



מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
المركز القطري للامتحانات والتقييم
מיסודן של האוניברסיטאות בישראל

דפוסי היבחנויות חוזרות בבחינה הפסיכומטרית כפונקציה של משתנים דמוגרפיים

דוח מחקר

RR
17-05

נובמבר 2017

דביר קלפר

דוח מחקר

RR-17-05

ISBN:978-965-502-208-7

**דפוס י היבחנויות חוזרות
בבחינה הפסיכומטרית
כפונקציה של משתנים
דמוגרפיים**

דביר קלפר

נובמבר 2017

תוכן עניינים

3.....	רשימת לוחות.....
4.....	רשימת איורים.....
5.....	1. תמצית.....
6.....	2. מבוא.....
6.....	2.1. היבחנויות חוזרות בבחינה הפסיכומטרית.....
6.....	2.2. מחקרי היבחנויות חוזרות במאל"ו.....
9.....	2.3. מטרות המחקר וחשיבותו.....
9.....	2.4. מבנה הדוח.....
10.....	3. שיטה.....
10.....	3.1. אוכלוסייה.....
10.....	3.2. משתני המחקר.....
11.....	3.3. נתונים חסרים.....
11.....	3.3.1. זקיפות פשוטות.....
12.....	3.3.2. זקיפות מרובות.....
12.....	3.4. מודלי המחקר כללי.....
13.....	3.5. מודל צמיחת הציון בבחינות חוזרות.....
13.....	3.5.1. מודל צמיחה לינארי – חותך אקראי.....
13.....	3.5.2. מודל צמיחה לינארי – חותך ושיפוע אקראיים.....
14.....	3.5.3. מידול מבנה השארית.....
14.....	3.5.4. מידול בקרה על משתני רקע אישיים.....
15.....	3.5.5. מודל צמיחה מקוטע (ספליין).....
16.....	3.5.6. חותך ושיפועים אקראיים במודל מקוטע.....
16.....	3.5.7. משתני רקע במודל מקוטע.....
16.....	3.6. מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת.....
17.....	3.7. מודל רגרסיה לחיזוי השיפור בציון הבחינה השנייה.....
18.....	4. תוצאות.....
18.....	4.1. סטטיסטיקה תיאורית.....
21.....	4.2. מודל צמיחת הציון בבחינות חוזרות.....

21.....	4.2.1	מודל צמיחה לינארי – חותך אקראי
22.....	4.2.2	מודל צמיחה לינארי – חותך ושיפוע אקראיים
22.....	4.2.3	מידול מבנה השארית
23.....	4.2.4	מידול בקרה על משתני רקע אישיים
24.....	4.2.5	מודל צמיחה מקוטע (ספליין)
24.....	4.2.6	חותך ושיפועים אקראיים במודל מקוטע
25.....	4.2.7	משתני רקע במודל מקוטע
26.....	4.3	מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת
27.....	4.4	מודל רגרסיה לחיזוי השיפור בציון הבחינה השנייה
29.....	5	דיון
29.....	5.1	סיכום הממצאים
29.....	5.1.1	מודל צמיחת הציון בבחינות חוזרות
30.....	5.1.2	מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת
31.....	5.1.3	מודל רגרסיה לחיזוי שיפור הציון בהיבחנות השנייה
31.....	5.2	השוואת המחקר הנוכחי למחקרי עבר
33.....	5.3	מגבלות המחקר והצעות למחקרי המשך
34.....	6	מקורות
35.....	7	נספח
35.....	7.1	השלמות בשיטת זקיפה מרובה
37.....	7.2	הצגות גרפיות של לוחות 4-7
39.....	7.3	הצגה פורמלית של מודל צמיחה
39.....	7.3.1	רמה ראשונה
39.....	7.3.2	רמה שנייה
40.....	7.3.3	משוואה אחת
40.....	7.3.4	כתיב מטריצות
40.....	7.3.5	מידול מבנה השארית
41.....	7.3.6	מודל צמיחה מקוטעת (ספליין)

רשימת לוחות

- לוח 1 – סטטיסטיקה של חסרים לפי בנאדם לאחר השלמה פשוטה.....11
- לוח 2 – שכיחות הרשומות/אנשים לפי מספר ההיבחנות.....18
- לוח 3 – סטטיסטיקה של הזמן בין היבחנויות עוקבות.....18
- לוח 4 – סטטיסטיקה תיאורית עבור המשתנים ציון הבחינה, גיל ומין.....19
- לוח 5 – סטטיסטיקה תיאורית עבור השכלת אב לפני ההשלמות ולאחריהן.....20
- לוח 6 – סטטיסטיקה תיאורית עבור השכלת אם לפני ההשלמות ולאחריהן.....20
- לוח 7 – סטטיסטיקה תיאורית עבור מצב סוציאקונומי לפני ההשלמות ולאחריהן.....21
- לוח 8 – תוצאות מודל הצמיחה: חותך אקראי.....21
- לוח 9 – תוצאות מודל הצמיחה: חותך ושיפוע אקראיים.....22
- לוח 10 – תוצאות מודל הצמיחה: מידול מבנה השארית.....22
- לוח 11 – תוצאות האפקטים הקבועים עבור מודל הצמיחה: בקרה על משתני רקע.....23
- לוח 12 – תוצאות האפקטים האקראיים עבור מודל הצמיחה: בקרה על משתני רקע.....24
- לוח 13 – תוצאות מודל הצמיחה המקוטעת: ללא בקרה על משתני רקע.....25
- לוח 14 – תוצאות האפקטים הקבועים עבור מודל הצמיחה המקוטעת: בקרה על משתני רקע...26
- לוח 15 – תוצאות האפקטים האקראיים עבור מודל הצמיחה המקוטעת: בקרה על משתני רקע.....26
- לוח 16 – תוצאות מודל הרגרסייה הלוגיסטית לחיזוי הסיכוי לבחינה חוזרת.....27
- לוח 17 – תוצאות מודל הרגרסייה לחיזוי השיפור בציון הבחינה השנייה.....28
- לוח 18 – קידוד המשתנים במודל צמיחה מקוטעת.....41

רשימת איורים

- איור 1 – צמיחת הציון הממוצע כפונקציה של מספר היבחנות 15
- איור 2 – ציון מכפ"ל הממוצע כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית..... 37
- איור 3 – הגיל הממוצע כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית..... 37
- איור 4 – השכלת האב הממוצעת כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית 38
- איור 5 – השכלת האם הממוצעת כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית..... 38
- איור 6 – מצב סוציאקונומי הממוצע כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית. 38
- איור 7 – המחשה של הפרמטרים השונים המשתתפים במודל צמיחה מקוטעת..... 42

1. תמצית

הקבלה ללימודים במוסדות להשכלה גבוהה בישראל מתבססת בעיקר על ציוני תעודת הבגרות ועל הציון בבחינה הפסיכומטרית. לאור זאת, רבים מהמועמדים נבחנים בבחינה הפסיכומטרית יותר מפעם אחת כדי לשפר את סיכויי קבלתם לחוג שהם רוצים ללמוד בו.

מחקרים מצאו קשר בין נתוני רקע דמוגרפיים (מין, גיל, השכלת הורים ועוד) להישגים לימודיים ולציוני בחינות (למשל Liu, Lee & von Davier, 2012), ובהן גם הבחינה הפסיכומטרית (סער ואורן, 2014; קלפר ואחרים, 2015).

מטרת המחקר הנוכחי היא לבדוק את דפוס ההיבחנויות החוזרות אצל אנשים שנבחנים יותר מפעם אחת בבחינה הפסיכומטרית בהתחשב בנתוני הרקע הדמוגרפי שלהם, ובפרט לענות על שלוש השאלות שלהלן:

- מהו השינוי הממוצע בציון הבחינה הפסיכומטרית אצל אנשים שנבחנים בה יותר מפעם אחת, ומהם הגורמים המשפיעים עליו?
- האם משתני הרקע מסייעים לניבוי הסיכוי שנבחן ייבחן בבחינה הפסיכומטרית יותר מפעם אחת? אם כן, באיזה כיוון הם משפיעים?
- האם משתני רקע מסייעים לניבוי מידת השיפור בציון בהיבחנות חוזרות? אם כן, באיזה כיוון הם משפיעים?

בסיס הנתונים כולל את כל הנבחנים שנבחנו בבחינה הפסיכומטרית בשנים 2001-2010 בשפה העברית, להוציא נבחנים שנבחנו חמש פעמים או יותר (פחות מ-1% מכלל הנבחנים), שאצלם נמצא דפוס חריג. במחקר נכללו 436,784 רשומות (היבחנויות) של 336,193 נבחנים שונים.

בכמה מהרשומות לעיל חסרו נתונים רבים הנוגעים למשתני הרקע הדמוגרפיים: השכלת אב, השכלת אם ומצב סוציאקונומי. נתונים אלו הושלמו בשני שלבים: ראשית בעזרת נתונים מהיבחנויות אחרות של אותו הנבחן, ושנית בשיטת "זקיפה מרובה" (multiple imputation).

הותאמו שלושה מודלים שונים כדי לענות על שלוש שאלות המחקר:

- כדי לענות על השאלה הראשונה לגבי השינוי הממוצע בהיבחנויות חוזרות הותאם מודל צמיחה (growth model) היררכי דו-רמתי (multilevel) למידול נתוני אורך (longitudinal data). הרמה הראשונה במודל היא המדידות החוזרות, והרמה השנייה היא האנשים.
 - נמצא כי הצמיחה מקוטעת, ולכן הותאם מודל ספלייני (spline) עם נקודת שבירה אחת בהיבחנות השנייה.
- כדי לענות על השאלה השנייה לגבי האפשרות לנבא סיכוי של נבחן להיבחן יותר מפעם אחת בבחינה הפסיכומטרית בעזרת משתני רקע הותאמה רגרסיה לוגיסטית. משתנה המטרה ברגרסיה זו הוא מי שנבחן פעם אחת בלבד, וזאת לעומת מי שנבחן יותר מפעם אחת.
- כדי לענות על השאלה השלישית לגבי האפשרות לנבא את רמת השיפור בציון בהיבחנות חוזרות נאמד השיפור בציון בהיבחנות השנייה בעזרת הפעלת רגרסיה לינארית על משתני הרקע.

נמצא כי השיפור הממוצע בציון הבחינה בין ההיבחנות הראשונה לשנייה הוא 45, והשיפור הממוצע בין ההיבחנות השנייה לרביעית הוא 26 (בכל היבחנות). כמו כן נמצא כי בהיבחנויות חוזרות המשתנים הדמוגרפיים משפיעים על כל ההיבטים הנבדקים.

2.1. היבחנויות חוזרות בבחינה הפסיכומטרית

האוניברסיטאות בישראל מקבלות מועמדים ללימודים בעיקר על בסיס הישגיהם בתעודת הבגרות ובבחינת הכניסה הפסיכומטרית לאוניברסיטאות (מכפ"ל). בתחילת תקופת המחקר התיר המרכז הארצי לבחינות ולהערכה (מאל"ו), האחראי לבחינה זו, לכל אדם לחזור ולהיבחן בבחינה הפסיכומטרית בתנאי שעברו לפחות שישה חודשים מאז היבחנותו הקודמת, ובמהלך תקופת המחקר השתנה התנאי לשני מועדי בחינה מאז היבחנותו הקודמת (ב-1.9.2017 הוסרה מגבלה זו לחלוטין). היות שמרבית המוסדות נוהגים להשתמש בציון הגבוה ביותר של הנבחן, רבים מהנבחנים בבחינה הפסיכומטרית נבחנים בה יותר מפעם אחת כדי לשפר את סיכויי קבלתם לחוג שהם רוצים ללמוד בו. שיפור בציון הבחינה הפסיכומטרית שאינו מקרי יכול לנבוע משני גורמים: האחד הוא רכישת ניסיון בפתרון שאלות הבחינה, והשני הוא למידה אינטנסיבית ושיפור אמיתי ביכולת הנמדדת בה. אף שסביר להניח כי לפחות חלק מן השיפור הוא אמיתי, אין להתעלם מכך שמחקרים מסוג זה, הבודקים את מידת השיפור, לוקים בהטיה הנובעת מברירה עצמית (self-selection): הנבחנים שבחרו להיבחן שנית הם רק אלה שמעריכים שיכולתם האמיתית לא השתקפה בציונם בהיבחנות הראשונה או שהם מסוגלים לשפר את היכולות הנבדקות בבחינה, והדבר עלול להטות את השיפור הנמדד כלפי מעלה. אם אומנם הערכתם נכונה, כי אז אפקט הבחירה העצמית, המסתייע ברגרסיה אל הממוצע, עשוי להיות מרכיב חשוב בשיפור ציונם בהיבחנות חוזרות. מחקר זה אינו נוגע בשאלה האם השיפור אמיתי, והוא מתאר את דפוס השיפור בלי להביא הטיה זו בחשבון.

בממוצע, כ-30% מהנבחנים בכל מועד בחינה הם נבחנים חוזרים (היבחנות שנייה ומעלה).

2.2. מחקרי היבחנויות חוזרות במאל"ו

מחקרים שנערכו במאל"ו מראים כי נבחנים חוזרים אומנם משפרים את הציון הפסיכומטרי הכללי שלהם בהיבחנות השנייה, ושיפור זה עומד על 30-50 נקודות. את המחקרים שנערכו במאל"ו ועסקו בהיבחנויות החוזרות בבחינה הפסיכומטרית אפשר לחלק לשני סוגים:

I. סוג אחד של מחקרים, הדומה ברוחו לזה הנוכחי (בן-סימון וכהן, 1988; סטולר ואללוף, 2002), בודק את דפוס ההיבחנות בצורה טהורה.

