"אכן, ברישום התאריכים לא השתמשנו בצורה "בא בתשרי". שמנו לב לכך בעקבות ההערות שלכם ושל אדם נוסף. לצערנו, איננו יכולים לשאול כעת את הבלשן יעקב אורבך ז"ל מדוע לא כלל בהצעתו גם את הצורה הזאת. אך כדי להסיר כל ספק, אנו מצהירים כאן, כי קביעת רישום התאריכים נעשתה ע"י יעקב אורבך, בדיוק כפי שהם כתובים בפרה-פרינט הראשון, עוד לפני ביצוע המבחן לרשימה הראשונה. אנו השתמשנו בתאריכים בדיוק באותה הצורה גם לגבי הרשימה השניה.

אם מישהו חושד, שהיתה כוונה נסתרת בהשמטת הצורה "בא בתשרי", על מנת לשפר את התוצאות, הרי הוא מוזמן לערוך בדיקה, כפי שעשינו לאחר ששמענו טענה (בעל פה) בעניין זה מד"ר דרור בר-נתן.

בדקנו, מה היה קורה אם היינו משתמשים גם בצורה "בא בתשרי".

כזכור, אמת המידה היחידה שלנו להצלחה, גם ברשימה הראשונה וגם בשניה, כפי שפורסם בפרה-פרינט הראשון ("חוברת לבנה") וגם בפרה-פרינט השני ("חוברת כחולה"), וגם בפרה-פרינט ששלחנו בתחילה ל-PNAS, היתה ההסתברויות הכוללות שחישבנו: P1 ו-P2.

מבחן הרנדומיזציה של פרופ' דיאקוניס הוצע בשלב מאוחר יותר.

(i) התוצאות שחושבו למידגם הראשון הן: P1=0.000000001334 P2=0.00000000145

אם היינו משתמשים בצורה "בא' בתשרי" היינו מקבלים

P'1=0.000000000349 P'2=0.00000000207

כלומר, התוצאה הטובה ביותר השתפרה פי 3.8.

(ii) התוצאות שחושבו למידגם השני הן: P1=0.00000000331 P2=0.00000000201

אם היינו משתמשים בצורה "בא' בתשרי" היינו מקבלים

$P^2=0.00000000171$ $P^1=0.00000000507$

כלומר, התוצאה הטובה ביותר השתפרה פי 1.18. "עד כאן הציטוט.

כלומר, מוכח מכאן, שצורות התאריך נקבעו באופן **א-פריורי**, ושהבחירה שלא לכלול את "בא' בתשרי" היתה **לא כדאית**. מוזר מאד, שבר הלל בחרה להמשיך ולטעון את טענתה המופרכת לא רק בכנס, אלא גם חודשים רבים אחר כך – במאמר ב"גלילאו".

בחירה 10: כנ"ל, לגבי הרשימה השניה.

תשובה: כלולה בתשובה הקודמת. מלבד זאת, שים לב שלגבי הרשימה השניה לא היתה זו כלל בחירה (כמו לגבי בחירות 2, 4, 6 ו-8).

בחירה 11: לגבי הרשימה הראשונה: נפלה טעות בהערכתנו את גודל הערך באנצ'י המוקצב לאחד מן האישים ברשימה הראשונה (חסירה שורה להשלמת 3 טורים). פרופ' בר הלל מונה זאת כבחירה ביחס לאפשרות להכין רשימה מתוקנת, ללא טעות כזאת.

בחירה 12-13: הועלתה טענה דומה לגבי הרשימה השניה. אלא, "שפרופ' בר הלל השוותה את מה שעשינו, לשתי אלטרנטיבות מסובכות, ולא הצביעה על האפשרות הפשוטה: להשתמש ברשימה השניה המתוקנת" (מדברי פרופ' ישראל אומן, במכתבו לפרופ' בר הלל מיום ט' בשבט התשנ"ז).

