

## צירופי-מקרים וסטטיסטיקה בתורה ובלשון הקודש\*

רופ' חיים שור

אוניברסיטת בן-גוריון בנגב

### 1. מבוא

"כך היה הקב"ה מביט בתורה ובורא את העולם" (מדרש רבה, בראשית, פרשה א);

"יודע היה בצלאל לצרף אותיות שנבראו בהם שמים וארץ" (ברכות, נה, א);

"יודע מהספרות הקבלית, שאותיות הא"ב נבראו תחילה. לאחר מכן, על ידי שימוש באותיות, ברא הקב"ה את כל העולמות, שנאמר "בראשית ברא את" (אור תורה, הרב דוב בער, המגיד ממזריץ').

עיון במקורותינו מגלה שוב ושוב דבר מפתיע: אבותינו האמינו כי קיים קשר הדוק בין העולם הפיזי המוכר לנו ובין מילות שפת הקודש. אמונה זאת באה לידי ביטוי לא רק באמירות כלליות כמו אלו שציטטתי בפתיח אלא במאמצים שונים, פזורים במקורות יהודיים שונים, להפיק מידע ממשי על העולם הפיזי מניתוח המבנה והערך המספרי של מלים בעברית. ייאמר מייד: אין הכוונה כאן לגימטריה. האחרונה, מתייחסת להנחה, שהייתה מקובלת על חז"ל ועל פרשני מקרא יהודיים, שאם ערכים מספריים של שתי מילים שונות הם שווים - כנראה שיש "סוד" משותף המאחד אותם, ותפקיד הגימטריה לגלות ההיבט משותף. אולם עיון במקורותינו מגלה שהייתה אמונה נוספת, נפרדת ובלתי קשורה לגימטריה, והיא שמילות לשון הקודש מספקות מידע נסתר על הנושא הגלוי בו עוסקת

\* מבוסס על הספר:

Shore, Haim (2007, 2008; 2nd Edition Dec., 2012). Coincidences in the Bible and in Biblical Hebrew. iUniverse, Bloomington, IN 47403, USA.

את בטחוני המוחלט כי אמונת חז"ל היא אכן אמונה בלבד. אביא שתי דוגמאות נוספות אליהן נחשפתי. אין בהן כדי לספק "הוכחה מדעית" אך הן מעניינות כשלעצמן ובהצטרפן לאחרות אולי יש בהן כדי לעודד חקירה יותר מעמיקה. הדוגמה הראשונה דומה לדוגמת ה"הריון" שנתנה קודם. כך גם המילה "שנה" בערכה המספרי, 355, שווה בדיוק לאורך השנה העברית (שנה המבוססת, כידוע, על מחזור הירח, ושארורה הממוצע הוא 354.4 יום). דוגמה שנייה היא המילה "אוזן" הנגזרת מאותו שורש כמו המילה "איזון", אם כי התגלית שאיזון הגוף היא פונקציה מרכזית של האוזן נמצאה רק בשלהי המאה ה-19.

כאמור, דוגמאות אלו ואחרות משכו את תשומת לבי במשך שנים רבות, ואם כי תחילה התייחסתי אליהן כקוריוז שאין לו כל בסיס מדעי עם השנים בטחוני שהדוגמאות מייצגות רק "מקרים של מקריות" (צירופי מקרים) התחיל להתערער. בקיץ 2006 הגעתי למסקנה שמספר הדוגמאות שאספתי במשך עשרות שנים עבר "מסה קריטית" מסוימת אשר מצדיקה העלאתן בכתב. תוצאות מסקנה זאת הוא ספר שיצא לאור באנגלית בארה"ב בחורף של 2007, בשם: "מקרים של מקריות במקרא ובלשון המקרא". השם שבחרתי לספר נועד להצביע על עמדותי ביחס לרב הדוגמאות המובאות בו: צירופי מקרים, ללא הוכחה מדעית, אך בכל זאת במספר כזה שאולי מחייב התייחסות קצת יותר רצינית.

במהלך כתיבת הספר קרה דבר נוסף: הרקע שלי בסטטיסטיקה יישומית ובמידול סטטיסטי האיצ בי לבצע בדיקה יותר מעמיקה אחת ולתמיד האם בערכים המספריים של מלים בלשון הקודש חבוי רובד נוסף של מידע, הרלבנטי לנושא הגלוי של המילה. הנחתי שניתן לגלות רובד זה על ידי ניתוח סטטיסטי מתאים. לא היו אלו מחשבות בעלמא. היה לי מושג ברור למדי מה אני רוצה לבדוק ומה היא הגישה הסטטיסטית הנכונה לבדיקה. מיותר לציין שאם ניתוח סטטיסטי מגלה שהנתונים המנותחים הם בעלי מובהקות סטטיסטית, והמובהקות היא ברמה המקובלת במדע, הרי שאותם "מקרים של מקריות" מפסיקים להיות כאלו ובפועל נוצר **בסיס מדעי מוצק** לאמונה שקיים מידע חבוי במילות לשון הקודש. להפתעתי הרבה, התגלה לי כי **בלשון הקודש כנראה חבוי רובד של מידע נסתר**, שמספק מידע נוסף על האובייקט שהמילה מייצגת. התוצאה הסופית של הבדיקות הסטטיסטיות שבצעתי היו למעלה מעשרים ניתוחים סטטיסטיים, כולם מובהקים סטטיסטית, המובאים בספר תוך המחשות גרפיות המאפשרות הבנה גם למי שאינו אומן על המתודולוגיה הסטטיסטית. בדצמבר 2012 יצאה מהדורה שנייה של הספר שבו הובאו ניתוחים סטטיסטיים נוספים, עם מתודולוגיה המבוססת על ביצוע סימולציה ממוחשבת.