○ מחקרים של בן-סימון וכהן הקיף את כל מי שנבחנו בבחינה הפסיכומטרית פעמיים או יותר בין דצמבר 1983 לדצמבר 1986. בתקופה זו היה אחוז הנבחנים שניגשו לבחינה בפעם השנייה בשנתון בחינה בין 13% ל-17.6%. אחוז הנבחנים שניגשו לבחינה בפעם השלישית בשנתון בחינה היה בין 1% אחד ל-2% בקירוב. מאז, כפי ששיערו בזמנו החוקרים, גדל מאוד אחוז הנבחנים שניגשו לבחינה יותר מפעם אחת בשנתון. ממצאי מחקרים מראים כי בקרב הנבחנים בנוסח העברי של הבחינה הנבחנים פעמיים ($N=8,071$) שיפרו את ציונם הפסיכומטרי הכללי בבחינה השנייה ב-32.5 נקודות בממוצע, ואילו הנבחנים שלוש פעמים ($N=374$) שיפרו את ציונם

בבחינה השנייה ב-25.6 נקודות בממוצע, ובבחינה השלישית ב-23.8 נקודות בממוצע (השיפור הממוצע בבחינה השלישית ביחס לבחינה הראשונה היה 49.4 נקודות). כמו כן נמצא כי הציון הפסיכומטרי הממוצע בבחינה הראשונה של הנבחנים פעמיים (478.8) והנבחנים שלוש פעמים (449.2) נמוך מהממוצע בקרב אוכלוסיית הנבחנים בנוסח העברי בין דצמבר 1983 לדצמבר 1986 (500). כך נמצא גם לגבי ציון ההיבחנות השנייה (474.8) והיבחנות השלישית (498.6) של הנבחנים שלוש פעמים ("משולשים"), ורק ציונם של הנבחנים פעמיים ("כפולים") בהיבחנותם השנייה (511.3) עולה במעט על הממוצע. פיזור הציונים של הנבחנים החוזרים בכל ההיבחנויות (סטיית תקן בין 78.9 ל-84.1) נמצא קטן מפיזור באוכלוסייה הכללית (סטיית תקן 100). בהשוואה בין המינים נמצא כי שיעור הגברים גבוה משיעור הנשים הן בקרב נבחנים "כפולים" (53.9% לעומת 46.1%) והן בקרב נבחנים "משולשים" (57.2% לעומת 42.8%). בקרב נבחנים "כפולים" לא נמצא הבדל בין רמת שיפור הציון בהיבחנות השנייה אצל גברים ובין רמה זו אצל נשים.

○ מחקרם של סטולר ואללוף עסק בנבחנים חוזרים "כפולים" בלבד, לצורך זה קבע שנבחנים "כפולים" הם מי שנבחנו בפעם השנייה בשנים 1995-2001. אוכלוסיית הנחקרים כללה נבחנים שנבחנו פעמיים בבחינה הפסיכומטרית באחת משלוש השפות האלה: עברית, ערבית ורוסית. תוצאות מחקרם לגבי ההיבחנות בעברית הן הרלוונטיות למחקרנו. נמצא כי בשנים אלו נבחנו 72,421 איש פעמיים: מחציתם נבחנו בשנית בתוך 18 חודשים למן ההיבחנות הראשונה, כ-75% בתוך 36 חודשים, וכ-95% בתוך 60 חודשים. השיפור הממוצע בציונם בהיבחנות השנייה עומד על 49.1 נקודות (סטיית התקן של השיפור היא 41.7). כדי לחקור כיצד הציון הראשון קשור לגודל השיפור בציון השני, חילקו החוקרים את אוכלוסיית הנבחנים בעברית לשלוש קבוצות על בסיס ציונם בהיבחנות הראשונה. "הנמוכים" (עד 400) השתפרו במידה מועטה יחסית, "הבינוניים" (400-600) – השתפרו במידה רבה יחסית, ו"הגבוהים" (מעל 600) השתפרו במידה מועטה (ככל שהיה ציונם בהיבחנות הראשונה גבוה יותר, הוא השתפר בהיבחנות השנייה במידה מועטה יותר). השיפור הגדול ביותר הופיע אצל נבחנים שטווח ציונם בהיבחנות הראשונה היה 500-550 (שיפור של 56 נקודות), והגדול ביותר אחריו אצל נבחנים שטווח ציונם בהיבחנות הראשונה היה 450-500 (שיפור של 55 נקודות). כדי לחקור את השפעת מעבר הזמן בין שתי ההיבחנויות על השיפור בציון חולקו הנבחנים לשלוש קבוצות: "פרק זמן קצר" (עד 36 חודשים בין ההיבחנויות; כ-75% מהנבחנים), "פרק זמן בינוני" (37 חודשים עד 60 חודשים; כ-20% מהנבחנים) ו"פרק זמן ארוך" (61 חודשים ויותר בין שתי ההיבחנויות; כ-5% מהנבחנים). נמצא כי קבוצת פרק הזמן הקצר שיפרה את ציונה ב-42.3 נקודות בממוצע, קבוצת פרק הזמן הבינוני, שיפרה את ציונה ב-62.5 נקודות בממוצע, וקבוצת פרק הזמן הארוך שיפרה את ציונה ב-65.1 נקודות בממוצע.

II. סוג אחר של מחקרים (אורן וכהן, 1988; קנת-כהן ואורן, 1989) בדק את שאלת ההיבחנויות החוזרות בכוונה לברר את יעילות הציון ואת הוגנותו (יעילות הציון היא תוקף הניבוי שלו, והוגנות הציון עומדת ביחס הפוך להשפעת טעויות שיטתיות על ניבוי הישגי הקריטריון של חברי קבוצה כלשהי). במחקרים אלו נבדקה ההוגנות רק לגבי הנבחנים פעמיים לעומת הנבחנים פעם אחת בלבד. שלא במחקרים מהסוג הקודם, במחקרים אלו נתוני השיפור בציון הם תוצר לוואי של המחקר ולא עיקרו.

○ במחקרם של אורן וכהן כללה אוכלוסיית המחקר 677 תלמידים שלמדו במכינת הטכניון בשבעה מחזורים, מיולי 1984 ועד ספטמבר 1985, ונבחנו בבחינה הפסיכומטרית בשנים 1984-1986. מקרב 677 התלמידים 321 נבחנו בבחינה הפסיכומטרית פעמיים, ו-356 נבחנו בה פעם אחת בלבד. ממוצע ציוניהם של הנבחנים פעם אחת היה 567.0 עם סטיית תקן 79.3. ממוצע ציוניהם של הנבחנים פעמיים היה 499.7 עם סטיית תקן 60.7 בהיבחנות הראשונה ו-544.0 עם סטיית תקן 61.3 בהיבחנות השנייה – כלומר, ציונם השתפר ב-44.3 נקודות בממוצע.

○ במחקרם של קנת-כהן ואורן כללה אוכלוסיית המחקר 5,632 סטודנטים מארבעה מוסדות להשכלה גבוהה בישראל. 65% מן הסטודנטים היו תלמידי שנה ראשונה בשנת הלימודים תשמ"ו (1985/1986), ו-35% מן הסטודנטים היו תלמידי שנה ראשונה בשנת הלימודים תשמ"ז (1986/1987). במחקר נכללו רק סטודנטים שנבחנו בבחינה הפסיכומטרית בשפה העברית ולמדו בפקולטות למדעי הרוח, למדעי החברה, למדעי הטבע או הנדסה. 1,539 מהסטודנטים אלו נבחנו פעמיים. ממוצע ציונם של הנבחנים פעם אחת היה 564.7 עם סטיית תקן 88.8. ממוצע ציונם של הנבחנים פעמיים היה 496.2 עם סטיית תקן 76.2 בהיבחנות הראשונה ו-532.8 עם סטיית תקן 79.1 בהיבחנות השנייה – כלומר, ציונם השתפר ב-36.6 בממוצע.

מחקרם של בן סימון וכהן הוא המחקר היחיד שבדק את דפוס ההיבחנויות החוזרות ביותר משתי היבחנויות. מחקריהם של סטולר ואללוף ושל בן-סימון וכהן בדקו, בין השאר, את הקשר שבין משתני רקע (ציון בהיבחנות ראשונה ומשך הזמן בינה לבין ההיבחנות השנייה אצל סטולר ואללוף; מין, ציון בהיבחנות ראשונה ומשך הזמן בינה לבין ההיבחנויות שאחריה במחקרם של בן-סימון וכהן) לבין שיפור הציון בהיבחנות חוזרות, אם כי בדיקה זו הייתה מצומצמת ביותר וכללה בעיקר סטטיסטיקה תיאורית ללא מידול סטטיסטי.

לאור הממצאים בדבר קשר בין נתוני רקע דמוגרפיים (מין, גיל, השכלת הורים ועוד) להישגים לימודיים ולציוני בחינות (למשל Liu et. al., 2012), ובפרט לציוני הבחינה הפסיכומטרית (סער ואורן, 2014; קלפר ואחרים, 2015), ראוי לחקור את דפוסי ההיבחנויות החוזרות בבחינה הפסיכומטרית ואת הקשר בין משתנים מגוונים לבין מידת שיפור הציון בהן באופן מעמיק ושיטתי, בעזרת מודלים סטטיסטיים. נושא זה יש לחקור לא רק במסגרת השוואה בין היבחנויות ראשונה לשנייה, אלא גם במסגרת מקיפה יותר. לכך נועד המחקר הנוכחי.

2.3. מטרת המחקר וחשיבותו

מטרת המחקר הנוכחי היא לאפיין את דפוס ההיבחנויות החוזרות בבחינה הפסיכומטרית ואת הקשר שלו לרקע הדמוגרפי של הנבחנים, ובפרט לענות על שלוש השאלות האלה:

- מהו השינוי הממוצע בציון הבחינה הפסיכומטרית של אנשים שנבחנים בה יותר מפעם אחת, ומהם הגורמים המשפיעים עליו?
- האם משתני רקע מסייעים לניבוי הסיכוי שנבחן ייבחן בבחינה הפסיכומטרית יותר מפעם אחת? אם כן, באיזה כיוון הם משפיעים?
- האם משתני הרקע מסייעים לניבוי מידת השיפור בציון בהיבחנות חוזרות? אם כן, באיזה כיוון הם משפיעים?

רוב המחקרים הממדלים מדידות חוזרות עושים זאת בשיטת ניתוח שונות של מדידות חוזרות (repeated measures ANOVA). מחקר זה הוא הראשון הממדל את השיפור בציוני הבחינה הפסיכומטרית בהיבחנויות חוזרות בשיטת הניתוח הרב־רמתי. שיטה מתקדמת זו מאפשרת להתמודד עם התלות בין ציונים של אותו נבחן, עם השונות בין נבחנים ועם חוסר האיזון של מערך בסיס הנתונים (כלומר עם נתונים המעידים על יותר מדפוס היבחנויות אחד). נוסף על כך, מחקר זה מראה כיצד נתוני הרקע הדמוגרפי משפיעים על דפוסי ההתנהגות של נבחנים בהיבחנויות חוזרות ומשלים מחקר אחר שערכו קלפר ואחרים (2015).

לבסוף, כל המחקרים הקודמים שבחנו את ההיבחנויות החוזרות בבחינה הפסיכומטרית והשיפור בציון הנלווה להן נערכו לפני זמן רב (שלושה מחקרים נערכו לפני 30 שנה, ומחקר אחד נערך לפני 15 שנה), ולכן יש צורך לחזור על הבדיקות על שנים יותר עדכניות¹.

2.4. מבנה הדוח

פרק 3, המופיע לאחר המבוא, מתאר את שיטת המחקר: שלבי הכנת הנתונים והשלמתם ושלוש שיטות לניתוח הנתונים בהתאם לשלוש שאלות המחקר. פרק זה מפרט את שלביה של בניית המודל ההיררכי, החל בפשוט ביותר – מודל לינארי עם חותך אקראי בלבד, שאינו מכיל משתנים מסבירים – וכלה במודל ספליין, שמכיל משתנים מסבירים בשתי הרמות. בפרק זה המודלים הסטטיסטיים (בעיקר מודל ההשלמה המרובה והמודל ההיררכי) מותאמים להצגת הנתונים במחקר הנוכחי, ואילו השיטה הכללית, על נוסחאותיה המתמטיות, מובאת בנספח כדי שלא לקטוע את רצף ההבנה. פרק 4 מפרט את תוצאות יישומם של כל אחד מהמודלים לפי סדרם בפרק השיטה. פרק 5 דן בתוצאות אלו. פרק 7 כולל נספח העוסק במתמטיקה שמאחורי המודלים הסטטיסטיים, וכן איורים הממחישים כמה מן הנתונים הסטטיסטיים התיאוריים המובאים בלוחות בגוף הדוח.

¹ ואכן, כמתואר בתת-פרק 5.2 העוסק בהשוואה למחקרי העבר, כמה מתוצאות המחקר הנוכחי שונות מאוד מאלו שהתקבלו בעבר.

3. שיטה

3.1. אוכלוסייה

בסיס הנתונים המקורי הכיל את כל הנבחנים בבחינה הפסיכומטרית בשנים 2001-2010 ואשר נבחנו רק² בשפה העברית, למעט נבחנים שנבחנו חמש פעמים או יותר (ששיעורם נמוך מ-1% מהנבחנים³), וזאת משום שעבורם התקבל דפוס שונה. נעיר שבטווח זמן זה התאפשר לנבחנים להיבחן כמה פעמים, ובנוסף אפשר היה לדעת גם עבור נבחני סוף התקופה האם נבחנו יותר מפעם אחת (על ידי הצצה "לעתידי"⁴).

בסה"כ התקבלו 436,784 רשומות (היבחנויות) עבור 336,193 אנשים שונים⁵.

3.2. משתני המחקר

כל היבחנות נחשבת כרשומה והיא מכילה את השדות הבאים:

המשתנה התלוי (קריטריון לחיזוי):

הבחינה הפסיכומטרית היא סוללת מבחני ברירה, המיועדים למדוד היבטים שונים של יכולות לימודיות ולחזות הישגים אקדמיים עתידיים. בחינה זו משמשת את מוסדות הלימוד האקדמיים במיון מועמדיהם לחוגים השונים.

סולם הציונים של הבחינה נע מ-200 עד 800, עם ממוצע היסטורי של 500 וסטיית תקן של 100. המבחן מורכב משלושה תחומים, ומשקלוליהם של תחומים אלו בציון הכללי הרב-תחומי הם 40% לחשיבה כמותית, 40% לחשיבה מילולית ו-20% לאנגלית.

הציון הכולל בבחינה הינו המשתנה התלוי לחיזוי עבור המודל הראשון, ומהווה בסיס למשתנה התלוי אותו נבא במודלים האחרים.

משתנים בלתי תלויים (חזאים):

המשתנים הבלתי תלויים הבאים ישמשו כחזאים במודלים השונים:

- מספר היבחנות נוכחי ומספר היבחנות מקסימלי
- מין הנבחן (1=זכר, 2=נקבה)
- גיל נוכחי וגיל בהיבחנות ראשונה
- מצב סוציואקונומי (1=גבוה בהרבה מהממוצע...6=נמוך בהרבה מהממוצע)

² לצורך זה הסתכלנו גם על שנים שלפני שנת 2001.

³ ראה לוח 2 בנושא.

⁴ וראה הערת שוליים 2 לגבי הצצה "לעבר".

⁵ מספרים אלו הינם לאחר הוצאות רשומות שעבורן לא הצלחנו להשלים נתונים עבור מין וגיל שהוצאו מן המדגם (ראה תת פרק 3.3.1 בנושא).