תשובה ל-11-13: בט"ו טבת התשנ"ז, יותר משבועיים לפי הכנס בו העלתה פרופ' בר הלל את טענותיה, היא קבלה מאתנו תשובה בכתב לטענות אלו. (ראה מיסמך 2: "בר הלל ובר נתן שואלים – ויצטום וריפס משיבים"). אני מצטט כאן תשובה זו:

"נשאלנו מדוע נכללו ר' אהרן מקרלין, ר' יהודה עיאש, ור' יהוסף הנגיד ברשימה השניה למרות שערכם באנציקלופדיה מרגליות, לפי דעת השואלים, הוא פחות מטור וחצי. כמו כן נשאלנו מדוע לא נכלל ר' מאיר אייזנשטט ברשימה השניה למרות שערכו לדעתם הוא טור וחצי בדיוק. נשאלנו גם מדוע נכלל ר' דוד גאנז ברשימה הראשונה ולא בשניה, למרות שלדעתם ערכו הוא מעט פחות מ-3 טורים.

התשובה לכך פשוטה: השואלים מדדו את גודל הערך לפי שורות. דורון, שבחר את 34 האישים לרשימה הראשונה, עשה זאת (עד כמה שהוא מסוגל לשחזר) בטביעת עין. וכן עשה אחרי כן בבחירת 32 האישים שברשימה השניה. מתברר, שהיתה זאת טעות לסמוך על טביעת העין, ושהמודד של השואלים טוב יותר.

אבל אם מישהו חושד, שהיה כאן נסיון ליצור הצלחה יש מאין באמצעות מניפולציה, יעיין בבקשה בתוצאות הבדיקה הבאה, שערכנו עם היוודע לנו טענה זאת.

(i) חישבנו את המידגם הראשון גם ללא ר' דוד גאנו. הבה נשווה:

התוצאות שחושבו למידגם הראשון הן: $P1=0.000000001334$ $P2=0.00000000145$

אם מחשבים אותו ללא ר' דוד גאנו מקבלים: $P'1=0.00000001336$ $P'2=0.00000000276$

כלומר, התוצאה הטובה ביותר נעשתה גרועה פי 2.07.

(i i) חישבנו את המידגם השני בהשמטת ר' אהרן מקרלין, ר' יהודה עיאש ור' יהוסף הנגיד,

ובהוספת ר' דוד גאנו ור' מאיר אייזנשטט (השמות והכינויים השייכים לו נמסרו לנו בי"ב טבת

התשנ"ז ע"י פרופ' הבלין, והם: רבי מאיר, איזנשטט, איזנשטאט, מאיר איזנשטט, מאיר

איזנשטאט, מהר"ם א"ש, "פנים מאירות", בעל "פנים מאירות"). לשם השוואה:

התוצאות שחושבו למידגם השני הן: $P1=0.00000000331$ $P2=0.00000000201$

אם מחשבים אותו עם השנויים הנ"ל מקבלים: $P'1=0.00000000422$ $P'2=0.00000000129$

כלומר, התוצאה הטובה ביותר השתפרה פי 1.56.

בסך-הכל, רואים בבירור, שמתקבלות תוצאות מאותו סדר גודל. "עד כאן הציטוט.

רואים בבירור, שהטעות ברשימה הראשונה היתה באמת לטובתנו, אך הטעות ברשימה השניה

היתה **לרעתנו** באותו סדר גודל.

לסיכום: מכל 13 הבחירות שהציגה פרופ' בר הלל, היא הצליחה להראות בחירה **אחת ויחידה**

שהיתה לטובתנו!

הקורא מתפלא: איך יתכן שהיא טעתה כל כך?!

התשובה היא, שהיא טעתה, ולא הבחינה בבחירות הנובעות מהנחת המחקר ומצרכיו.