במאמר זה בכוונתי לפרט, בלשון לא סטטיסטית ומובנת לכל נפש, מהי גישת הניתוח בה נקטתי ולהביא מספר דוגמאות בולטות לממצאים שהתגלו באותם ניתוחים. אין בכוונתי

לשכנע את הקורא כי הממצאים מהווים בסיס מדעי מספיק לאמונת חז"ל כי בלשון הקודש חבוי רובד של מידע נסתר. בכוונתי רק לפתוח וויכוח. במאמר זה אציג מספר ממצאים (המובאים בפירוט רב יותר בספרי), וישפוט הקורא היש בהם כדי לבסס את אותה אמונה, חלק מהותי ובלתי נפרד מהמסורת היהודית.

בסעיף 4 מובא תיאור קצר של שיטת ניתוח סטטיסטי שנקטה ביחס לסיפור הבריאה בספר בראשית כדי לעמת סיפור זה עם ממצאי המדע בן-ימינו.

### 2. הסבר הגישה

נשער שבידינו שתי קבוצות של נתונים. קבוצה אחת מייצגת מדידות של טמפרטורת הסביבה בסולם צלזיוס שנלקחו במהלך מספר ימים (נניח שישנן N מדידות). בקבוצה השנייה אותו מספר של נתונים (N ערכים מספריים), אך לא ידוע מה הם מייצגים. מדען מסוים טוען כי הנתונים בקבוצה השנייה הם מדידות טמפרטורה שנלקחו במהלך אותם ימים כמו בקבוצה הראשונה (כלומר, באותו יום נמדדה אותה טמפרטורה אך בשני מכשירי מדידה נפרדים וכנראה גם בסקאלות שונות). כיצד נבדוק את טענתו של המדען? התשובה היא פשוטה: אם הקבוצה השנייה מייצגת אותן מדידות טמפרטורה, אך בסקלה אולי שונה, למשל, פרנהייט, אזי אם נסדר את שתי הקבוצות בסדר עולה, נקבל N זוגות של נתונים. עכשיו נצייר את הנתונים על פני גרף, כאשר בציר האופקי נרשום את הערך מהקבוצה הראשונה ובציר האנכי נרשום את הערך המקביל מקבוצת המדידות השנייה. יתקבלו N נקודות בסך הכל (כמספר הנתונים בכל קבוצה). אם הנקודות יסתדרו על קו ישר, אזי ברור לכל שנתוני קבוצת המדידות השנייה (שמשמעותם, כאמור, אינה ידועה) מייצגים אותן טמפרטורות שנמדדו בקבוצה הראשונה, ואולם בסקלה אחרת.

מה החשיבות של העובדה שהמדידות יסתדרו על קו ישר? נזכר כיצד עוברים מסקאלת צלזיוס לסקאלת פרנהייט:

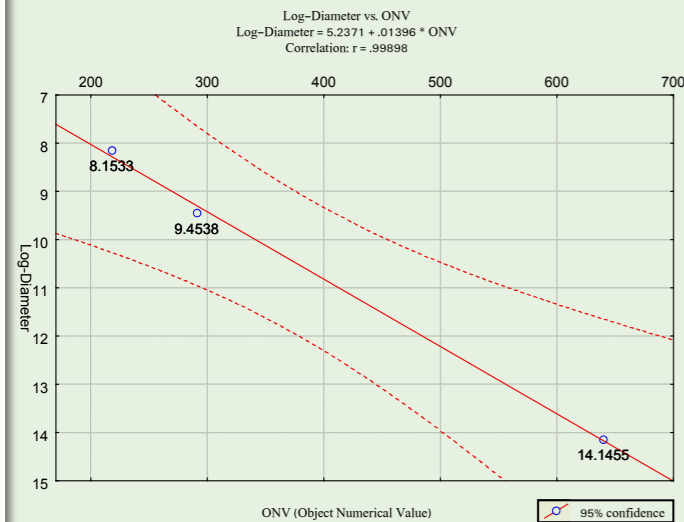
**מדידה בסקאלת פרנהייט = (מדידה בסקלה של צלזיוס) \* 1.8 + 32**

כלומר, משוואה של קו ישר. מסקנה זאת נכונה באופן כללי: מעבר מסקלה לסקלה של מדידה פרושה טרנספורמציה ליניארית. על כן אם מדידות מסוימות נמדדות בשתי סקאלות שונות, הצגתן בגרף תמיד תיתן קו ישר. המסקנה ההפוכה תקפה גם היא: אם קבוצה של זוגות של נתונים, שצוירו בגרף, תראה קשר ליניארי, ניתן להניח כי הנתונים מייצגים מדידה של אותו דבר, אם כי בסקאלות שונות. מסקנה זאת ניתן להחיל כדי לבדוק באופן מדעי את אמונת חז"ל, כפי שהתייחסתי אליה קודם. בדוגמאות שנביא להלן נבדוק קבוצות של מלים בעברית הקשורות ביניהן במובן מסוים, ונגדיר תכונה פיזיקאלית עיקרית בעלת ערך מספרי מדיד ביחס לכל אחד מהאובייקטים המיוצגים על ידי המלים בקבוצה. נציג את הנתונים המתקבלים