- השכלת אב (1=לא למד...7=תואר שני ומעלה)
- השכלת אם (1=לא למד...7=תואר שני ומעלה)
- ציון בגרות
- מרווח הזמן מהבחינה הקודמת בשנים (לא קיים עבור היבחנות ראשונה)
- ציון בחינה נוכחי וציון בחינה ראשונה

3.3. נתונים חסרים

השדות של מספר היבחנות (נוכחי ומקסימלי), ציון בבחינה (נוכחי וראשון) וזמן מהבחינה הקודמת לא כללו נתונים חסרים.

המשתנים מין וגיל הכילו אחוזים בודדים של נתונים חסרים (כ-2% לגיל ו-6% למין).

בניגוד למשתנים הקודמים, בסיס הנתונים הכיל מספר רב של נתונים חסרים לגבי משתני הרקע הדמוגרפיים: השכלת אב, השכלת אם ומצב סוציאקונומי ועבור המשתנה בגרות.

3.3.1. זקיפות פשוטות⁶

הנתונים החסרים הושלמו תחילה בשיטה בסיסית. השיטה שנבחרה הייתה השלמה לפי נתוני היבחנויות אחרות, כאשר במקרים שהתגלתה סתירה נבחר המידע המוקדם יותר. רשומה שלא כללה מין ו/או גיל לאחר ביצוע השלמה פשוטה הוצאה מהמדגם, כך שבסופו של דבר עבור משתנים אלו אין נתונים חסרים.

גם לאחר ההשלמה הבסיסית רבות מהרשומות לא היו מלאות מאחר שבשלושת שדות הפרטים הדמוגרפיים ובשדה הבגרות חסרו נתונים רבים.

בלוח 1 מוצגות הסטטיסטיקות של הנתונים החסרים לפי בנאדם כאשר הסימן (+) מסמן "אין".

לוח 1 – סטטיסטיקה של חסרים לפי בנאדם לאחר השלמה פשוטה

המשתנים כבודדים	אחוז רשומות חסר	המשתנים בזוגות	אחוז רשומות חסר	המשתנים בשלוש וכולם	אחוז רשומות חסר
השכלת אב	13.56%	אב+אם	13.89%	אב+אם+סוציו	19.31%
השכלת אם	13.01%	אב+סוציו	19.09%	אב+אם+בגרות	42.70%
סוציאקונומי	16.96%	אב+בגרות	42.57%	אב+סוציו+בגרות	44.40%
ציון בגרות	41.65%	אם+סוציו	18.62%	אם+סוציו+בגרות	44.15%
		אם+בגרות	42.30%		
		סוציו+בגרות	43.88%	אב+אם+סוציו+בגרות	44.51%

⁶ המחבר מבקש להודות לעמיתו אליוט טורוול על העזרה בכתיבת קוד המחשב לצורך ביצוע הזקיפות הפשוטות.

ניתן לראות בלוח 1 שקיימת חפיפה די גדולה בין החסרים לגבי 3 המשתנים הדמוגרפיים. בנוסף, המשתנה ציון בגרות חסר מאוד ובדרך כלל כשהוא חסר אז גם שלושת המשתנים הדמוגרפיים חסרים. לפיכך, הוחלט לוותר על המשתנה ציון בגרות במחקר זה.

3.3.2. זקיפות מרובות

בשלב השני בוצעו השלמות של המשתנים הדמוגרפיים: השכלת אב, השכלת אם ומצב סוציאקונומי בשיטה מתקדמת המוכרת בשם "זקיפה מרובה" (MI=Multiple Imputation, Rubin, 1976, Schafer, 1997; 1978). תיאור מפורט של השיטה ניתן למצוא בנספח 7.1.

המשתנים ששימשו לצורך ביצוע הזקיפות הינם:

- מספר היבחנות נוכחי + מספר היבחנות מקסימלי
- מין הנבחן
- גיל נוכחי
- ציון בחינה נוכחי

ההשלמות בוצעו ברמת הנבחן ועבור הרשומה הראשונה הקיימת עבור כל נבחן, כאשר המידע שהושלם עבורה הועתק לשאר הרשומות הקיימות של אותו נבחן.

בוצעו בסך הכול 5 סטים של השלמות (ראה נספח 7.1).

אפשר לראות בלוח 5, לוח 6 ובלוח 7 כי שיטה זו אינה משנה בהרבה את הממוצעים ואת סטיות התקן של המשתנים אותם היא משלימה.

3.4. מודלי המחקר כללי

הותאמו שלושה מודלים שונים וזאת על מנת לענות על כל אחת משלוש שאלות המחקר:

- כדי לענות על השאלה הראשונה לגבי השינוי הממוצע, הותאם מודל דו-רמתי (bi-level) למידול נתוני אורך (longitudinal data) הנקרא גם מודל של צמיחה (growth model), כאשר הרמה הראשונה היא המדידות החוזרות והרמה השנייה היא האנשים.
 - נמצא כי הצמיחה אינה לינארית ולכן הותאם כמודל הסופי מודל ספלייני (spline) עם נקודת שבירה אחת.
- כדי לענות על השאלה השנייה לגבי האפשרות לנבא סיכוי של נבחן להיבחן יותר מפעם אחת בבחינה הפסיכומטרית בעזרת משתני רקע, הותאמה רגרסיה לוגיסטית כאשר משתנה המטרה דיכוטומי: נבחן פעם אחת בלבד=0; נבחן יותר מפעם אחת=1.
- כדי לענות על השאלה השלישית לגבי האפשרות לנבא את רמת השיפור מבחינה לבחינה, נבא את השיפור מבחינה ראשונה לשנייה בעזרת רגרסיה לינארית רגילה כאשר החזאים הם משתני הרקע והמשתנה התלוי הוא השיפור.

3.5. מודל צמיחת הציון בבחינות חוזרות⁷

כאמור, ניתן להתייחס אל נתוני אורך כנתונים היררכיים כאשר מדידות חוזרות משוכנות בתוך בן אדם. שימוש במודל רב-רמתי מאפשר להתמודד עם התלות הקיימת בציונים השונים של אותו נבחן, עם השונות שקיימת בין נבחנים שונים ועם העובדה שמערך בסיס הנתונים אינו מאוזן (קיימים דפוסים שונים של היבחנויות בנתונים שברשותנו).

נראה, אם כך שמודל דו-רמתי עבור ניתוח נתוני אורך יכול להתאים לנתונים שבידינו. מודלים רב-רמתיים (multilevel) למידול נתוני אורך (longitudinal data) נקראים גם מודלים של צמיחה (growth models). כאמור הרמה הראשונה במודל היא המדידות החוזרות והרמה השנייה היא האנשים.

כאן המקום להעיר, כי פרק זה מכיל את הנוסחאות של המודלים עבור המשתנים של המחקר הנוכחי. הצגה יותר כללית של המודלים ההיררכיים ניתן למצוא בנספח 7.3.

אנו נבנה את המודלים הלינאריים בשלבים כאשר לצורך נוחות פירוש התוצאות היבחנות ראשונה יתחשב כזמן אפס (היבחנות שנייה כזמן 1, וכו').

3.5.1. מודל צמיחה לינארי – חותך אקראי

ההכללה הפשוטה ביותר למודל הרגרסיה הלינארית הקלאסית מתקבלת כשמאפשרים לחותך להיות אקראי, מודל הנקרא מודל צמיחה עם חותך אקראי (random intercept model) ומוצג ע"י המשוואות הבאות בהתאם לשתי הרמות:

- רמה ראשונה: $e_{ij} + \text{היבחנות} * \beta_{1j} + \beta_{0j} = \text{ציון}$

- רמה שנייה: $\beta_{1j} = \gamma_{10} ; \beta_{0j} = \gamma_{00} + u_{0j}$

כאשר האינדקס j הוא האינדקס של הנבחן ואינדקס i הוא אינדקס ההיבחנות.

אם נרשום את שתי הרמות ביחד נקבל (על ידי הצבה של הרמה השנייה בראשונה):

$$\text{ציון} = \gamma_{00} + \gamma_{10} * \text{היבחנות} + (u_{0j} + e_{ij})$$

3.5.2. מודל צמיחה לינארי – חותך ושיפוע אקראיים

מודל חותך אקראי מניח שציוני הבחינה הראשונה שונים עבור אנשים שונים אבל שקצב הצמיחה זהה. כמובן שבמציאות לאנשים שונים יש שיפור שונה. לכן נציע את המודל הבא בו גם החותך וגם השיפוע אקראיים:

- רמה ראשונה: $e_{ij} + \text{היבחנות} * \beta_{1j} + \beta_{0j} = \text{ציון}$

⁷ תת-פרק זה מתבסס על תיאור השיטה וסדר בניית המודלים של Wang et al., 2012.

• רמה שנייה: $\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{10} + u_{1j}$

כאשר u_{0j} ו- u_{1j} מתפלגים לפי $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01}^2 \\ \sigma_{u01}^2 & \sigma_{u1}^2 \end{pmatrix} \right]$

נעיר שהרמה הראשונה זהה למודל הקודם – מודל עם חותך אקראי (בלבד), וההבדל הוא רק במשוואה השנייה עבור הרמה השנייה (זו עבור β_{1j}).

אם נרשום את שתי הרמות ביחד נקבל (על ידי הצבה של הרמה השנייה בראשונה):

$$\text{ציון} = \gamma_{00} + \gamma_{10} * \text{היבחנות} + (u_{0j} + u_{1j} * \text{היבחנות} + e_{ij})$$

3.5.3. מידול מבנה השארית

המידול עד כה היה מידול עם מקדמים אקראיים⁸. קיימת אופציה נוספת (במקום/בנוסף) של מידול השארית⁹. תיאור השיטה מפורט בנספח 7.3.5. נעיר רק שהתוצאות, כפי שהן מופיעות בפרק התוצאות (תת-פרק 4.2.3), מראות שהתוספת הזאת למודל לא משפרת את איכות התוצאות ולכן לא נמדל את השארית בהמשך.

3.5.4. מידול בקרה על משתני רקע אישיים

עד כה המודלים לא כללו משתנים מסבירים. ננסה להסביר את הרמה השנייה (רמת הבן אדם) על ידי הכנסת המשתנים הבאים למודל, כולם קבועים לאורך התקופה:

- מין
- גיל בבחינה הראשונה
- השכלת אב
- השכלת אם
- מצב סוציאקונומי

בהתאם לכך משוואות המודל יהיו:

• רמה ראשונה: $e_{ij} + \text{היבחנות} * \beta_{1j} + \beta_{0j} = \text{ציון}$

• רמה שנייה:

$$u_{0j} + \text{סוציו} * \gamma_{05} + \text{אם} * \gamma_{04} + \text{אב} * \gamma_{03} + \text{גיל ראשון} * \gamma_{02} + \text{מין} * \gamma_{01} + \gamma_{00} = \beta_{0j}$$

$$u_{1j} + \text{סוציו} * \gamma_{15} + \text{אם} * \gamma_{14} + \text{אב} * \gamma_{13} + \text{גיל ראשון} * \gamma_{12} + \text{מין} * \gamma_{11} + \gamma_{10} = \beta_{1j}$$

ואם נרשום את שתי הרמות ביחד נקבל (על ידי הצבה של הרמה השנייה בראשונה):

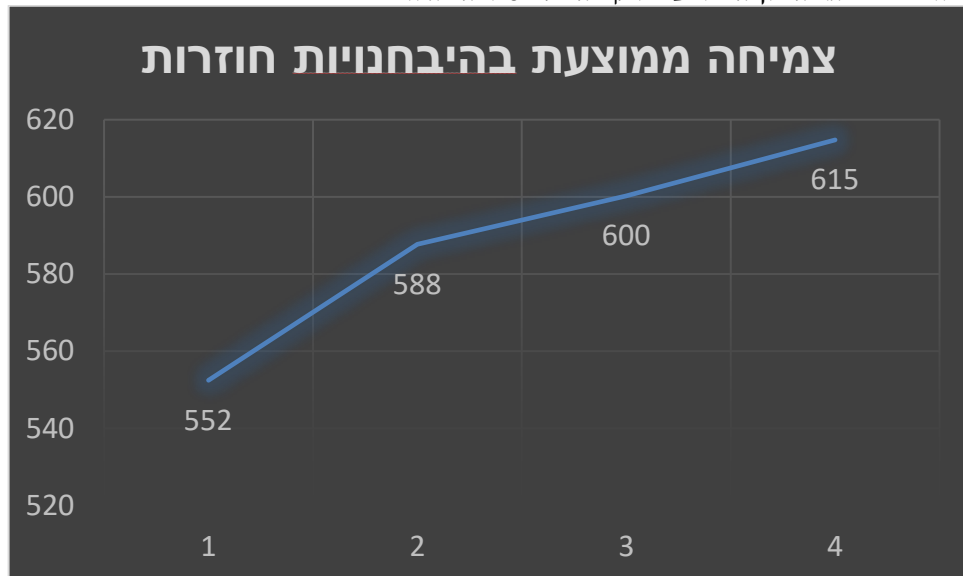
⁸ גישה זו ממדלת את השונות הבין נבחני, ונקראת (Littell et al., 2006) random coefficient model.
⁹ גישה זו ממדלת את השונות התוך נבחני, ונקראת (Fitzmaurice et al., 2004; Hedeker & Gibbons, 2006). למעשה השיטה שממדלת את השונות הבין נבחני הינה מקרה פרטי של השיטה הזו (Jennrich & Schluchter, 1986).

$$\begin{aligned}
& \text{מין} * (\gamma_{01} + \gamma_{11} * \text{היבחנות}) + \gamma_{00} + \gamma_{10} * \text{היבחנות} = \text{ציון} \\
& + (\gamma_{02} + \gamma_{12} * \text{היבחנות}) * \text{גיל ראשון} + (\gamma_{03} + \gamma_{13} * \text{היבחנות}) * \text{אב} \\
& + (\gamma_{04} + \gamma_{14} * \text{היבחנות}) * \text{אם} + (\gamma_{05} + \gamma_{15} * \text{היבחנות}) * \text{סוציו} \\
& + (u_{0j} + u_{1j} * \text{היבחנות} + e_{ij})
\end{aligned}$$

3.5.5. מודל צמיחה מקוטע (ספליין)

באיור 1 מוצג הציון הממוצע כפונקציה של מספר ההיבחנות.

איור 1 – צמיחת הציון הממוצע כפונקציה של מספר היבחנות



באיור 1 ניתן לראות שהצמיחה מתחלקת לשני חלקים : צמיחה גדולה בין היבחנות ראשונה לשנייה, ולעומתה צמיחה מתונה יותר ולינארית בין היבחנות שנייה לרביעית. בהתאם לכך נתאים בשלב זה מודל ספלייני (מקוטע). את המתמטיקה מאחורי מידול זה ניתן למצוא בנספח 7.3.6. נעיר רק שבניגוד למידול לינארי בו השתמשנו במשתנה זמן יחיד, כאשר ממדלים גידול עם שבירה אחת יש להשתמש בשני משתני זמן (ראה נספח למשמעות המתמטית שלהם).