יתר על כן, היא חישה את ה"כדאיות" שבבחירה לפי שיטת החישוב של פרופ' דיאקוניס,

שהוצעה רק **לאחר** שבוצעו הניסויים הנידונים, וכלל אינה רלוונטית לניתוח הבחירות הללו. סביר

להניח, שאם היתה בודקת אותן כהוגן, לא היינו שומעים כלל על הסיפור של "13 הבחירות"...

אך תמיהה גדולה נוספת עולה מעבודתה זו: כיוון שחישה את הכדאיות לפי שיטת החישוב של

פרופ' דיאקוניס, **היה עליה לדעת** כי התוצאה עבור הרשימה השניה המתוקנת (בחירה 12-13)

משתפרת בצורה ניכרת: במקום הסתברות של 4 למיליון שהתקבלה בניסוי המקורי, מתקבלת

הסתברות של **אחד לחמישה מיליון!** מדוע נשמטה ממנה עובדה חשובה זו הן בכנס במרכז לחקר

הרציונליות, והן במאמר ב"גלילאו"? ומדוע במקום חישוב פשוט זה הציגה "שתי אלטרנטיבות מסובכות, ולא הצביעה על האפשרות הפשוטה: להשתמש ברשימה השנייה המתוקנת" (מדברי פרופ' ישראל אומן, במכתבו הנ"ל).

ב.הבחירות של מק-קי

ד"ר מק-קי מדווח על 20 ווריאציות על שיטת המדידה, בדו"ח שפירסם ביום 22 אפר. 197. הניסויים שערך מסומנים שם ב-V1 עד V20. להלן נדון בכל אחת מן הבחירות (הווריאציות) שנבדקו על ידיו. מטבע הדברים, הדיון בפרטי שיטת המדידה הוא טכני, ומניח שהקורא בקי במאמר שפורסם ב- Statis. Sci. אך לפני שאגלוש לדיון כה טכני, ברצוני להביא דוגמה בולטת במיוחד, פשוטה ומאלפת ביותר:

שיטת המדידה בה השתמשנו, העניקה לכל זוג ביטויים, ערך מספרי בין 0 ל-1. אם המיפגש היה מוצלח, היה ערכו של המיפגש קרוב ל-0. ואם המיפגש היה גרוע, ערכו המספרי התקרב ל-1. כדי להעריך את "הנטייה הכוללת למיפגשים" בין זוגות הביטויים ברשימה הראשונה, השתמשנו בשני מדדים P1 ו-P2. המדדים הללו נועדו לבחון האם ישנם "הרבה" מיפגשים מוצלחים, כלומר, האם ישנם "הרבה" מיפגשים שערכם קטן וקרוב ל-0.

המדד P1 היה פשוט ביותר: הוא ספר כמה מיפגשים היו כה מוצלחים כך שערכם היה בין 0 ל-0.2. זו דרך פשוטה לחשב את ההסתברות. נניח שהכל מתרחש במקרה. אזי, ערכי המיפגשים צריכים להיות מפוזרים באופן פחות או יותר שווה בין ה-0 ל-1, ולכן, מכל 100 ערכים כ-20 צריכים "ליפול" בין 0 ל-0.2. כל זה כאמור, צפוי להתרחש במקרה. יהיה מאד לא סביר למצוא ש-40 ערכים נמצאים בין 0 ל-0.2 במקום 20 ערכים, ואת ההסתברות המדוייקת לכך אפשר לחשב (באמצעות ההתפלגות הבינומאלית).

עבור הרשימה הראשונה היו 152 ערכי מיפגשים. מתוכם 63 ערכים "נפלו" בין 0 ל-0.2, במקום כ-30 ערכים בלבד, שהיו צפויים להמצא שם, אם הכל מקרי. ההסתברות לסטייה כה גדולה, זעירה ביותר: 0.000000001334.

והנה, ד"ר מק-קי טוען (ובצדק) שהסף שלקחנו, דהיינו 0.2, אינו "קדוש", ושהיתה לנו אפשרות בחירה בספים אחרים. הוא בדק בעצמו כמה ספים כאלה.