**דוגמה שנייה: קוטר של "ירח", "ארץ", "שמש"**  
 בדוגמה זאת אנו בודקים את אמירת חז"ל כי במילות העברית חבוי מידע על עולמנו, ביחס לתכונה פיזיקאלית מרכזית המבדילה בין שלושת העצמים השמימיים שבקבוצה: גודלם, כפי שהדבר מתבטא בקוטר. כמו בדוגמה הקודמת, עקב הבדלים של סדרי גודל בין הקטרים של הירח, הארץ והשמש, יירשמו הקטרים בסקלה של לוג. טבלה 2 מציגה את הנתונים (הקטרים נלקחו מהאתר של סוכנות החלל האמריקאית, NASA), וציור 2 מציג אותם כאשר על הציר האופקי מיוצג הערך המספרי של המילה ואילו הציר האנכי מייצג את הקוטר (כאמור, בסקאלה של לוג).

מילה	ONV Object Numerical Value	קוטר (km)	קוטר (בסקלת לוג)
ירח	218	3474.8	8153292
ארץ	291	12756.28	9453779
שמש	640	1,391,000	14.14553

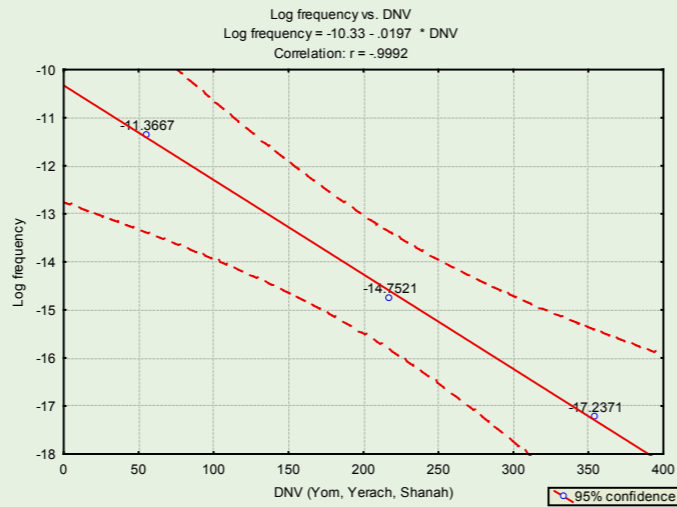
טבלה 2. נתונים עבור הדוגמה השנייה (קטרים של "ירח", "ארץ", "שמש"). ONV מייצג את הערך המספרי של המילה



ציור 2. הקוטר של "ירח", "ארץ", "שמש" (בסקאלת לוג, בציר האנכי) כפונקציה של ONV, ערכי המלים בעברית (על הציר האופקי).

התופעה שנצפתה בדוגמה הקודמת חוזרת על עצמה: שלושת הנקודות מסתדרות על קו ישר עם מתאם ליניארי של 0.999 (אם הקשר בין הערכים היה קשר ליניארי מתמטי היינו מקבלים מתאם של 1!). המובהקות המתקבלת עתה היא באותו סדר גודל כמו בדוגמה הראשונה (2.9%).

ציור 1 מציג את שלוש הנקודות שערכיהן מובאים בטבלה 1.



ציור 1. מחזוריות (בסקלת לוג) של "יום", "ירח", "שנה" כפונקציה של ערכי המלים בעברית (על הציר האופקי).

בציר האופקי נרשמו הערכים המספריים של המלים בעברית, ואילו על הציר האנכי נרשמה המחזוריות של ה"ישות" המיוצגת על ידי כל מילה, בסקלה של לוג (לפי בסיס טבעי). אנו מוצאים כי הנקודות מסתדרות על קו ישר. בפועל, אם הנקודות היו מסתדרות בקו ישר מתמטי המתאם שהיה מתקבל היה -1. המתאם שקבלנו הוא -0.9992 (!). המובהקות הסטטיסטית המתקבלת היא 2.5%. בלשון עממית ניתן לומר כי ההסתברות שהנקודות יסתדרו באופן מקרי, כפי שהסתדרו בפועל, היא הסתברות קטנה מאד (רק 2.5%). בדרך כלל, קבלת ערכי מובהקות נמוכים מ-5% מקובלת במדע כמעידה על חוסר מקריות (בלשון יותר מדעית, על כך שתחת השערה ההפוכה להשערה שאותה אנו בודקים, דהיינו השערת המקריות, ההסתברות לקבלת התוצאות שקבלנו בפועל היא רק 2.5%). באמצעות מתודולוגיה סטטיסטית שונה, המתבססת על סימולציה ממוחשבת, נמצא כי ההסתברות לקבלת הנקודות בקו ישר במקרה היא אפילו יותר נמוכה, 0.5% (טבלה 21.2 במהדורה החדשה של הספר).