3.5.6. חותך ושיפועים אקראיים במודל מקוטע

בשלב הראשון ניישם את המודל המקוטע ללא משתנים מסבירים (כלומר נכליל את המודל המתואר בתת-פרק 3.5.2). משוואות המודל בהתאם לקידוד בשני משתנים כפי שמתואר בנספח 7.3.6 הינן:

- רמה ראשונה: $\text{ציון}_{ij} = \beta_{0j} + \beta_{1j} * 1_{\text{היבחנות}} + \beta_{2j} * 2_{\text{היבחנות}} + e_{ij}$

- רמה שנייה: $\beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10} + u_{1j}; \beta_{2j} = \gamma_{20} + u_{2j}$

כאשר u_{0j}, u_{1j} ו- u_{2j} מתפלגים לפי:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01}^2 & \sigma_{u02}^2 \\ \sigma_{u01}^2 & \sigma_{u1}^2 & \sigma_{u12}^2 \\ \sigma_{u02}^2 & \sigma_{u12}^2 & \sigma_{u2}^2 \end{pmatrix} \right]$$

ואם נרשום את שתי הרמות ביחד נקבל (על ידי הצבה של הרמה השנייה בראשונה):

$$\begin{aligned} \text{ציון} &= \gamma_{00} + \gamma_{10} * 1_{\text{היבחנות}} + \gamma_{20} * 2_{\text{היבחנות}} + \\ & (u_{0j} + u_{1j} * 1_{\text{היבחנות}} + u_{2j} * 2_{\text{היבחנות}} + e_{ij}) \end{aligned}$$

3.5.7. משתני רקע במודל מקוטע

כעת נכניס גם את משתני הרקע למודל האחרון כדי לקבל את המודל הסופי שלנו. משוואות המודל בהתאם לקידוד בשני משתנים כפי שמתואר בנספח 7.3.6 הינן:

- רמה ראשונה:

$$\text{ציון}_{ij} = \beta_{0j} + \beta_{1j} * 1_{\text{היבחנות}} + \beta_{2j} * 2_{\text{היבחנות}} + e_{ij}$$

- רמה שנייה:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * \text{מין} + \gamma_{02} * \text{גיל ראשון} + \gamma_{03} * \text{אב} + \gamma_{04} * \text{אם} + \gamma_{05} * \text{סוציו} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * \text{מין} + \gamma_{12} * \text{גיל ראשון} + \gamma_{13} * \text{אב} + \gamma_{14} * \text{אם} + \gamma_{15} * \text{סוציו} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} * \text{מין} + \gamma_{22} * \text{גיל ראשון} + \gamma_{23} * \text{אב} + \gamma_{24} * \text{אם} + \gamma_{25} * \text{סוציו} + u_{2j}$$

3.6. מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת

כדי לענות על השאלה השנייה לגבי האפשרות לנבא סיכוי של נבחן להיבחן יותר מפעם אחת בבחינה הפסיכומטרית בעזרת משתני רקע, נבצע רגרסיה לוגיסטית כאשר משתנה המטרה הוא נבחן פעם אחת בלבד = 0 לעומת נבחן יותר מפעם אחת = 1.

המשתנים שהוכנסו למודל כמנבאים הם אותם המשתנים כמו במודל הצמיחה ובנוסף הוכנס גם הציון בבחינה הראשונה (שכמובן לא היה יכול להיכנס למודל הקודם), כך שבסך הכול הוכנסו המשתנים הבאים למודל:

- מין
- גיל בבחינה הראשונה
- ציון בבחינה הראשונה
- השכלת אב
- השכלת אם
- מצב סוציאקונומי

הפעם, בניגוד למודל הצמיחה, המודל הופעל ברמת נבחן, כלומר על בסיס נתונים שמכיל 336,193 רשומות זרות (ראה לוח 2 בהקשר זה).

3.7. מודל רגרסיה לחיזוי השיפור בציון הבחינה השנייה

כדי לענות על השאלה השלישית לגבי האפשרות לנבא את רמת השיפור מבחינה לבחינה, נסתכל על השיפור מבחינה ראשונה לשנייה. ראוי לציין שלכאורה מודל זה הינו מקרה פרטי של מודל הצמיחה בו עסקנו בפרק 3.5. כפי שנראה מיד, מודל זה מאפשר להכניס משתנים נוספים לניבוי שלא היה ניתן להכניסם למודל הקודם ובכך מגלם ערך מוסף. בנוסף, מודל זה פשוט מבחינה מתמטית בהרבה מזה של מודל הצמיחה בכך שהוא ממדל בעזרת רגרסיה "רגילה". בהתאם לכך, התוצאות שמתקבלות שונות מזו של מודל הצמיחה ובמקרה אחד אפילו הפוכות (ראה פרק 4.4 לתוצאות ופרק 0 לדיון בהם).

נסתכל על רשומות של אנשים שזו ההיבחנות השנייה שלהם (סה"כ 107,687) וננסה לנבא את השיפור בבחינה זו יחסית לבחינה הראשונה לפי המשתנים:

- מין
- גיל בבחינה הראשונה
- השכלת אב
- השכלת אם
- מצב סוציאקונומי

עד כאן אלו בדיוק אותם המשתנים כמו במודל הצמיחה. אך הפעם בנוסף לאילו (וכאן החידוש שלא היה אפשרי כאמור במודל הצמיחה) נכניס את שני זוגות המשתנים הבאים:

- זמן מהבחינה הראשונה + זמן מהבחינה הראשונה²
- ציון בחינה ראשונה + ציון בחינה ראשונה²

הסיבה לכך שעבור שני משתנים אלו אנו מכניסים גם את הריבוע שלהם היא שלמשל עבור המשתנה ציון בחינה ראשונה, אנשים עם ציון נמוך יתכן מאוד שלא מסוגלים לשפר הרבה ולעומתם אנשים עם ציון גבוה לא יכולים לשפר הרבה (אפקט תקרה).

לבסוף נדגיש שוב שהקריטריון שלנו לניבוי הוא ההפרש בין ציון הבחינה השנייה לציון הבחינה הראשונה.

4. תוצאות

בפרק זה נציג תחילה את הסטטיסטיקה התיאורית לאחר ההשלמות של החסרים (ראה פרק 3.3). לאחר מכן נציג את תוצאות השימוש במודלים השונים בהתאם למתואר בפרק 3.4.

4.1. סטטיסטיקה תיאורית

כאמור, בבסיס הנתונים 436,784 רשומות (היבחנות) מהם 336,193 זהויות שונות. בלוח 2 מוצגת הסטטיסטיקה של מספר הנבחנים כאשר הספירה נעשתה בשני אופנים: ספירה לפי מספר רשומות וספירה של אנשים.

לוח 2 – שכיחות הרשומות/אנשים לפי מספר ההיבחנות

ספירה של אנשים			ספירה של רשומות		
אחוזים	מספר אנשים	מספר היבחנות מקסימלית	אחוזים	מספר רשומות	מספר היבחנות
61.46%	206,618	1	69.36%	302,954	1
29.77%	100,096	2	24.65%	107,687	2
7.17%	24,119	3	5.14%	224,36	3
1.59%	5,360	4	0.85%	3,707	4
100.00%	336,193	סה"כ	100.00%	436,784	סה"כ

ניתן לראות בלוח 2 שככל שמספר הבחינה גבוה יותר כך שכיחות הנבחנים נמוכה יותר, כך שעבור בחינה רביעית יש כאחוז אחד בלבד, מה שמצדיק את הסינון שעשינו עבור היבחנות חמישית ומעלה (ראה פרק 3.1 בנושא).

בלוח 3 מוצג מרווח הזמן בין היבחנויות עוקבות של אנשים.

לוח 3 – סטטיסטיקה של הזמן בין היבחנויות עוקבות

בחינות	מתבסס על מספר רשומות	ממוצע הזמן בשנים	סטיית התקן של הזמן בשנים	חציון הזמן בשנים
1→2	107,687	2.45	2.24	1.67
2→3	22,436	1.51	1.88	0.87
3→4	3,707	1.20	1.69	0.70

ניתן לראות בלוח 3 כי מרווח הזמן הגדול ביותר בין הבחינות הינו בין הבחינה הראשונה לשנייה שעומד על ממוצע של 2.45 שנים בעוד שהזמן בין בחינה שנייה לשלישית הינו בממוצע 1.51 והזמן בין בחינה שלישית לרביעית הינו בממוצע 1.20 שנים. עוד ניתן להסיק מלוח 3 שמאחר שיש לנו

מידע עד סוף 2015 נוכל לדעת בקירוב טוב אם בנאדם נבחן לפחות פעמיים גם עבור נבחני סוף התקופה – מה שמצדיק בדיעבד את בחירת השנים¹⁰.
 לוח 4 מציג סטטיסטיקה תיאורית עבור המשתנים: ציון הבחינה, גיל ומין לפי מספר היבחנות מקסימלי ובתוכו מספר היבחנות.
 המחשה גרפית של הנתונים בלוח 4 ניתן למצוא באיור 2 ובאיור 3 בנספח 7.2.

לוח 4 – סטטיסטיקה תיאורית עבור המשתנים ציון הבחינה, גיל ומין לפי מספר היבחנות מקסימלי

מספר היבחנות מקסימלי	מספר היבחנות	מספר רשומות	ציון בחינה		גיל		אחוז נשים
			ממוצע	סטיית תקן	ממוצע	סטיית תקן	
1	1	206,618	560.30	106.29	21.68	3.44	55.32%
2	1	75,003	537.79	87.21	19.97	2.51	57.70%
	2	84,905	592.71	92.96	22.24	2.73	56.20%
3	1	17,421	527.62	85.51	19.38	2.30	57.60%
	2	18,660	568.99	85.73	21.27	2.42	56.22%
	3	18,477	601.75	89.73	22.86	2.81	56.58%
4	1	3,912	528.84	90.99	19.19	2.26	55.65%
	2	4,122	569.42	92.23	20.92	2.64	55.34%
	3	3,959	593.17	92.76	22.30	2.71	55.04%
	4	3,707	614.76	93.32	23.54	3.17	54.98%
סה"כ		436,784	564.12	100.11	21.43	3.17	56.08%

ניתן לראות בלוח 4 (וביתר קלות באיורים 2-3) כי ממוצע הציון בבחינה הראשונה של אלו שנבחנו בבחינה הפסיכומטרית רק פעם אחת, גבוה מציון הבחינה הראשונה של אלו שנבחנו יותר מפעם אחת. למעשה, ככל שנבחן חוזר על בחינה מספר פעמים רב יותר, ציונו בבחינה הראשונה נמוך יותר¹¹. הדבר נכון גם להיבחנות שנייה ושלישית. עם זאת, ככל שאתה נבחן יותר פעמים כך משתפר הציון שלך ועולה על זה שנבחן מספר פעמים קטן יותר. כך למשל, נבחן שנבחן בסך הכול פעמיים בממוצע ישיג ציון בבחינה השנייה הגבוה יותר מזה שהשיג נבחן בחינה יחידה. ממצא זה שונה מתוצאות מחקרי העבר; ראה תת-פרק 5.2 לדיון בנושא.
 לגבי הגיל, ניתן לראות (אולי שלא במפתיע) שככל שנבחנת יותר פעמים גילך בהיבחנות הראשונה היה צעיר יותר וכנ"ל לגבי ההיבחנות השנייה והשלישית.
 לבסוף, נשים נבחנות יותר מגברים בכל אחת מהקטגוריות.
 באופן דומה ללוח 4, בלוח 5, בלוח 6 ובלוח 7 מוצגים הסטטיסטיים התיאוריים של המשתנים הדמוגרפים: השכלת אב, השכלת אם ומצב סוציאקונומי בהתאמה לפי מספר היבחנות מקסימלי ובתוכו מספר היבחנות; התוצאות מוצגות טרם ההשלמה ולאחר ההשלמה.
 המחשה גרפית של הנתונים בלוח 5, בלוח 6 ובלוח 7 ניתן למצוא באיורים 4-6 בנספח 7.2.

¹⁰ וראה הערת שוליים 4 עבור הצצה "לעיתיד".

¹¹ אפשר גם להסתכל על זה בפרספקטיבה הפוכה: ככל שהציון בבחינה הראשונה נמוך יותר כך מספר ההיבחנות גדול יותר.

לוח 5 – סטטיסטיקה תיאורית עבור השכלת אב לפני ההשלמות ולאחריהן

אחרי השלמות			לפני השלמות			מספר היבחנות	מספר היבחנות מקסימלי
סטיית תקן	ממוצע	מספר רשומות	סטיית תקן	ממוצע	מספר רשומות		
1.49	4.78	206,618	1.53	4.84	168,704	1	1
1.48	4.95	75,003	1.50	4.97	69,586	1	2
1.49	4.90	84,905	1.50	4.91	80,134	2	
1.48	5.06	17,421	1.49	5.07	16,616	1	3
1.48	5.02	18,660	1.49	5.03	18,007	2	
1.49	4.98	18,477	1.50	4.99	17,932	3	
1.48	5.17	3,912	1.49	5.19	3,730	1	4
1.49	5.17	4,122	1.50	5.18	3,982	2	
1.50	5.13	3,959	1.50	5.14	3,849	3	
1.51	5.10	3,707	1.51	5.10	3,624	4	
1.49	4.88	436,784	1.51	4.91	386,164	סה"כ	

לוח 6 – סטטיסטיקה תיאורית עבור השכלת אם לפני ההשלמות ולאחריהן

אחרי השלמות			לפני השלמות			מספר היבחנות	מספר היבחנות מקסימלי
סטיית תקן	ממוצע	מספר רשומות	סטיית תקן	ממוצע	מספר רשומות		
1.43	4.87	206,618	1.47	4.92	170,106	1	1
1.44	5.00	75,003	1.46	5.02	69,888	1	2
1.44	4.92	84,905	1.45	4.94	80,413	2	
1.46	5.09	17,421	1.47	5.10	16,675	1	3
1.46	5.03	18,660	1.47	5.04	18,067	2	
1.48	4.97	18,477	1.48	4.98	17,988	3	
1.50	5.14	3,912	1.51	5.16	3,738	1	4
1.50	5.13	4,122	1.51	5.14	3,991	2	
1.51	5.09	3,959	1.52	5.10	3,859	3	
1.52	5.05	3,707	1.52	5.06	3,628	4	
1.44	4.93	436,784	1.47	4.97	388,353	סה"כ	

אחרי השלמות			לפני השלמות			מספר היבחנות	מספר היבחנות מקסימלי
סטיית תקן	ממוצע	מספר רשומות	סטיית תקן	ממוצע	מספר רשומות		
1.07	3.31	206,618	1.09	3.30	160,241	1	1
1.07	3.28	75,003	1.08	3.28	67,521	1	2
1.06	3.23	84,905	1.06	3.23	78,096	2	
1.09	3.32	17,421	1.10	3.32	16,304	1	3
1.08	3.29	18,660	1.08	3.29	17,697	2	
1.07	3.27	18,477	1.07	3.27	17,662	3	
1.09	3.31	3,912	1.09	3.30	3,689	1	4
1.09	3.29	4,122	1.09	3.29	3,945	2	
1.09	3.28	3,959	1.09	3.28	3,812	3	
1.07	3.27	3,707	1.08	3.27	3,593	4	
1.07	3.29	436,784	1.08	3.28	372,560	סה"כ	

ראשית וכפי שצוין כבר בתת-פרק 3.3.2, ניתן לראות בלוח 5, בלוח 6 ובלוח 7 שבסך הכול הן הממוצע והן סטיית התקן לא משתנים בהרבה לאחר תהליך ההשלמה. לוח 5 ולוח 6 מלמדים אותנו שכלל שהוריו של נבחן משכילים יותר (הן אב והן אם) הוא נוטה להיבחן יותר. לעומת זאת, לוח 7 מלמד אותנו שהתופעה הזו איננה קיימת עבור המשתנה מצב סוציאקונומי, או במילים אחרות, לא קיים קשר ברור בין המצב הסוציאקונומי של הנבחן ומספר הפעמים שבחר להיבחן.