התברר לפי החישובים של מק-קי עצמו, כי אם הסף היה נקבע 0.33 התוצאה היתה משתפרת פי 1,000, ואם הסף היה נקבע ל-0.5, התוצאה היתה משתפרת פי 3,300! (למעשה, החישובים שלו

לא היו מדויקים, והתוצאה עבור סף של 0.33 היתה משתפרת פי 2,200, ואילו התוצאה עבור סף של 0.5 היתה משתפרת פי 7,500 --- כלומר היינו מגיעים להסתברות של **אחד לחמשת אלפים** (ביליון!)

בבדיקה זו, המסומנת בדו"ח שלו כ- V20 הצליח ד"ר מק-קי להוכיח בצורה נאה, כי אנו עבדנו אך ורק באופן א-פריורי וללא הטיית. אף אחד לא היה יכול לדעת מהו הסף שקבענו מראש, ואם בכלל קבענו אותו מראש. חוקר המבקש לשפר את התוצאות, לא היה מהסס לבחור בסף המשפר את התוצאה פי 7,500 ואינו "מותיר עקבות!" מתברר, שגם לגבי הרשימה השניה היה שינוי הסף גורם שיפור ניכר! ד"ר מק-קי עצמו חישב זאת. (אך כפי שצינתי לעיל, כל הפראמטרים של הניסוי הראשון נשמרו בשני, ולכן אין זה רלוונטי לחשבון הבחירות). ועתה נעבור לבדיקות האחרות. מהשוואות שעשיתי, ברור לי כי המספרים של ד"ר מק-קי אינם מדויקים. אולם לא עומדים לרשותי עוצמת המחשבים והמשאבים העומדים לרשותו, ולכן לא ניסיתי לשחזר את כל התוצאות שלו. בכמה מקרים בודדים (שאצייך) אסתמך על החישובים שלי, אך בדרך כלל אאלץ להסתמך על הנתונים שפירסם בדו"ח שלו. גם כאן, אתייחס לבחירות שהיו בניסוי עבור הרשימה הראשונה, אלא אם כן אציין במפורש אחרת.

בדיקה V1: בניסוי המקורי קבענו סף לגודל הדילוג, שבו מחפשים ELS לביטוי, כך שמספר ה-ELS הצפויים להופיע הוא 10. ד"ר מק-קי חקר מה יקרה אם במקום 10 נקבע 15, 20 וכו'. הוא מדווח כי את הניסוי עבור הערך 20 לא סיים. וחבל. כך היה נוכח לדעת כי התוצאה **משתפרת**, ואם היה ממשיך ובודק עבור הערך 25, היה מוצא שהתוצאה **ממשיכה להשתפר!**

בדיקה V2: כאן בודק ד"ר מק-קי מה היה קורה אם היינו מגדירים אחרת את הפונקציה "דלתה". מתברר, שאם "הקירבה" בין שני ELS היתה מוגדרת **כך: (המרחק בריבוע)/1** (ולא כפי שמופיע במאמרנו), התוצאה היתה **משתפרת בערך פי 30!**

בדיקה V3: אנחנו בדקנו בניסוי המקורי ביטויים שאורכם עד 8 אותיות. הסיבה לכך היתה פשוטה: ערכי המיפגשים נקבעו על ידי השוואה בין מיפגשים בדילוגים שווים לבין מיפגשים בדילוגים שאינם שווים. ביטויים שאורכם 9 אותיות או יותר אינם צפויים להופיע באקראי בספר בראשית, ולכן עבור ביטויים באורך זה צפוי ש"אין עם מי להתחרות" – המתחרים בדילוגים שאינם שווים לא יופיעו. לכן, סף זה כלל אינו שרירותי, ומוצדק מבחינת המחקר (ראה תנאי 2 במבוא). לעומת זאת, מק-קי "חתך" בצורה שרירותית והגביל את אורך הביטויים ל-7 או אפילו ל-6 אותיות. כך קיבל **פחות** זוגות ביטויים בניסוי, וממילא הורעה התוצאה. מק-קי כותב שלא

הספיק לבדוק את המקרה בו לוקחים גם ביטויים בני 9 אותיות. חבל! זו הבדיקה היותר מעניינת.