מהבחינה המדעית הטהורה אפוא אפשר להטיל ספק כי הסתדרות הנקודות בקו ישר היא מקרית. יצוין, כי לא כל הניתוחים בספר מצטמצמים לשלוש נקודות בלבד (כמו בשתי הדוגמאות הראשונות). חלקם מתייחס לקבוצות שכוללות עד 7 תצפיות! הדוגמה השלישית שנביא להלן תדגים את התופעה שאנו מתייחסים אליה כאשר במדגם הנתונים בסך הכול ארבע תצפיות.

### 3. שלוש דוגמאות

**דוגמה ראשונה: מחזוריות של "יום", "ירח", "שנה"**

(Day, Month, Year)

הייתכן שהערכים המספריים של קבוצת מלים זאת מייצגים את המחזוריות הקשורה באותן מלים? (מחזוריות, או תדירות (frequency)), היא התכונה הפיזיקאלית המרכזית המבדילה בין ה"ישויות" המיוצגות על ידי המלים).

כדי לבדוק את הדבר, עלינו לבטא את המחזוריות של כל ה"ישויות" ביחידת מדידה זהה. למשל, אם היינו בוחרים כי היחידה תהיה "מחזורים לשנה" אזי המחזוריות של יום תהיה:  $354.37 = (29.53059 * 12)$  (בממוצע אורך חודש הירח הוא 29.53059 יום), המחזוריות של חודש תהיה 12 ושל שנה 1. אם היינו בוחרים ב"יום" כיחידת המדידה אזי מחזוריות השנה תהיה  $1/354.37$  ליום. בפועל החלטנו להשתמש ביחידת מחזוריות המקובלת במדע והיא "הרץ" (מחזור אחד לשנייה). נשים לב, כי אבותינו חלקו את השעה ל-1080 חלקים (בעוד אנו מחלקים ל-3600 שניות). לבחירת היחידה בה נשתמש לצורך הניתוח הסטטיסטי אין כל חשיבות מבחינת התוצאות המתקבלות כל עוד נעשה שימוש עקבי (קונסיסטנטי) ביחידות השונות. טבלה 1 מציגה את הערכים המספריים של המלים השונות ואת המחזוריות הקשורה בכל מילה.

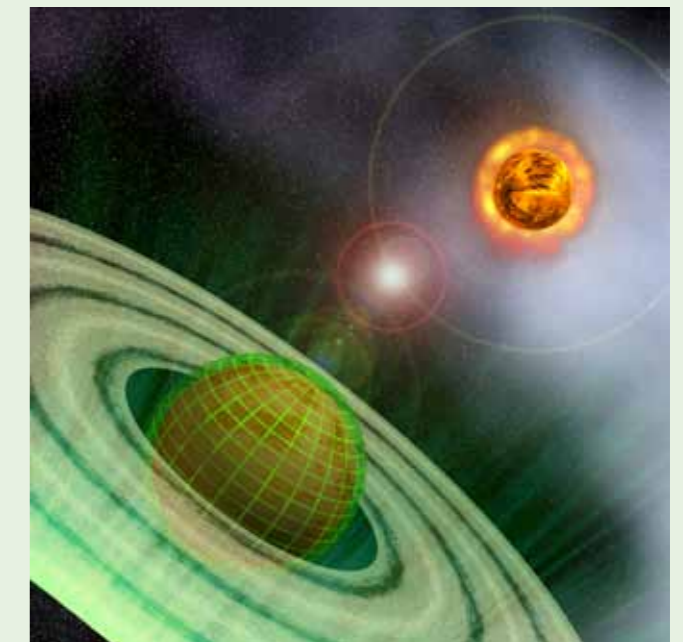
מילה	DNV (Date Numerical Value)	תדירות (מחזורים לשנה)	תדירות (בסקלת לוג)
יום	56	1.1574E-05	-11.3667
ירח	218	3.9194E-07	-14.7521
שנה	355	3.2661E-08	-17.2371

טבלה 1. נתונים עבור הדוגמה הראשונה (מחזוריות של "יום", "ירח", "שנה"). DNV מייצג את הערך המספרי של המילה. יחידת המדידה היא הרץ (מחזור לשניה).  $E-05$  משמעותו:  $10^{-5}$ .

שימו לב כי ערך מספרי כל שהוא יכול לקבל ייצוג על ידי המספר עצמו במערכת כלשהיא (למשל, המערכת העשרונית, כפי שאנו רגילים), או להיות מבוטא כחזקה. למשל: את המספר 7 ניתן לרשום בשני אופנים:  $7 = 10^{0.845}$  המספר 0.845 נקרא הלוג של המספר 7 לפי בסיס 10 (אני משחזר כאן חומר מלימודי ביה"ס התיכון ועם הקוראים הסליחה...). בפועל, כאשר ערכים מספריים של מדגם נתונים משתרעים על פני מספר סדרי גודל נהוג לייצגם, לצורך מידול סטטיסטי, על ידי ערכי לוג שלהם. כך גם נעשה כאן ביחס לשתי הדוגמאות הראשונות.