4.2. מודל צמיחת הציון בבחינות חוזרות

4.2.1. מודל צמיחה לינארי – חותך אקראי

בלוח 8 מוצגות תוצאות המודל עם חותך אקראי בהתאם למפורט בתת-פרק 3.5.1. אמידת הפרמטרים בוצעה בשיטת REML (REstricted Maximum Likelihood) כאשר דרגות החופש נקבעו לפי KR^{12} .

לוח 8 – תוצאות מודל הצמיחה: חותך אקראי

אפקט	מקדם	רכיב שונות	טעות תקן	ערך t/Z
קבוע	γ_{00}		0.1753	3,157
	γ_{10}		0.1217	313
אקראי	$\sigma_{u0}^2(u_{0j})$	8,969	24.89	360
	$\sigma^2(e_{ij})$	1,170	5.28	222

*-2LL= 5,110,384

¹² KR=Kenward-Roger (Kenward & Roger, 1977; Freund et al., 2002) הינה השיטה המומלצת עבור בסיס נתונים שאינו מאוזן (SAS Institute Inc, 2004).

נעיר מספר הערות על תוצאות המודל :

- 553 הוא הציון הממוצע בבחינה הראשונה ו-38 הוא הצמיחה הממוצעת מבחינה לבחינה (כמובן שזו תוצאה משוקללת במספר האנשים).
- המשמעות של σ_{u0}^2 מובהק היא שנבחנים אכן שונים זה מזה בציון הבחינה הראשון.
- המשמעות של σ^2 מובהק היא שיש שונות בתוך נבחן מסוים בין הבחינות השונות גם לאחר הכנסת החותך האקראי, יחד עם זאת, רוב השונות היא בין אנשים.
- ניתן לכמת את שתי הנקודות הקודמות וזאת על ידי חישוב מקדם המתאם הבין רמתי ICC (Intraclass Correlation Coefficient). החישוב נותן :

$$ICC = \frac{\sigma_{u0}^2(u_{0j})}{\sigma_{u0}^2(u_{0j}) + \sigma^2(e_{ij})} = \frac{8,969}{8,969 + 1,170} = 0.88$$

4.2.2. מודל צמיחה לינארי – חותך ושיפוע אקראיים

בלוח 9 מוצגות תוצאות המודל עם חותך ושיפוע אקראיים בהתאם למפורט בתת-פרק 3.5.2.

לוח 9 – תוצאות מודל הצמיחה: חותך ושיפוע אקראיים

אפקט		מקדם	רכיב שונות	טעות תקן	ערך t/Z
קבוע	γ_{00}	553.08		0.1753	3,155
	γ_{10}	40.19		0.1325	303
אקראי	σ_{u0}^2		9,156	25.99	352
	σ_{u1}^2		314	9.84	32
	σ_{u01}^2		-108	15.53	-7
	$\sigma^2(e_{ij})$		945	7.12	133

$$*-2LL = 5,108,601$$

נעיר מספר הערות על תוצאות המודל :

- הערך השלילי עבור הערך של σ_{u01}^2 (-108) אומר שככל שהציון הראשון גבוה יותר השיפור קטן יותר.
- אם נשווה את ערכי מינוס לוג הנראות בשני המודלים נקבל (ראה תחתית לוח 8 ולוח 9): $\Delta\chi^2(2) = 1,783$. היות וזה ערך מובהק (ברמה $p < 0.0001$), המודל מהווה שיפור יחסית לקודמו.

4.2.3. מידול מבנה השארית

בלוח 10 מוצגות תוצאות המודל המכיל מידול של השארית וזאת בהתאם לתת-פרק 3.5.3.

אנו מציגים בלוח 10 רק את השורה הרלוונטית לשלב זה.

לוח 10 – תוצאות מודל הצמיחה: מידול מבנה השארית

אפקט		רכיב שונות	טעות תקן	ערך Z	ערך p
אקראי	AR(1)	0.03326	0.0267	1.24	0.2139

$$*-2LL = 5,108,600$$

נעיר מספר הערות על תוצאות המודל :

- ערך p שהתקבל (0.2139) הוא כמובן ערך שאינו מובהק.
- אם נשווה את ערך מינוס לוג הנראות במודל זה לערך המתקבל מן המודל הקודם (שכאמור מודל זה מרחיב) נקבל: $\Delta\chi^2(1) = 1$, וזה נותן ערך $p = 0.32$, כלומר ערך לא מובהק, או במילים אחרות המודל הנוכחי לא משפר את דיוק הניבוי.

שתי הנקודות האחרונות מראות שתוספת השארית למידול לא משפרת את התוצאות ולכן לא נמדל את השארית בהמשך.

4.2.4. מידול בקרה על משתני רקע אישיים

בלוח 11 ובלוח 12 מוצגות תוצאות המודל המכיל בקרה על משתני רקע אישיים וזאת בהתאם לתת-פרק 3.5.4. לוח 11 מציג את תוצאות האפקטים הקבועים ולוח 12 מציג את תוצאות האפקטים האקראיים.

לוח 11 – תוצאות האפקטים הקבועים עבור מודל הצמיחה: בקרה על משתני רקע

משתנה	מקדם
חותך	571.75
היבחנות	110.13
מין	-35.94
מין*היבחנות	-2.51
גיל ראשון	-0.56
גיל ראשון*היבחנות	-3.00
השכלת אב	9.76
השכלת אב*היבחנות	-0.50
השכלת אם	11.31
השכלת אם*היבחנות	-0.71
סוציו-אקונומי	-16.41
סוציו-אקונומי*היבחנות	-0.58

$$p < 0.0001$$

לוח 12 – תוצאות האפקטים האקראיים עבור מודל הצמיחה: בקרה על משתני רקע

משתנה	רכיב שונות
σ_{u0}^2	7,338
σ_{u1}^2	284
σ_{u01}^2	-169
$\sigma^2(e_{ij})$	939

$$p < 0.0001$$

נעיר מספר הערות על תוצאות המודל :

- הערכים של החותך ושל ההיבחות הינם חסרי משמעות וזאת משום שלא מרכזנו.
- מין : הציון הראשון של נשים נמוך ב-35.9 מזה של גברים, ובנוסף גודל השיפור מבחינה לבחינה בקרב נשים נמוך ב-2.5 נקודות מזה שבקרב גברים.
- גיל : ככל שהבחינה הראשונה נעשית בגיל מבוגר יותר הציון נמוך יותר (0.6 לכל שנה) וכך גם גודל השיפור שקטן ב-3.0 לכל שנה נוספת של גיל בחינה ראשון.
- השכלת אב ואם : ככל שההורים משכילים יותר הציון הראשון גבוה יותר (9.8 ו-11.3 לכל שלב עבור אב ואם בהתאמה). לעומת זאת, השיפור בציון קטן ככל שההורים משכילים יותר (0.5 ו-0.7 לכל שלב עבור אב ואם בהתאמה).
- מצב סוציאקונומי : ככל שהמצב הסוציאקונומי יותר טוב הציון הראשון יותר גבוה (ב-16.4 לכל שלב), ובנוסף קצב צמיחת הציון גדול יותר (ב-0.6 לכל שלב).

4.2.5. מודל צמיחה מקוטע (ספליין)

בתת-פרק זה נציג את תוצאות מודל הצמיחה המקוטעת. נציג תחילה את תוצאות המודל ללא משתני רקע ולאחר מכן את תוצאות המודל עם משתנה הרקע. המודל האחרון הינו המודל הסופי שלנו והדיון בו יעשה בתת-פרק 5.1.1 העוסק בסיכום ודיון של הממצאים.

4.2.6. חותך ושיפועים אקראיים במודל מקוטע

בלוח 13 מוצגות תוצאות מודל הצמיחה המקוטעת ללא משתני רקע וזאת בהתאם לתת-פרק 3.5.6.

אפקט		מקדם	רכיב שונות	טעות תקן	ערך t/Z
קבוע	γ_{00}	552.35		0.1758	3,142
	γ_{10}	45.14		0.1609	281
	γ_{20}	25.63		0.2544	101
אקראי	σ_{u0}^2		9,246	29.07	318.06
	σ_{u1}^2		506	31.00	16.34
	σ_{u2}^2		166	20.51	8.11
	σ_{u01}^2		-187	23.46	-7.99
	σ_{u02}^2		-298	29.94	-9.95
	σ_{u12}^2		205	16.84	12.17
	$\sigma^2(e_{ij})$		861	14.57	59.06

נעיר מספר הערות על תוצאות המודל:

- $\gamma_{00}=552.35$ זה הממוצע בהיבחנות ראשונה.
- $\gamma_{10}=45.14$ זו הצמיחה הממוצעת בין בחינה ראשונה לשנייה.
- אומנם באיור 1 זה נראה כאילו זה 36, אבל צריך לזכור שאלו שנבחנו פעם אחת כנראה קיבלו ציון גבוה יותר, וממוצע היבחנות ראשונה של אלו שנבחנו פעמיים נמוך מ-552.
- $\gamma_{20}=25.63$ זו הצמיחה הממוצעת בין בחינה שנייה לרביעית (לכל בחינה).
- $\sigma_{u01}^2=-187; \sigma_{u02}^2=-298$, כלומר ככל שהציון הראשון גבוה יותר כך השיפור בהמשך קטן יותר.
- $\sigma_{u12}^2=205$ משמע שגודל השיפור בהתחלה מתואם חיובית עם גודל השיפור בהמשך.

4.2.7 משתני רקע במודל מקוטע

בלוח 14 ובלוח 15 מוצגות תוצאות מודל הצמיחה המקוטעת המכיל בקרה על משתני רקע אישיים וזאת בהתאם לתת-פרק 3.5.7. לוח 14 מציג את תוצאות האפקטים הקבועים ולוח 15 מציג את תוצאות האפקטים האקראיים.

לוח 14 – תוצאות האפקטים הקבועים עבור מודל הצמיחה המקוטעת: בקרה על משתני רקע

משתנה	חותך	היבחנות1	היבחנות2
	564.97	135.54	73.14
מין	-35.65	-3.10	-2.40
גיל ראשון	-0.34	-3.80	-2.38
השכלת אב	9.77	-0.55	לא מובהק
השכלת אם	11.39	-0.97	לא מובהק
סוציו-אקונומי	-16.29	-1.03	0.88

$p < 0.0001$

לוח 15 – תוצאות האפקטים האקראיים עבור מודל הצמיחה המקוטעת: בקרה על משתני רקע

משתנה	רכיב שונות
σ_{u0}^2	7,441
σ_{u1}^2	444
σ_{u2}^2	142
σ_{u01}^2	-277
σ_{u02}^2	-275
σ_{u12}^2	148
$\sigma^2(e_{ij})$	857

* $p < 0.0001$

היות וזהו המודל הסופי שלנו, ניתוח התוצאות יובא בתת-פרק 5.1.1 העוסק בסיכום הממצאים ודיון.

נעיר כאן בכל זאת הערה אחת לגבי תוצאות המודל:

- בדומה למודל המקביל (עם משתני הרקע אבל ללא השבירה) גם כאן, היות ולא מרכזנו, ערכים בשורה הראשונה של לוח 14 חסרי משמעות.

4.3. מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת

בלוח 16 מוצגות תוצאות מודל הרגרסיה הלוגיסטית לחיזוי הסיכוי לבחינה חוזרת וזאת בהתאם לתת-פרק 3.6.

משתנים	אומדן למקדם= β	95% רווח בר-סמך ליחס הסיכויים		
		נמוך	יחס הסיכויים= e^β	גבוה
חותך	8.479*			
מין	-0.337*	0.703	0.714	0.725
גיל בבחינה הראשונה	-0.313*	0.729	0.731	0.734
ציון בבחינה הראשונה	-0.004*	0.996	0.996	0.996
השכלת אב	0.040*	1.033	1.041	1.049
השכלת אם	0.010*	1.003	1.010	1.017
מצב סוציאקונומי	-0.072*	0.923	0.930	0.938

* $p < 0.001$, $c = 0.734$

ניתוח התוצאות יובא בתת-פרק 5.1.2 העוסק בסיכום הממצאים ודיון.
נעיר כאן בכל זאת הערה אחת לגבי תוצאות המודל:

- השפעה של השכלת האם על הסיכוי של נבחן להיבחן לפחות פעמיים נמוכה מזו של השפעתו של האב (אם כי באותו הכיוון). אפקט דומה שמראה שלמשתנה השכלת אב יש יותר השפעה מזה של השכלת האם נמצא גם במחקרים קודמים שנעשו במאליי (ראה למשל, קלפר ואחרים, 2015).

4.4. מודל רגרסיה לחיזוי השיפור בציון הבחינה השנייה

בלוח 17 מוצגות תוצאות מודל הרגרסיה לחיזוי השיפור בציון הבחינה השנייה וזאת בהתאם לתת-פרק 3.7.

לוח 17 – תוצאות מודל הרגרסיה לחיזוי השיפור בציון הבחינה השנייה

משתנים	מקדם b	טעות תקן
חותך	-93.05*	4.55
מין	-3.78*	0.27
גיל בבחינה הראשונה	-1.70*	0.06
השכלת אב	0.88*	0.12
השכלת אם	0.84*	0.12
מצב סוציאקונומי	-2.92*	0.14
זמן מהבחינה הראשונה	12.04*	0.12
זמן מהבחינה הראשונה ²	-0.66*	0.01
ציון בחינה ראשון	0.7430*	0.015532
ציון בחינה ראשון ²	-0.0008*	0.000015

* $p < 0.0001$, $R^2 = 0.2061$

ניתוח התוצאות יובא בתת-פרק 0 העוסק בסיכום הממצאים ודיון.