כך היה מגלה שאין ביטויים כאלה ברשימה הראשונה, המופיעים בדילוג שווה! כך היה מגלה

שלפי ההגדרה שלנו, כל הביטויים ברשימה, המופיעים בדילוג שווה, השתתפו בניסוי!

בדיקה V4: כאן בדק מק-קי פרט טכני אחר: כאשר חישבנו את אורך השורה בטבלת האותיות

עיגלנו 1/2 למעלה. מק-קי בדק מה היה קורה אם היינו מעגלים למטה. התוצאה שקיבל:

התוצאה היתה נשאר ללא שינוי!

(אגב, הוא מדווח שלגבי הרשימה השניה, עיגול למטה היה גורם לתוצאה מעט יותר גרועה. הוא

טעה בחישוב: התוצאה משתפרת!)

בדיקה V5: כאן הוא בודק מה היה קורה אם במקום לסכם על כל המיפגשים (TOTAL)

היינו בוחרים את המיפגש המוצלח ביותר (BEST). הוא מוצא שהתוצאה היתה משתנית לרעה.

בדיקה V6: כאן הוא בודק את האפשרות, שבסיכום על הטבלאות, ילקח כל אורך שורה רק פעם

אחת, ולא כפי שכתוב במאמרנו. מתברר מהחישוב שלו, שהתוצאה **משתפרת** לפי בחירה זו.

בדיקה V7: הוא מצא שלפי התוכנה ששימשה לניסוי המקורי, אורך השורה המינימלי לטבלה

היה 2, לעומת הכתוב במאמר שאינו מגביל אותו ל-2. הוא לא ידע שזה היה פשוט "באג" בתוכנה,

והחשיב זאת כבחירה. לכן, הוא ניסה סף של 1 וראה שהתוצאה **לא משתנה**. הוא ניסה סף של 10

וגילה שהתוצאה **משתפרת!**

בדיקה V8: כאן הוא ניסה הגדרה שונה של תחום המינימליות, וקיבל תוצאה גרועה יותר

במידת מה. (יתכן ואבדוק בעתיד את המספרים שלו).

בדיקה V9: כאן בדק מק-קי פרט טכני שולי ביותר, וגם הוא מודה שאין שום השלכה מעשית

לכך, והתוצאה לא משתנה.

בדיקה V10: כאן הוא בדק כמה אלטרנטיבות לעניין ה"משקל" בו השתמשנו (הוא לא סיים

לבדוק את כולן). אם נאמין למספרים שלו, הרי לגבי הרשימה הראשונה אכן התוצאה הורעה

במקצת (וגבי הרשימה השניה השתפרה).

בדיקה V11: כאן הוא שינה את בחירת קבוצת "המתחרים" בדילוגים שאינם שווים, וקיבל

שיפור לתוצאות (גם לגבי הרשימה השניה).

בדיקה V12: כאן הוא בדק מה יקרה אם נגדיל את קבוצת "המתחרים" בדילוגים שאינם

שווים, וקיבל **שיפור** לתוצאות (גם לגבי הרשימה השניה).

בדיקה V13: כאן הוא בדק מה היה קורה אם בהגדרת הפונקציה "סיגמה", היינו מסכמים לא

עד 10 טבלאות, אלא עד 5. הוא קיבל **שיפור** משמעותי בתוצאות!