בגרף, כאשר על הציר האופקי יוצג הערך המספרי של המילה ועל הציר האנכי הערך המספרי של התכונה הפיזיקאלית, כפי שידועה ביחס לאובייקט המיוצג על ידי המילה. נבדוק אם הנקודות שהתקבלו מסתדרות על קו ישר (כל זוג תצפיות ייוצג על ידי נקודה בגרף). כיוון שהקשר האפשרי בין ערך מילה בעברית והתכונה הפיזיקאלית של האובייקט שהיא מייצגת אינו קשר מתמטי (בניגוד לדוגמת ההתמרה של טמפרטורה מצלזיוס לפרנהייט), נזדקק לניתוח סטטיסטי שיקבע האם קירבת הנקודות המצוירות לקו ישר מספיק טובה כדי להניח שהסתדרות הנקודות בגרף אינה מקרית אלא היא מייצגת קשר ליניארי אמיתי בין המשתנה המיוצג על הציר האופקי לזה המיוצג על הציר האנכי. ניתוח סטטיסטי שיראה כי התכונה הפיזיקאלית מתייחסת באופן ליניארי לערכים המספריים של המלים המקבילות בעברית יוביל למסקנה ששני הצירים מייצגים אותה "ישות" ובפועל "מודדים" אותו דבר.

ואם יימצא כי התופעה הזאת חוזרת שוב ושוב בשפת המקרא הרי שאמונת חז"ל, כי חבוי רובד נוסף של מידע במילות שפת הקודש (אמונה שהדגמתי קודם ממקורותינו), אמונה זאת יש לה כנראה על מה שתבסס גם מהבחינה המדעית. בספרי מובאים למעלה מ-20 ניתוחים סטטיסטיים מבוססים על הגישה שהצגתי. להלן נביא שלוש דוגמאות.



באמצעות מתודולוגיה סטטיסטית שונה, המתבססת על סימולציה ממוחשבת, נמצא כי ההסתברות לקבלת הנקודות בקו ישר במקרה היא אפילו יותר נמוכה, 0.2% (טבלה 21.2 במהדורת החדשה של הספר).

נציין, כי המובהקות המתקבלת בניתוחים אחרים שבוצעו עבור מדגמים גדולים יותר (מספר גדול יותר של תצפיות, למשל, N=7) היא נמוכה הרבה יותר, מה שנותן משנה תוקף למסקנה שאולי עולה מכלל הניתוחים שבוצעו.

### דוגמה שלישית: צבעים בעברית

הייתכן כי הערכים המספריים של שמות צבעים בעברית מייצגים את תדירות הגל של אותם צבעים? לצורך הבדיקה בחרנו ארבעה שמות צבעים המופיעים במקרא שלגביהם אין ספק אלו צבעים הם מייצגים: צהב, ירוק (ירקון, בשפת המקרא) וארגמן. נעיר מיד שהחלטה אם לכלול האות ו' במלה "צהב" נקבעה לפי השכיחות הגבוהה ביותר במקרא, והכללת ה-ו' במילה או השמטתה לא תשנה את משמעות תוצאות הניתוחים הסטטיסטיים שנביא מיד. כמו כן, המילה "ירוק" (במשמעותה המודרנית) אינה מופיעה כלל במקרא. המילה היחידה הקרובה לתאר צבע זה היא כנראה "ירקון" (ירמיהו ל, ו). לדעת מרבית הפרשנים היהודיים (כמו מצודת דוד ורד"ק) משמעותה היא הצבע "ירוק". נעיר, כי השימוש בצבע "ירוק" לתיאור חיוורון, כמו במקור המוזכר מירמיהו, מקובל גם בשפות אחרות, למשל באנגלית (כשאומרים "פניו הפכו ירוקות" מתכוונים לתיאור חיוורון). המשמעות המודרנית של ארגמן היא המילה המקבילה למילה "Purple" באנגלית, אך גם המילה הקרובה מבחינת הגוון, "Magenta", יכולה להתאים. להלן נניח כי זאת היא המשמעות של המילה ארגמן המופיעה במקרא (ראה פרטים על ההיגיון שבבחירה זאת בספר). שוב נציין כי בחירת הגוון השני לא תשנה באופן משמעותי את תוצאות הניתוחים שלהלן.

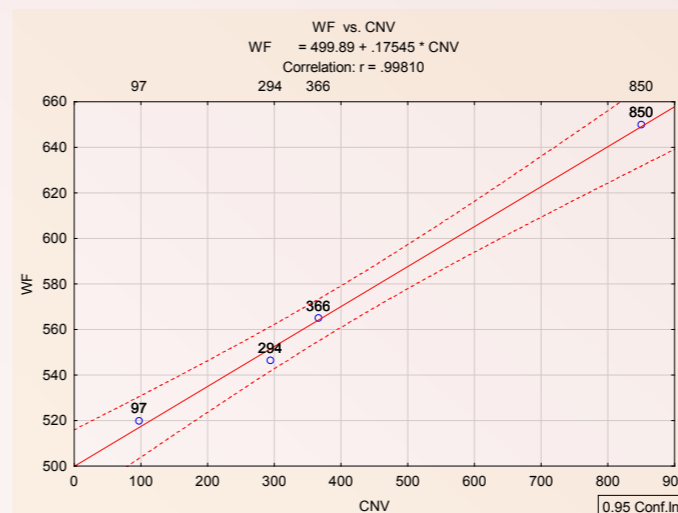
טבלה 3 מציגה את תדירות הגל (ביחידות של טרה-הרץ,  $10^{12}$  הרץ, כאשר הרץ זהו מחזור לשנייה). הערכים המספריים של המילים בעברית מוצגים גם הם בטבלה. כיוון שלכל צבע יש טווח אופייני של תדירויות, בחרנו (שרירותית) את אמצע הטווח כערך מייצג. כמו כן, בעוד שירוק, צהב ותכלת הם צבעי יסוד (צבעים אלמנטריים), עם תדירות ברורה, לצבע ארגמן אין תדירות משלו כיוון שהוא נוצר מערבוב של הצבעים אדום וכחול. התדירות של ארגמן בטבלה מייצגת "תדירות אקוילונטית", שחושבה על סמך שיקולי אנרגיה של תדירויות האדום והכחול (ראה פירוט בספר, סעיף 12.3.3). הדבר הבולט מהתבוננות בטבלה הוא שערכי המילים בעברית, כאשר מסדרים אותם בסדר עולה, מסתדרים באותו סדר כמו ערכי תדירות הגל המתאימים.