5.1. סיכום הממצאים

פרק זה עונה על השאלות שנשאלו בפרק 2.2.

- תת-פרק 5.1.1 דן בתוצאות המודל הסופי, המודד את הצמיחה בציון בהיבחנות חוזרות, ועונה על שאלה 1 בדבר השינוי הממוצע בציון הבחינה הפסיכומטרית ובדבר הגורמים המשפיעים על שינוי זה.
- תת-פרק 5.1.2 דן בתוצאות מודל הרגרסיה הלוגיסטית ועונה על שאלה 2 בדבר האפשרות לנבא את הסיכוי להיבחן יותר מפעם אחת בבחינה הפסיכומטרית ובדבר הגורמים המשפיעים על סיכוי זה.
- תת-פרק 0 דן בתוצאות מודל הרגרסיה הרגילה ועונה על שאלה 3 על בדבר האפשרות לנבא את מידת השיפור בהיבחנות השנייה ובדבר הגורמים המשפיעים על מידה זו.

5.1.1. מודל צמיחת הציון בבחינות חוזרות

תת פרק זה מסכם את ממצאי השימוש במודל "חותך ושיפועים אקראיים במודל מקוטע" (המודל מתואר בתת-פרק 3.5.7, והממצאים מובאים בתת-פרק 4.2.7).

- השיפור הממוצע בציון בין ההיבחנות הראשונה לשנייה הוא 45.1, והשיפור הממוצע בציון בין ההיבחנות השנייה לרביעית (לכל היבחנות) הוא 25.6 (זוהי מסקנה העולה מהמודל ללא משתני בקרה, כפי שמוצג בלוח 13).
- מין: הציון הראשון של נשים נמוך ב-35.7 מזה של גברים. נוסף על כך, קצב השיפור בציון של נשים בין ההיבחנות הראשונה לשנייה נמוך ב-3.1 נקודות מזה של גברים, ובהיבחנות נוספות הוא נמוך ב-2.4 נקודות מזה של גברים.
- גיל: ככל שההיבחנות הראשונה של הנבחן נערכת בגיל מבוגר יותר, הציון בה נמוך יותר (0.3 נקודות לכל שנה). נוסף על כך, לכל שנת גיל קצב השיפור בין היבחנות ראשונה לשנייה קטן ב-3.8 נקודות, ובין ההיבחנות השנייה לרביעית ב-2.4 נקודות.
- השכלת אב ואם: ככל שהורי הנבחן משכילים יותר, ציונו בהיבחנות הראשונה גבוה יותר (9.8 ו-11.4 נקודות לכל רמת השכלה של האב ושל האם בהתאמה). לעומת זאת, ככל שההורים משכילים יותר קצב צמיחת הציון בין היבחנות ראשונה לשנייה קטן יותר (0.6 ו-1.0 נקודות לכל רמת השכלה של האב ושל האם בהתאמה). השכלת ההורים אינה משפיעה על קצב הצמיחה בין ההיבחנות השנייה לרביעית.
- מצב סוציאקונומי: ככל שהמצב הסוציאקונומי של הנבחן טוב יותר, ציונו בהיבחנות הראשונה גבוה יותר (16.3 נקודות לכל רמה סוציאקונומית). נוסף על כך, לכל עלייה ב-1 בסולם הרמות הסוציאקונומיות קצב צמיחת הציון גדול יותר ב-1.0 נקודות בין ההיבחנות הראשונה לשנייה וקטן יותר ב-0.9 נקודות בין ההיבחנות השנייה לרביעית.

5.1.2. מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת

תת פרק זה מסכם את ממצאי השימוש במודל הרגרסיה הלוגיסטית לחיזוי הסיכוי לבחינה חוזרת (המודל מתואר בתת-פרק 3.6, והממצאים מובאים בתת-פרק 4.3).

- מין: נשים נוטות לחזור ולהיבחן פחות מגברים (יחס הסיכויים להיבחנות חוזרת של אישה לעומת היבחנות חוזרת של גבר הוא 0.714). (נזכר כי לפי הקידוד שלנו המין מקודד כזכר=1 ונקבה=2).
- תוצאה זו אינה טריוויאלית כלל ולמעשה היא מנוגדת לתוצאה הנובעת מן הסטטיסטיקה התיאורית¹³ המופיעה בלוח 4. ככל הנראה היא נובעת מהשפעה של משתנים אחרים. במילים אחרות, נטייתם של גברים להיבחן יותר מנשים אינה נובעת מהיותם גברים אלא מכך שהוריהם משכילים יותר, מצבם הסוציאקונומי טוב יותר וכדומה. ראה בהמשך על השפעתם של גורמים אחרים.
- גיל: ככל שגיל הנבחן בהיבחנות הראשונה גדול יותר, הסיכוי שהוא ייבחן פעמיים לפחות קטן יותר (יחס סיכויים של 0.731 לכל שנה).
- ככל שהציון בהיבחנות הראשונה של נבחן גבוה יותר, הסיכוי שהוא ייבחן פעמיים לפחות קטן יותר (יחס סיכויים של 0.996 לכל נקודה בבחינה הראשונה).
- ככל שהשכלת אב הנבחן גבוהה יותר, הסיכויים שהוא ייבחן פעמיים לפחות גדולים יותר (יחס סיכויים של 1.041 לכל עלייה ב־1 בסולם רמות ההשכלה).
- ככל שהשכלת אם הנבחן גבוהה יותר, הסיכויים שהוא ייבחן פעמיים לפחות גדולים יותר (יחס סיכויים של 1.010 לכל עלייה ב־1 בסולם רמות ההשכלה).
- ככל שהמצב הסוציאקונומי של הנבחן גבוה יותר, הסיכויים שהוא ייבחן פעמיים לפחות גדולים יותר (יחס סיכויים של 1.075 לכל עלייה ב־1 בסולם הרמות הסוציאקונומיות).

¹³ במפתיע, עצם העובדה שיותר נשים נבחנות בכל קטגוריה אינה מעידה על נטייתן וסיכוייהן להיבחן פעמיים או יותר. ההסבר מצוי בטבלה הבאה (ספירה לפי אנשים):

	אישה	גבר
נבחן פעם אחת בלבד	114,308 (55.3%)	92,310 (44.7%)
נבחן לפחות פעמיים	72,147 (55.7%)	57,428 (44.3%)

ההסתברות שגבר ייבחן פעמיים לפחות היא זו: $0.3835 = \frac{57,428}{92,310 + 57,428}$; ובהתאם הסיכוי הוא: $0.622 = \frac{0.3835}{1 - 0.3835}$
ההסתברות שאישה תיבחן פעמיים לפחות היא זו: $0.3869 = \frac{72,147}{114,308 + 72,147}$; ובהתאם הסיכוי הוא: $0.631 = \frac{0.3869}{1 - 0.3869}$
לפיכך יחס הסיכויים הוא $1.014 = \frac{0.631}{0.622}$ לטובת אישה!

5.1.3. מודל רגרסיה לחיזוי שיפור הציון בהיבחנות השנייה

תת-פרק זה מסכם את ממצאי השימוש במודל הרגרסיה לחיזוי השיפור הציון בהיבחנות השנייה (המודל מתואר בתת-פרק 3.7 והתוצאות מובאות בתת-פרק 4.4).

- מין: נשים משפרות את ציון פחות מגברים (השיפור קטן ב-3.78 נקודות). (נזכר כי לפי הקידוד שלנו המין מקודד כזכר=1 ונקבה=2).
- גיל: ככל שגיל הנבחן בהיבחנותו הראשונה מבוגר יותר, השיפור הממוצע בציונו קטן יותר (ב-1.70 נקודות לכל שנה).
- משתנים דמוגרפים: ככל שהורי הנבחן (הן האב והן האם) משכילים יותר ומצבו הסוציאקונומי טוב יותר, השיפור הממוצע בציונו גדול יותר (0.88/0.84/2.92 נקודות לכל רמת השכלת אב, השכלת אם ומצב סוציאקונומי בהתאמה).
- הממצאים הנוגעים להשכלת ההורים מנוגדים לאלו שנמצאו בעזרת מודל הצמיחה, וייתכן שהדבר נובע מהכנסת משתנים אחרים למודל. על כל פנים, גודל האפקט כאן קטן מאוד.

כאשר מכניסים למודל את המשתנים "זמן מאז ההיבחנות הראשונה" ו"ציון בהיבחנות הראשונה" בלויית ריבועיהם הם יוצרים פרבולות הפוכות, כלומר פונקציות בעלות מקסימום, משום ש- a , המקדם של המשתנה המועלה בריבוע, קטן מאפס. לכן אפשר לחשב את הנקודה המרבית של נתוני המחקר בעזרת הנוסחה לחישוב נקודת מקסימום של פרבולה: $x_{max} = \frac{-b}{2a}$ $y = ax^2 + bx \Rightarrow$

- עבור המשתנה "זמן מאז ההיבחנות הראשונה": מרב השיפור מופיע בערך של 9.12 שנים. מאחר שאצל 99% מהאוכלוסייה הזמן שבין ההיבחנות הראשונה לשנייה קצר מערך זה (ראו לוח 3), עבור כמעט כל הנבחנים ככל שעובר יותר זמן מהבחינה הראשונה השיפור גדול יותר.

○ נתון זה מעיד כי מרכיב חשוב בשיפור הציון בהיבחנות חוזרת הוא צמיחתם האישית של הנבחנים (רכישת ידע וניסיון בבית הספר או מחוצה לו) בזמן שחלף בין היבחנות להיבחנות. זאת משום שאילו נבע כל השיפור בציון בהיבחנות חוזרת מהחשיפה לבחינה הראשונה, היה צפוי שבחלוף הזמן ילך השיפור ויקטן.

- עבור המשתנה "ציון בהיבחנות הראשונה": מרב השיפור מופיע בציון 465. פירוש הדבר הוא שעד לציון זה בהיבחנות הראשונה השיפור בציון בהיבחנות השנייה גדל, ומציון זה ומעלה השיפור בציון בהיבחנות השנייה קטן.

5.2. השוואת המחקר הנוכחי למחקרי עבר

ככלל, מחקרי העבר כמעט שלא עסקו בהשפעתם של משתני רקע על דפוס ההיבחנות החוזרת (סטולר ואללוף עסקו רק בציון בהיבחנות הראשונה ובמרווח הזמן שחלף מאז ההיבחנות הראשונה, ובן-סימון וכהן הוסיפו לשני אלה גם את משתנה המין). שני המחקרים הללו דנו במשתנים אלו בעזרת סטטיסטיקה תיאורית, חלוקה למקבצים או חישוב קתאמים, וחסר בהם מידול סטטיסטי מדויק המפקח על שאר המשתנים.

מהשוואה בין תוצאות המחקר הנוכחי למחקרי עבר במאלי"ו (בהקשר זה ראו את פרק 2.2) עולים ממצאים בולטים אלו:

- המחקרים הקודמים הראו שאף שציונם הפסיכומטרי של הנבחנים החוזרים משתפר בהיבחנותם השנייה, ציונם המשופר נופל במוצע מציונם של אלו הנבחנים פעם אחת בלבד.¹⁴ מחקר זה מראה שהמצב השתנה: אומנם ציונם של נבחנים חוזרים בהיבחנות הראשונה נמוך מזה של אלו הנבחנים פעם אחת בלבד, אך בהיבחנותם השנייה ציונם גבוה יותר מזה של אלו הנבחנים פעם אחת בלבד.
- שני מחקרים (קנת-כהן ואורן ובן-סימון וכהן) הצביעו על שיפור של 36.6 ו-32.5 נקודות בהתאמה בהיבחנות השנייה, שיפור הנמוך מזה שנמצא במחקר הנוכחי – 45.1 נקודות. סטולר ואללוף מצאו שיפור של 49.1 נקודות, שיפור הגבוה מן השיפור שנמצא במחקר זה. במחקרם של אורן וכהן, שהתמקד באוכלוסייה קטנה וייחודית, התקבל ממוצע שיפור של 44.3, ערך הדומה לערך שהתקבל במחקר זה. רק בן-סימון וכהן חקרו את השיפור בין ההיבחנות השנייה לשלישית, והם מצאו כי הוא שווה ל-23.8 נקודות, ערך הקרוב מאוד לערך שהתקבל במחקר הנוכחי (25.6).
- הממצאים שלהלן נוגעים להשוואה בין המחקר הנוכחי ובין מחקרים של סטולר ואללוף לגבי משתני הרקע "ציון בהיבחנות הראשונה" ו"זמן מאז ההיבחנות הראשונה".
 - ציון בהיבחנות הראשונה: במחקרם של סטולר ואללוף הופיע השיפור המרבי אצל נבחנים שציונם בהיבחנות הראשונה היה 450–550. במחקר הנוכחי הופיע השיפור המרבי אצל נבחנים שציונם בהיבחנות הראשונה היה 465, ערך המצוי בטווח זה.
 - זמן מאז ההיבחנות הראשונה: סטולר ואללוף הסיקו כי ככל שמרווח הזמן בין ההיבחנות ארוך יותר, כך השיפור גדול יותר. במחקר הנוכחי התקבלה מסקנה זהה.
- הממצאים שלהלן נוגעים להשוואה בין המחקר הנוכחי ובין מחקרים של בן-סימון וכהן לגבי משתני הרקע "מין", "ציון בהיבחנות הראשונה" ו"זמן מאז ההיבחנות הראשונה":
 - מין: שני המחקרים הסיקו כי גברים נוטים לחזור על הבחינה יותר מנשים. בן-סימון וכהן הסיקו זאת מניתוח הסטטיסטיקה התיאורית: הם מצאו כי מספר הגברים שנבחנים פעמיים או שלוש פעמים, בהתאמה, נוסף על כך, הם מצאו כי אחוז הגברים בקרב הנבחנים פעמיים גבוה במובהק מאחוז הנשים בקבוצה זו, בהתחשב ביחס בין נשים וגברים באוכלוסיית הנבחנים כולה (וכך גם לגבי הנבחנים שלוש פעמים).¹⁵ במחקר הנוכחי הסטטיסטיקה התיאורית מראה את ההפך: נמצא כי מספר הנשים שנבחנו פעמיים עולה על מספר הגברים שנבחנו פעמיים, וכך גם לגבי הנבחנים שלוש פעמים. אף על פי כן, כאמור, המחקר הנוכחי מסיק שגברים נוטים להיבחן פעמים

¹⁴ מחקרים של סטולר ואללוף מציג רק את השיפור בהיבחנות השנייה, ואינו מביא בחשבון את הנבחנים פעם אחת בלבד.