בדיקה V14: בניסוי המקורי לא הובאו בחשבון אותם ערכי המפגשים שבהם היו מעט "מתחרים" בדילוגים שאינם שווים – פחות מ-10. היתה לכך סיבה ברורה: תאר לעצמך שקיים מיפגש מוצלח ביותר בדילוגים שווים בספר בראשית: מיפגש בו ה-ELS הנפגשים הם לא רק קרובים מאד, אלא גם "נדירים" – הסיכוי להופעתם כ-ELS במקרה נמוך מאד. בגלל סיכוי נמוך זה, לא יימצאו "מתחרים" בדילוג שאינו שווה: הם פשוט לא יופיעו! ה-ELS ישתתפו לבדם בתחרות, וערכו של המפגש יהיה: $1/1$ (כי מספר המתחרים הצטמצם לאחד). לפי שיטת החישוב שלנו תוצאה זו מורה על כישלון גמור! (זכור, ככל שהערך קרוב ל-0 זו הצלחה, וככל שהוא קרוב ל-1 זה כישלון).

גם אם יהיה עוד מתחרה, יהיה עיוות בלתי נסבל של התוצאה: $1/2$ הוא ערך שאינו מצביע על הצלחה! כדי למנוע עיוות זה קבענו סף של 10 מתחרים. ד"ר מק-קי לא הבין זאת, והציע כבחירה אלטרנטיבית סף של שני מתחרים או חמישה. זו דוגמה לבחירה שנובעת מאי הבנת המחקר. (אגב, מתברר שהשינוי הנגרם בתוצאה --קטן). מעניין לציין, שבחירתו סף של 15 או 20 מתחרים העלתה, שאין שינוי בתוצאה.

בדיקה V15: כאן הוא הציע לנסות והגביל את גודל הדילוג, כך שיהיה לפחות 3. בדיקתו העלתה שהתוצאה **השתפרה!**

בדיקה V16: כאן בדק ד"ר מק-קי אפשרות אחרת לחישוב המרחקים עבור ההופעות בדילוג שאינו שווה. גם כאן **השתפרה** התוצאה.

בדיקה V17: כאן הוא בדק עניין הרלוונטי דווקא לרשימה השניה: מה היה קורה אם במבחן הפרמוטציות של דיאקוניס, היינו משמיטים את כינויי האישים שאין להם תאריך. (בניסוי המקורי לא השמטנו אותם). הוא קיבל שהתוצאה **השתפרה!**

בדיקה V18: כאן בדק ד"ר מק-קי מה קרה אם יוסיפו את צורת התאריך "בא' בתשרי". גם כאן התברר, שהתוצאה **השתפרה** (גם לגבי הרשימה השניה)! (ראה לעיל תשובה לבחירה 9 של בר-הלל).

בדיקה V19: כאן הוא בדק מה היה קורה אם לא היינו כוללים בתאריכים את המיספור $1/1$. ראה לעיל, תשובה לבחירה 7 של בר-הלל, שזו בחירה הנובעת מהנחת המחקר.

בדיקה V20: בדוגמה זו פתחנו את סידרת הבדיקות של מק-קי.

סיכום 20 הבדיקות

מתוך 20 הבדיקות התברר:

א. 3 בחירות נבעו מהנחת המחקר וצרכיו: 3, 14, 19.

ב.ב- 12 בחירות התוצאה **השתפרה**: 1, 2, 6, 7, 11, 12, 13, 15, 16, 17, 18, 20.

ג.ב- 2 בחירות התוצאה לא השתנתה: 4, 9.

ד. ב- 3 בחירות היתה הרעה בתוצאה: 5, 8, 10.

הרי לפנינו מאזן המצביע בצורה חותכת על כך שהבחירות שלנו היו אובייקטיביות, א-פריוריות ובלתי מוטות!

א.בחירות אחרות שהועלו על-ידי מבקרינו

פרק זה נועד לקבץ בחירות אחרות שהועלו ע"י מבקרינו, והוזכרו במיסמכים הקיימים באתר. כעת אין שעת פנויה לכך, ולפי שעה אפשר למצוא את החומר במקומו במיסמכים הנ"ל.