ציור 3 מציג את הנקודות כאשר על הציר האופקי ערכי המילים בעברית ואילו על הציר האנכי התדירויות (ביחידות של THz).

שוב מתקבלת בבירור תופעת הסתדרות הנקודות על קו ישר (עם מובהקות סטטיסטית של 0.2%).

צבע	WF (THz)	CNV
צהב	520	97
ארגמן	546.5	294
ירקון ("ירוק" במקרא)	565	366
תכלת	650	850

טבלה 3. תדירויות גל (WF, ביחידות THz) של "צהב, ארגמן, ירקון, תכלת" וערכי המילים בעברית (CNV-Color numerical value)



ציור 3. תדירויות הצבעים "צהב, ארגמן, ירקון, תכלת" (WF) כפונקציה של ערכי המילים בעברית (CNV).

### 4. ניתוח סטטיסטי של הקשר בין סקלת הזמן של סיפור הבריאה בספר בראשית וסקלת הזמן הקוסמית-מדעית

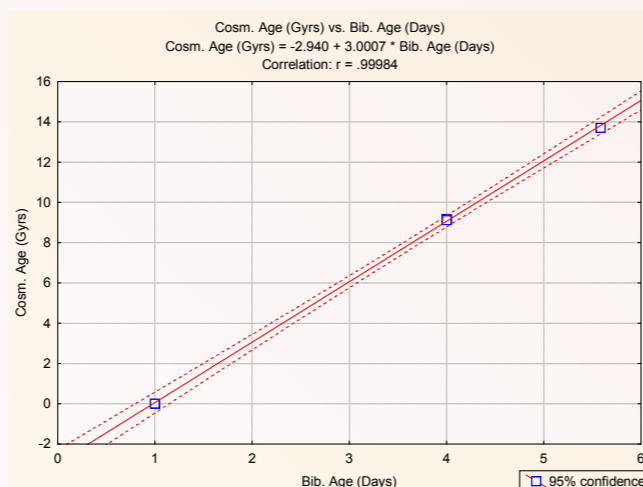
אחד הניתוחים המעניינים ביותר המובאים בספרי הוא ניתוח הקשר בין סקלת הזמן של התורה וזאת של המדע המודרני. ניתוח זה מובא במלואו בפרק 22 בספרי וכאן אביא ניתוח חלקי המתבסס על חלק ממדגם הנקודות המופיעות ונדונות שם. הרעיון המרכזי של הניתוח אינו שונה מזה בדוגמאות הקודמות: אם סיפור הבריאה בספר בראשית מציג סקלת זמן תואמת את הסקלה המדעית הידועה לנו כיום, אז הצגת אירועים המופיעים בסיפור הבריאה בשתי סקלות הזמן תייצר גרף בו סידרה של נקודות המייצגות אותם אירועים מסתדרת על קו ישר.

לצורך הניתוח המובא כאן הסתפקנו בארבעה אירועים (ארבע נקודות) שאין לגביהם ספק מבחינת ההתייחסות בפרק א' בספר בראשית ומבחינת סקלת הזמן המדעית הידועה היום: \* **יצירת האור**; \* **גיל השמש**; \* **גיל הירח**; \* **"בריאת האדם"**. לגבי שלושת האירועים הראשונים:

**יצירת האור** – לפי המדע, זמן קצר לאחר המפץ הגדול נוצרה קרינה מורגשת (כאשר מהפלסמה הראשונית החלו להיווצר אטומים "מחזירי אור", כ-380,000 שנה אחר המפץ הגדול שהתרחש לדעת המדענים לפני 13.7 מיליארדי שנים). האור נוצר, לפי תהליך הבריאה שבתורה, ביום הראשון.

**יצירת השמש והירח** – ההערכה המדעית כיום היא שגיל השמש והירח הם, בהתאמה, 4.57 ו-4.53 מיליארדי שנים. "המאורות הגדולים" נוצרו ביום הרביעי.

**"בריאת האדם"** – לפי הכתוב בתורה, בריאת האדם התרחשה לפני כ-6000 שנה. (ואפילו לפי תאוריית האבולוציה המדברת על פרק זמן ארוך יותר, על ציר הזמן הקוסמי מזמן המפץ הגדול הנמדד במיליארדי שנים לא יראה כמעט כל הבדל. לכן אין כאן חשיבות למספר המדויק).