¹⁵ את החישוב המוצג בהערת שוליים 13 אפשר להפעיל על תוצאות טבלה מס. 16 במחקרם של בן-סימון וכהן. חישוב שכזה מראה כי יחס הסיכויים להיבחנות יותר מפעם אחת בין נשים לגברים הוא 0.78. פירוש הדבר הוא שלפי הסטטיסטיקה התיאורית ובלי לפקח על משתנים אחרים סיכוייה של אישה להיבחן בבחינה חוזרת הוא 0.78 מסיכויי של גבר לעשות זאת – תוצאה הזוהה לתוצאות המחקר הנוכחי, הכולל פיקוח על משתנים אחרים.

רבות יותר מנשים, וזאת בהתחשב בתוצאותיו של מודל סטטיסטי המפקח על משתני רקע אחרים (בהקשר זה ראו הערת שוליים 13).

שלא כבמחקרם של בן-סימון וכהן, במחקר הנוכחי נמצא הבדל מובהק בין מידות השיפור בציון אצל נשים וגברים: אצל נשים השיפור היה קטן יותר.

○ ציון בהיבחנות הראשונה: במחקרם של בן-סימון וכהן לא נמצא שיפור מרבי הנגזר מהציון בהיבחנות הראשונה, אלא נמצאה ירידה עקבית בשיפור ככל שהציון בהיבחנות הראשונה גבוה יותר. במחקר הנוכחי נמצא שהציון 465 בהיבחנות הראשונה גורר אחריו שיפור מרבי, כלומר עד ציון זה בהיבחנות הראשונה השיפור גדל, ומציון זה ומעלה השיפור קטן.

○ זמן מאז ההיבחנות הראשונה: לרוב לא נמצא במחקרם של בן-סימון וכהן קשר מובהק בין מרווח הזמן לבין גודל השיפור. המתאמים המובהקים המועטים שנמצאו בו מצביעים על קשר שלילי, כלומר על כך שהזמן פועל לרעתו של הנבחן: ככל שמרווח הזמן בין היבחנות ארוך יותר, השיפור בציון קטן יותר. ממצא זה מנוגד לממצא של המחקר הנוכחי, ולפיו שככל שמרווח הזמן בין היבחנות ארוך יותר, השיפור בציון גדול יותר.

תוצאות מחקרם של בן-סימון וכהן שונות מתוצאות המחקר הנוכחי בהיבטים מרובים יותר מתוצאות שאר המחקרים הקודמים. מחקרם זה נערך לפני קרוב לשלושים שנה, ונראה שדפוסי ההיבחנות החוזרות השתנו מאז במידה ניכרת.

5.3. מגבלות המחקר והצעות למחקרי המשך

כפי שראינו בתת-פרק 3.3 העוסק בחסרים במחקר ובפרט בלוח 1, המשתנים הדמוגרפיים (השכלת האב, השכלת האם והמצב הסוציאקונומי) והמשתנה בגרות חסרו ברשומות רבות. את המשתנים הדמוגרפיים השלמנו במידת האפשר (קודם השלמות פשוטות ולאחר מכן השלמות מרובות), ואילו על המשתנה בגרות, שחסר מאוד, נאלצנו לוותר. במחקרים עתידיים מומלץ לאסוף נתונים רבים יותר הנוגעים למשתנים אלו כך שאפשר יהיה לכלול את המשתנה בגרות במחקר ועל מנת להפחית את אחוז ההשלמות. ראוי לציין כי המשתנים הדמוגרפיים מתבססים על דיווח עצמי, ולא ברור עד כמה הם אמין; התשובה לשאלת המצב הסוציאקונומי סובייקטיבית במיוחד באופייה. בהתאם לכך, במחקרים עתידיים מומלץ להשיג נתונים אלו ממקור חיצוני, כגון מאגר המידע של הלשכה המרכזית לסטטיסטיקה (הלמ"ס).

מחקר זה היה כרוך בניתוחים סטטיסטיים מורכבים, ועל כן התמקד באוכלוסיית הנבחנים בשפה העברית בלבד ובניתוח נתוני הציון הכללי בבחינה הפסיכומטרית בלבד. מומלץ לערוך מחקרים נוספים שיעסקו גם באוכלוסיות נבחנים אחרות, כמו הנבחנים בשפה הערבית, שהיא האוכלוסייה השנייה בגודלה, וכן יעסקו בניתוח ציוני התחומים בבחינה הפסיכומטרית: חשיבה כמותית, חשיבה מילולית ואנגלית.

כמוסבר במבוא, מחקר זה מתמקד בניתוח דפוס ההיבחנות החוזרות בלי לדון בשאלת הוגנות השימוש בציוניהן. יש מקום למחקר עתידי שישלב את חקר ההיבחנות החוזרות עם חקר השפעתן על יעילות השימוש בציוני הנבחנים החוזרים ועל הוגנותו.

6. מקורות

אורן, כ. וכהן א. (1988). *הוגנות השימוש בבחינה הפסיכומטרית השנייה של נבחנים חוזרים בקרב תלמידי מכינת הטכניון* (דוח מס' 67). ירושלים: מרכז ארצי לבחינות ולהערכה.

בן-סימון, ע., וכהן י. (1988). השפעה של היבחות חוזרת על הישגים במבחנים פסיכומטריים. ב: נבו, ב., וכהן, י. (עורכים). *בעיות נבחרות בהערכה ובמידה*, פרק 5, ירושלים: מרכז ארצי לבחינות ולהערכה.

סטולר, ר. ואללוף, א. (2002). *שיפור בציוני בחינה חוזרת בקרב נבחנים כפולים בשמות בחינה: עברית, ערבית ורוסית (1995-2001)* (דוח טכני פנימי מס' 137). ירושלים: מרכז ארצי לבחינות ולהערכה.

סער, י. ואורן, כ. (2014). דוח סטטיסטי לשנים 1999-2013. ירושלים: מרכז ארצי לבחינות ולהערכה.

קלפר, ד., אללוף, א., טורוול, א., אורן, כ. ופרונטון, מ. (2015). *בקרת איכות דו-רמתית על ציונים לפי משתנים דמוגרפיים* (דוח מס' 415). ירושלים: מרכז ארצי לבחינות ולהערכה.

קנת-כהן, ת. ואורן, כ. (1989). על היעילות וההוגנות של הגדרות ציונים קובעים שונים, *בברירת נבחנים חוזרים במכפ"ל* (דוח מס' 110). ירושלים: מרכז ארצי לבחינות ולהערכה.

- Allison P.D. (2002). *Missing Data (Quantitative Applications in the Social Sciences)*. London: SAGE Publications, Inc. (Book 136).
- Fitzmaurice, G., Laird N. & Ware J. (2004). *Applied Longitudinal Analysis*. New York: Wiley.
- Freund, R., Littell C. & Stroup W.W. (2002). *SAS for Linear Models*, Cary, NC: SAS Institute Inc.
- Hedeker, D. & Gibbons R. (2006). *Longitudinal Data Analysis*. New York: Wiley.
- Jennrich, R. & Schluchter M. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*. Vol. 42: 805-820.
- Kenward, M.G. & Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. Vol. 53:983-997.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D. & Schabenberger, O. (2006). *SAS for Mixed Models, Second Edition*. Cary, NC: SAS Institute Inc.
- Liu, M., Lee, Y. H. & von Davier, A. A. (2012). *Detection of unusual administrations using a linear mixed effects model*. Paper presented at the international meeting of the Psychometric Society, Lincoln, NE.
- Rubin, D. (1976). Inference and missing data. *Biometrika*. Vol. 63: 581-592.
- Rubin, D. (1987). *Multiple Imputation for Nonresponse in Survey*. New York: Wiley.
- SAS Institute Inc. (2004). *SAS/STAT® 9.1 User's Guide*. SAS Institute Inc.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
- Wang, J., Xie, H. & Fisher, J.H. (2012). *Multilevel Models: Applications Using SAS*. Berlin: Walter de Gruyter.

7.1. השלמות בשיטת זקיפה מרובה

זקיפה מרובה (MI=Multiple Imputation, Rubin, 1976, 1978; Schafer, 1997) הינה שיטה מתוככמת להשלמת נתונים חסרים. הרעיון של השיטה הוא להחליף ערכים חסרים עם מספר ערכים אפשריים, ואז לנתח את בסיס הנתונים "השלם" בשיטות רגילות. השיטה שונה בצורה מהותית מזו שמבצעת השלמה בודדת בצורה המסורתית. השלמה בודדת משלימה ערך יחיד עבור כל ערך חסר, ובהתאם לכך, שיטת השלמה זו אינה משקפת את אי-הוודאות שקיימת בחיזוי ההשלמה של הערך החסר, וכתוצאה מכך טעויות התקן עבור הפרמטרים הנאמדים מן המודל מוטות כלפי מטה. בניגוד להשלמה בודדת, זקיפה מרובה הינה שיטת מונטה קרלו היוצרת בעזרת סימולציה מדגם מקרי או התפלגות של הערכים החסרים. פרוצדורה זו גורמת לכך שהשימוש בהסקה סטטיסטית ישקף בצורה אמינה את אי-הוודאות הקיימת בקיום הערכים החסרים. המודל המקובל הוא המודל הנורמלי הרב מימדי שמניח כי:

❖ לכל המשתנים יש התפלגות נורמלית

❖ כל משתנה יכול להיות מיוצג כפונקציה לינארית של שאר המשתנים עם טעות

הומוסקדסטית נורמלית

למרות ההנחות החזקות של המודל, נמצא כי בפועל המודל "עובד" גם אם התנאים מופרים. נתאר להלן את הפרוצדורה לביצוע ההשלמה המשתמשת באלגוריתם של הרחבת נתונים (data augmentation) בשישה שלבים¹⁶:

1. בחר ערכים התחלתיים עבור הפרמטרים שבמקרה של המודל הנורמלי הם הממוצעים ומטריצת השונות המשותפות. ערכים התחלתיים אלו יכולים להתקבל ע"י שימוש בשיטות פשוטות יותר כמו מחיקת שורה (listwise deletion), מחיקה בזוגות (pairwise deletion) ועדיף, בשיטה של אלגוריתם ציפיה - מקסום (EM algorithm).
2. השתמש בערכים התחלתיים אלו לקבלת אומדנים למקדמי הרגרסיה במשוואות שבהן כל אחד מהמשתנים עם הערכים החסרים הינו פונקציה של כל שאר המשתנים. דבר זה יש לעשות עבור כל תבנית של חוסר בנתונים.
3. השתמש במשוואות הרגרסיה כדי לחזות את הערכים החסרים. לכל ערך חזוי הוסף טעות מקרית מההתפלגות הנורמלית של השארית.
4. בעזרת שימוש בבסיס הנתונים ה"שלם" חשב מחדש את הממוצעים ומטריצת השונות המשותפות.
5. בעזרת הממוצעים ומטריצת השונות המשותפות החדשים שחושבו, הגרל ערכים חדשים של ממוצעים ושונות משותפת מתוך ההתפלגות החדשה (posterior distribution).
6. חזור לצעד 2 עם הערכים החדשים שהתקבלו ובצע את תהליך האיטרציה הזה עד להתכנסות (מספקת) של התהליך (נשים לב שמדובר בהתכנסות הן של המקדמים והן של ההתפלגות). ההשלמה שמתקבלת בשלב הסופי באיטרציה הינה ההשלמה המבוקשת.

¹⁶ בתיאור המודל אנו עוקבים אחרי הפיתוח המופיע ב- Allison, 2002.

הערה: כדי לקבל את ההתפלגות החדשה יש צורך בהתפלגות טרומית (prior distribution). בדרך כלל מניחים התפלגות טרומית "ללא מידע" ("noninformative" prior). בגלל אופי התהליך, בכל פעם מתקבלת השלמה שונה, ולכן נהוג לעשות מספר השלמות (ומכאן השם: "השלמה מרובה"). במקרה שלנו בוצעו 5 סטים של השלמות (שהינו ברירת המחדל של התוכנית בה השתמשנו – ראה מיד). כדי לקבל אומדן אחד מחשבים ממוצע פשוט של האומדנים שהתקבלו. לעומת זאת על מנת לחשב אומדן אחד לטעות התקן משתמשים בנוסחה הבאה (עבור M שכפולים):

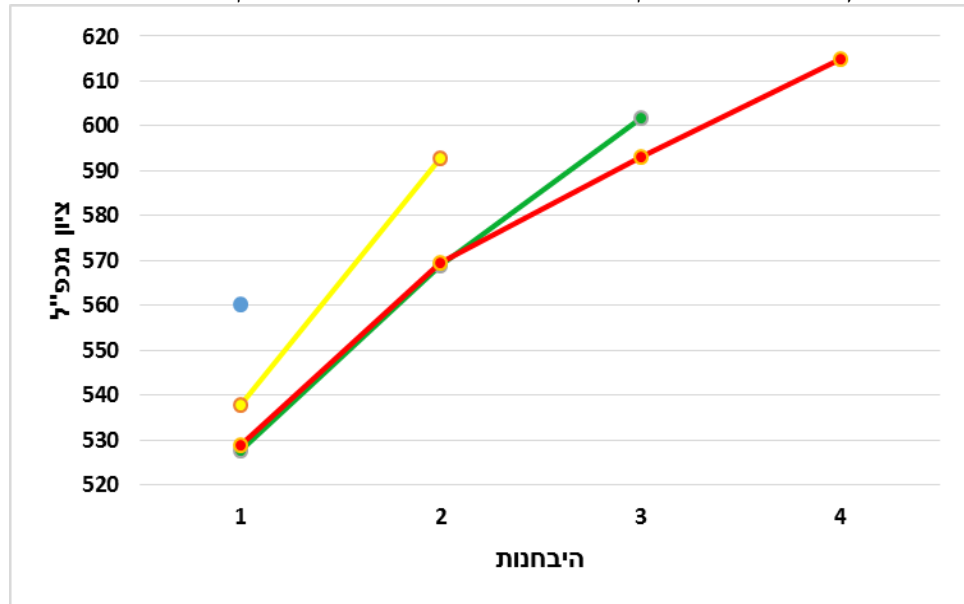
$$S.E.(\bar{\theta}) = \sqrt{\frac{1}{M} \sum_{k=1}^M s_k^2 + \left(1 + \frac{1}{M}\right) \left(\frac{1}{M-1}\right) \sum_{k=1}^M (\theta_k - \bar{\theta})^2}$$

כדי ליישם השיטה אנו השתמשנו בשתי פרוצדורות שעושות את כל זאת בצורה אוטומטית ב-SAS בעזרת שימוש ב: PROC MI ו־PROC MIANALYZE.

7.2. הצגות גרפיות של לוחות 4-7

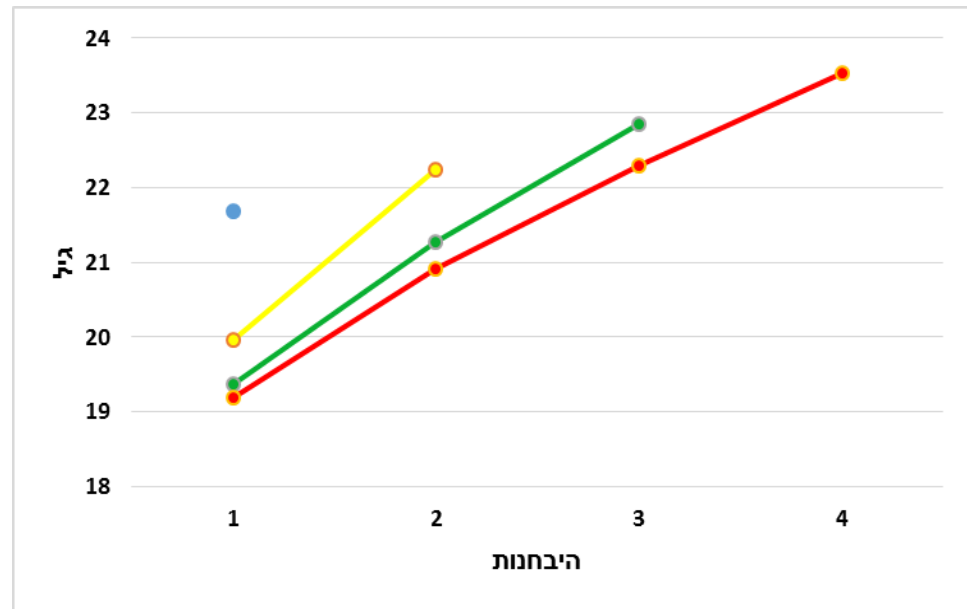
באיורים 2-5 מוצגת הסטטיסטיקה התיאורית של ציון המכפ"ל הממוצע, הגיל הממוצע (לוח 4), השכלת האב הממוצעת (לוח 5), השכלת האם הממוצעת (לוח 6), ומצב סוציאקונומי (לוח 7) בהתאמה כפונקציה של מספר ההיבחנות ולפי מספר מקסימלי של היבחנות.

איור 2 – ציון מכפ"ל הממוצע כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית



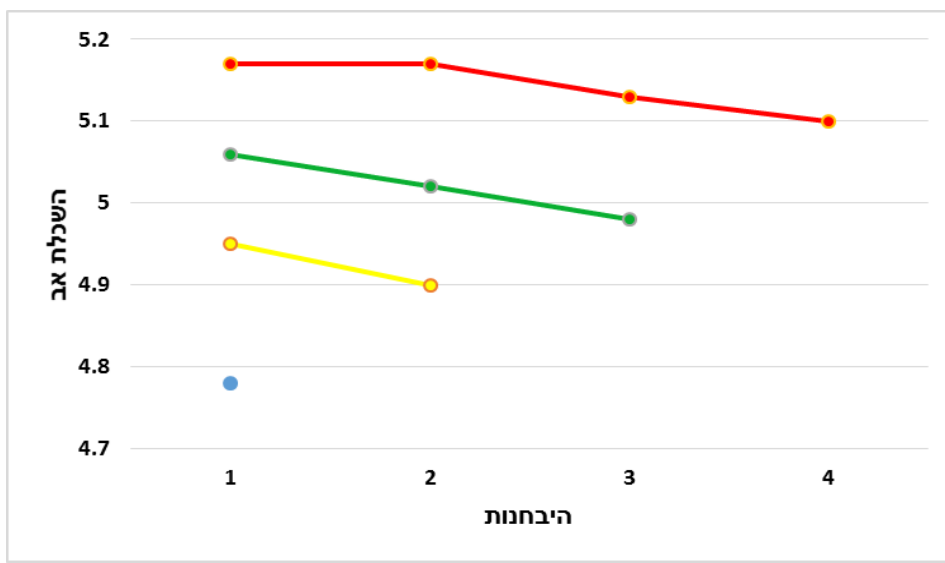
- מקסימום 1 היבחנות
- מקסימום 2 היבחנות
- מקסימום 3 היבחנות
- מקסימום 4 היבחנות

איור 3 – הגיל הממוצע כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית

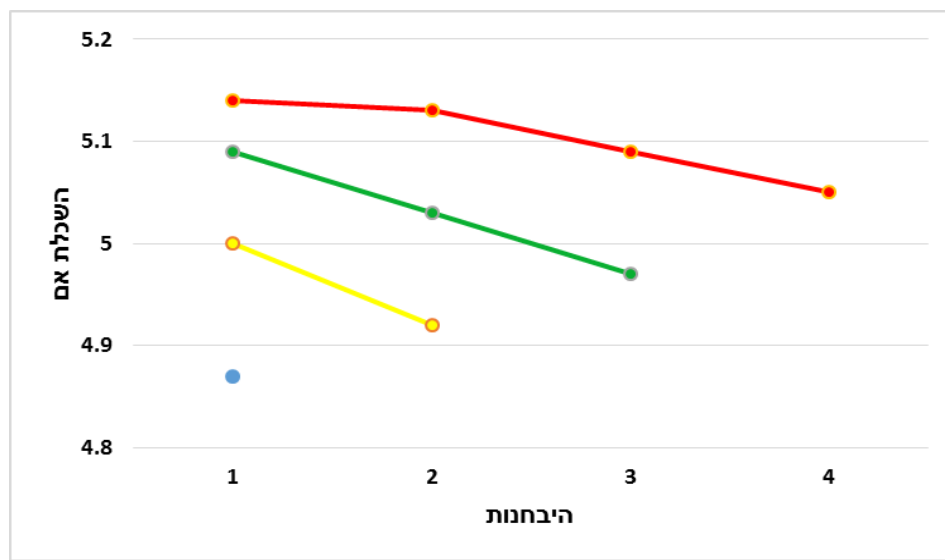


- מקסימום 1 היבחנות
- מקסימום 2 היבחנות
- מקסימום 3 היבחנות
- מקסימום 4 היבחנות

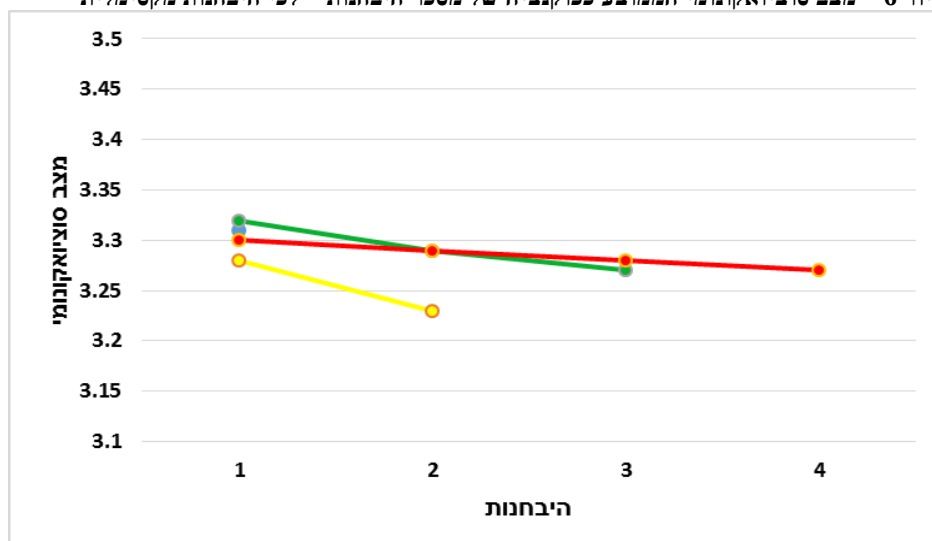
איור 4 – השכלת האב הממוצעת כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית



איור 5 – השכלת האם הממוצעת כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית



איור 6 – מצב סוציאקונומי הממוצע כפונקציה של מספר היבחנות – לפי היבחנות מקסימלית



7.3. הצגה פורמלית של מודל צמיחה

כאמור, ניתן להתייחס אל נתוני אורך כנתונים היררכיים כאשר מדידות חוזרות משוכנות בתוך נבחן. שימוש במודל רב-רמתי מאפשר להתמודד עם התלות הקיימת בציונים השונים של אותו נבחן, עם השונות שקיימת בין נבחנים שונים ועם העובדה שמערך בסיס הנתונים אינו מאוזן (קיימים דפוסים שונים של היבחנויות בנתונים שברשותנו).

בנספח זה נתאר מודל דו-רמתי עבור מידול נתוני אורך. מודלים רב-רמתיים (multilevel) למידול נתוני אורך (longitudinal data) נקראים גם מודלים של צמיחה (growth models). כאמור הרמה הראשונה היא המדידות החוזרות והרמה השנייה היא האנשים.

אנו מתארים בנספח זה את המקרה הכללי כאשר ההתאמה למשתנים במחקר מובאת בגוף הדוח.

7.3.1. רמה ראשונה

נהוג למדל את הרמה הראשונה בהתאם למשוואה הבאה:

$$y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + e_{ij}$$

כאשר:

- y_{ij} הוא ההתרחשות ה- i ית (רמה ראשונה) עבור המדידה של אדם j (רמה שנייה)
- β_{0j}, β_{1j} הם חותך אקראי ושיפוע אקראי בהתאמה (המשמעות של האקראיות היא שהערך ההתחלתי וקצב השינוי שונים עבור אנשים שונים)
- t_{ij} הוא הזמן
- e_{ij} הוא טעות ומניחים שהוא מתפלג נורמלית עם תוחלת 0 ושונות σ^2

7.3.2. רמה שנייה

את הרמה השנייה ממדלים בהתאם לשתי המשוואות הבאות:

$$\beta_{0j} = \beta_{00} + \beta_{01}x_j + u_{0j}$$

$$\beta_{1j} = \beta_{10} + \beta_{11}x_j + u_{1j}$$

כאשר:

- β_{00}, β_{10} הם החותך הממוצע הכולל והשיפוע הממוצע הכולל (בהתאמה) מותנים בחזאי x_j (מרמה 2)
 - β_{01}, β_{11} הם מקדמי אפקטים קבועים עבור החזאי x_j
 - u_{0j}, u_{1j} הם מקדמי אפקטים אקראיים לאדם j על מקדמי החותך והשיפוע מרמה 1
- בהתאמה. מניחים שהם מתפלגים נורמלית עם תוחלת 0 ושונות $\sigma_{u0}^2, \sigma_{u1}^2$ בהתאמה

7.3.3. משוואה אחת

אם נציב את המודלים מרמה 2 לתוך רמה 1 נקבל משוואה אחת המייצגת את המודל:

$$y_{ij} = (\underbrace{\beta_{00} + \beta_{01}x_j + \beta_{10}t_{ij} + \beta_{11}x_jt_{ij}}_{\text{רכיב קבוע של המודל}}) + (\underbrace{u_{0j} + u_{1j}t_{ij} + e_{ij}}_{\text{רכיב אקראי של המודל}})$$

השארית
טעויות
מרמה 1
מרמה 2

7.3.4. כתיב מטריצות

באופן כללי נוכל לרשום בכתיב מטריצות:

$$Y = X\beta + ZU + e$$

כאשר: Y מסמל וקטור של תוצאות נצפות, β וקטור לא ידוע של אפקטים קבועים עם מטריצת עיצוב X (design) ידועה, U וקטור לא ידוע של אפקטים אקראיים עם מטריצת עיצוב Z , ולבסוף e הוא וקטור שאריות.

7.3.5. מידול מבנה השארית

המודלים שפורטו עד כה (ובפרט המודל שמופיע בתת-פרק 7.3.4) התמקדו במידול השונות בין אנשים ע"י הכנסת חותכים ושיפועים אקראיים, שיטה הנקראת מידול עם מקדמים אקראיים (random coefficient model, Littell et al., 2006).

אפשרות נוספת היא למדל את השונות בתוך אנשים ע"י כפיית מבנה שארית עבור מטריצת השונות המשותפת R . שיטה זו נקראת מידול עם מבנה שונות משותפת (Covariance pattern model, Fitzmaurice et al., 2004; Hedeker & Gibbons, 2006). הדבר בא לידי ביטוי כאשר במשוואת המודל הכללית

$$Y = X\beta + ZU + e$$

אנו ממדלים הפעם גם את e וזאת בנוסף למידול U , כך שאנו מניחים הפעם ש U, e מתפלגים

$$\text{נורמלית עם תוחלת } 0 \text{ ושונות } \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}. \text{Var} \begin{pmatrix} U \\ e \end{pmatrix}$$

המודל שבו מקובל להשתמש (ובו גם אנו נשתמש) עבור מידול המטריצה R הוא מודל חלקי שמיני 2 פרמטרים חופשיים והוא מידול אוטו-רגרסי מדרגה 1 המוכר בשם: AR(1). מודל זה מקובל כמידול של סדרות עתיות ומניח את מבנה השארית הבא:

$$R = \begin{pmatrix} \sigma & \rho & \rho^2 & \rho^3 \\ \rho & \sigma & \rho & \rho^2 \\ \rho^2 & \rho & \sigma & \rho \\ \rho^3 & \rho^2 & \rho & \sigma \end{pmatrix}$$

7.3.6. מודל צמיחה מקוטעת (ספליין)

באיור 1 ניתן להבחין למעשה בשני חלקי צמיחה בציון : צמיחה גדולה בין היבחנות ראשונה לשנייה וצמיחה מתונה יותר (ויחסית דומה) בין היבחנות 2 ל-4. במקרים שכאלו מקובל להתאים מודל מקוטע (piecewise growth curve) המוכר גם בשם ספליין (spline). על מנת לקבל את האפקט של השבירה לאחר הבחינה השנייה נקודת את מספר ההיבחנות בעזרת שני משתנים (היבחנות 1 + היבחנות 2) לפי הלוח הבא :

לוח 18 – קידוד המשתנים במודל צמיחה מקוטעת

משתני היבחנויות			מספר היבחנות בפועל
היבחנות 2	היבחנות 1	היבחנות	
0	0	0	1
0	1	1	2
1	1	2	3
2	1	3	4

כאשר העמודה היבחנות הינה המשתנה במודל ללא שבירה (שכאמור מקודד כאחד פחות מספר ההיבחנות בפועל). עבור המשוואה הבאה :

$$\text{ציון} = \text{חותך} + \beta_1 * \text{היבחנות 1} + \beta_2 * \text{היבחנות 2}$$

ובהתאם לקידוד שלנו נקבל (ראה איור 7 להמחשה) :

- החותך (כשמו) מציין את נקודת החיתוך עם ציר ה-y
- β_1 הוא השיפוע של המקטע הראשון
- β_2 הוא השיפוע של המקטע השני