ציור 4. ארבעה "אירועים" המוזכרים בפרק א' בספר בראשית בסקלה מקראית (הציר האופקי; בימים) ובסקלה הקוסמית-מדעית (הציר האנכי; מיליארדי שנים).

ארבעת ה"אירועים" הם: "בריאת האור", "בריאת השמש", בריאת הירח, "בריאת האדם". שתי הנקודות הקשורות בבריאת "המאורות הגדולים" ביום הרביעי מופיעות כנקודה אחת (עקב קירבתן). המיתאם הליניארי מתקבל: 0.9998.

מבחינת מועד הבריאה לפי ספר בראשית, מסורת בידינו כי האדם נברא ביום שישי בסוף השעה ה-14. הדבר נקרא וי"ד, ומשמש בסיס לחישוב אורך החודש העברי בדיוק של עד 5

ספרות אחרי הנקודה העשרונית; ראו פירוט בפרק 18 בספרי. תיאור של ארבע הנקודות, כאשר הציר האופקי הוא ציר הזמן לפי "סיפור הבריאה" והציר האנכי הוא ציר הזמן המדעי (ביחידות של מיליארדי שנים) מוצג בציור 4. אנו רואים כי הנקודות מסתדרות על קו ישר ממש (עם מיתאם ליניארי של 0.9998!!). ההסתברות לקבל את התוצאות שהתקבלו (דהיינו, על קו ישר) במקרה היא 0.0165% (ראו סעיף 22.4.2 בספרי). התוצאה המעניינת ביותר מופיעה בנוסחה שבכותרת הגרף, לפיה "יום" לפי סיפור הבריאה בספר בראשית שווה ערך ל-3.0007 מיליארדי שנים (שימו לב כי המספר השלם כמעט, עם שלושה אפסים לאחר הנקודה העשרונית, התקבל מהמחשב כתוצאה של ניתוח סטטיסטי של נתונים!!).

### 5. ניתוח לשוני-סטטיסטי של ספר בראשית לבחינת "השערת המקורות"

בשנת 1985 התפרסם באנגלית ספר משותף\* לפרופ' רדאי ז"ל ושלי שהציג ממצאי המחקר שביצענו בדבר האחידות הלשונית של ספר בראשית. הנחת היסוד של מחקר זה, כמו של מאמרים קודמים שפרסמנו במשותף, הייתה כי לכל מחבר של טקסט ספרותי "טביעת אצבעות" לשונית, וכי ניתן לאפיין את התבנית הלשונית של מחבר על ידי מדדים כמותיים לשוניים. באמצעות ניתוח סטטיסטי מתאים אפשר לבחון האם טקסטים, שנכתבו לכאורה על ידי אנשים שונים, הם למעשה ממוצא משותף, או להיפך, טקסטים שלכאורה נכתבו על ידי אותו מחבר הם בפועל כל כך שונים מבחינת הסגנון הלשוני שכנראה נכתבו על ידי אנשים שונים.

דוגמאות למדדים סגנוניים כמותיים של טקסט לשוני הוא מספר מלים ממוצע למשפט, מספר פונימות למילה, תאחוז ה' הידיעה או ה' החיבור במלים שעשויות לשאת אותן, תאחוז המופע של בניין-פועל מסוים מבין כל מילות הפועל הקיימות בטקסט (למשל, באיזה אחוז ממילות-פועל מופיע בניין-הפועל "התפעל"), שכיחות המופע של מעברים בין קבוצות של מילים, למשל, משם עצם לשם עצם, משם עצם לשם תואר, מפועל לשם עצם. יש לשים לב שמדדים אלו אינם מתייחסים כלל לתוכן הטקסט אלא רק ל"צורתו הפורמלית" (מבחינת סגנון). כיוון שסגנון של מחבר נקבע בדרך כלל באופן בלתי מודע, ניתן להניח שניתוח סטטיסטי משווה של טקסטים שכביכול נכתבו על ידי אנשים שונים יוכל להכריע בשאלה האם באמת מקור הטקסטים הוא מחבר יחיד או שאכן נכתבו על ידי אנשים שונים.

בעבודתנו המשותפת (שתרמו לה פרקים גם פרופ' דיטר וויקמן מגרמניה, פרופ' משה פולצ'ק, פרופ' חיים רבין ופרופ' שמריהו טלמון), בקשנו לבדוק באופן לשוני-סטטיסטי את השערת המקורות (Documentary Hypothesis) מבית מדרשו של וולהאוזן (Wellhausen, 1844-1918). האחרון היה פרופסור



המחקר שביצענו עורר התעניינות רבה בעולם ועד היום מפורסמים מאמרים, עבודות תיזה ועבודות דוקטורט, שמתייחסים לממצאים של אותו מחקר. רשימה חלקית של אותם מאמרים (שאותרו על ידי חיפוש ב"סקולר" של גוגל, עם מילות מפתח "Shore" ו-"Radday") מובאת בסוף מאמר זה. הקורא יכול למצוא מאמרים נוספים ברשימות הספרות המתפרסמות באותם מאמרים. המאמרים ברשימה (כולל דוגמה של עבודת תיזה ושל עבודת דוקטורט) נבחרו באופן אקראי מבין למעלה מ-100 פרסומים המתייחסים למחקר שפורסם בספר ב-1985.<sup>1</sup>

### 6. סיכום

מסורת היא מחז"ל כי המלים בשפת הקודש לעתים מוסרות מידע נסתר שמוסיף למידע הגלוי של המילה. הבאנו מספר דוגמאות למסורת זאת מהמקורות. בחיבור זה ניסינו להביא, בלשון פשוטה ככל האפשר, מספר ניתוחים כמותיים שמלמדים כי ייתכן שהתופעה של ייצוג אופיינים מרכזיים של האובייקט המבוטא במילה באמצעות הערך המספרי של המילה בעברית - ייתכן שתופעה זאת נרחבת בלשון הקודש מששיערנו עד כה. מגוון הנושאים בהם אותה תופעה חוזרת שוב ושוב, כפי שהצגתי בספרי, אינה מאפשרת להתייחס אליה כקוריוז נטול חשיבות. במהדורה המעודכנת של הספר, שיצאה לאור בתחילת 2008, נוסף פרק חדש המחיל ניתוח סטטיסטי, בגישה דומה לזאת שאת קווי המתאר העיקריים שלה הצגנו כאן, על נתונים המובאים בפרק א' של ספר בראשית. מעט מניתוחים אלו הוצגו בסעיף 4.<sup>2</sup>

פרוטסטנטי ללימודי המזרח בגרמניה, ובעבודתו המחקרית, שסיכמה גם מחקרים קודמים, הניח כי לספר בראשית תרמו מקורות שונים שכוננו מקור E (קיצור של Elohist), מקור J (קיצור של Jahwist) ומקור P (קיצור של Priestly Writer). במלים אחרות: ההנחה הבסיסית של אסכולת וולהאוזן (שכונתה השערת המקורות) היא שספר בראשית הוא קומפילציה של טקסטים מקראיים ממקורות שונים. השערה זאת עומדת כמובן בניגוד לעמדה (מאד בסיסית) של היהדות כי חמשת חומשי תורה נכתבו על ידי מחבר יחיד. אסכולת וולהאוזן הפכה לזרם המרכזי בחקר המקרא באקדמיה בכל העולם.

במחקר שבוצע על ידינו חולק הטקסט של ספר בראשית (כ-19,200 מילה) ל-96 מדגמים אחידים על פי השערת המקורות (כלומר, מילות כל מדגם נלקחו כולן מאותו מקור, בייחוסו על פי השערת המקורות). ביחס לכל מדגם של מלים חושבו 54 מדדים כמותיים של סגנון, כגון המדדים שהובאו לפני כן. באמצעות ניתוח רב משתני בדקנו האם ישנה מובהקות סטטיסטית להשערת המקורות, כלומר, האם ניתן לדחות את השערת האפס כי ספר בראשית הוא אחד מבחינת מחברו. הניתוח הסטטיסטי הראה כי אין כל הוכחה סטטיסטית להשערת המקורות.

במילים אחרות, אותן טכניקות סטטיסטיות שבמחקרים קודמים (הן של רדאי ושלי והן של חוקרים אחרים) הראו בבירור כי ביחס לטקסטים מקראיים שאין חולק שנכתבו על ידי מחברים שונים הדבר הוכח גם באופן סטטיסטי, אותן טכניקות כשלו להראות מובהקות סטטיסטית שתצביע על כך כי ספר בראשית אכן נכתב על ידי מחברים שונים. ומכאן שלא ניתן, על סמך הניתוח הסטטיסטי, לדחות את המקובל במסורת היהודית שספר בראשית נכתב על ידי מחבר יחיד.

1. Radday, YT, Shore, H. *Genesis - an Authorship Study in Computer Assisted Statistical Linguistics*. With contributions by Wickman D, Pollatschek MA, Rabin C, Talmon S. *Analecta Biblica*-103. Rome Biblical Institute Press (Romae E Pontificio Instituto Biblico). 1985.

2. עבודת מחקר זאת בוצעה בזמנו הפרטי של פרופ' שור והיא אינה קשורה לעבודתו המחקרית באוניברסיטת בן-גוריון. האתר האישי בו ניתן למצוא מאמר זה (בעברית או באנגלית): <http://www.bgu.ac.il/~shor/index.htm>

רשימה ביבליוגרפית (חלקית; ראו סעיף 5)

- [1] Forsyth, RS. *Stylistic Structures - A computational approach to text classification*. PhD research report. Faculty of Science at the University of Nottingham, UK. 1995.
- [2] Glover, AD. *Automatically detecting stylistic inconsistencies in computer-supported collaborative writing*. MSc research report. Department of Education, University of Toronto. 1996.
- [3] Grieve, J. *Quantitative authorship attribution: an evaluation of techniques*. *Literary and Linguistic Computing* 2007. 22(3): 251-270.
- [4] Holmes, DI. *Authorship attribution*. *Computers and the Humanities* 1994. 28: 87-106.
- [5] Houk, CB. *Syllable-word patterns in Esther*. *ZAW* 2003. 115: 578-585.
- [6] Houk, CB. *Statistical analysis of Genesis sources*. *Journal for the Study of the Old Testament* 2002. 27(1): 75-105.
- [7] José Alviar J. *Recent advances in computational linguistics and their application to biblical studies*. *New Testament Studies* 2008. 54: 139-159.
- [8] Stern, D. *Recent trends in biblical source criticism*. *Jewish Bible Quarterly* 2008