

מדינת ישראל
משרד ראש הממשלה



**מודלים סטוכסטיים לחיזוי מורים חדשים
במערכת החינוך בישראל
דוח מתודולוגי**

אלון שפירא
תחום ניתוח סטטיסטי
אגף מדען ראשי

ד"ר אלברט וקסלר
תחום ניתוח סטטיסטי
אגף מדען ראשי

ירושלים, נובמבר 2003

תוכן העניינים

3	מבוא	1
4	הגדרת משתנה מורה חדש	2
5	בניית רווחי סמך לפרופורציית מורים חדשים	3
8	טיפול בזקיפת נתונים חסרים	4
8	התאמת מודלים לוגיסטיים לאמידת פרופורציות מורים חדשים בשנים 2001-1995	5
10	זקיפת פרופורציות מורים חדשים בשנים 1994-1991	6
10	בדיקת שקילות המודלים	6.1
12	חישוב פרופורציות מורים חדשים לשנים 1994-1991	6.2
13	התאמת מודל אגרסיבי לחיזוי פרופורציות מורים חדשים לשנים 2009-2003	7
13	התאמת מודל Logit	7.1
14	בדיקת נורמליות	7.2
14	חיזוי מספר מורים לשנים 2007-2002	7.3
16	נספח מודלים ותרשימים	8

1 מבוא

דוח מתודולוגי זה מתאר את תהליך פיתוח מודלים לחיזוי מספר מורים חדשים במערכות חינוך בישראל. חיזוי מורים חדשים מהווה חלק בלתי נפרד מפרוייקט לחיזוי מצבת כח אדם בהוראה במערכות חינוך (ראה דוח מעגן ווקסלר 2002, וקסלר ושפירא 2003)

תהליך חיזוי מורים חדשים, בדומה לתהליך חיזוי כוח אדם בהוראה, התבסס על שימוש במודלים לוגיסטיים שהותאמו על סמך נתוני פרט של המורים במערכות החינוך בישראל המתייחסים לתקופה 1991-2001. המודלים הלוגיסטיים נבנו על סמך מאפיינים מקצועיים ואישיים של המורים כגון: מין, גיל, ותק, בסיס משרה, שעות הוראה וכד'. תהליך הפיתוח נשען על הנחת ההטרוגניות במערכות חינוך במגזרים שונים, בשלבי חינוך ובמחוזות גיאוגרפיים שונים. המודלים לחיזוי נבנו לשני מגזרים של מערכת החינוך – המגזר העברי והמגזר הערבי. במגזר העברי התחזיות ניתנו ל-18 תאים נפרדים (3 שלבי חינוך – חינוך יסודי, חטיבות ביניים וחטיבות עליונות ו-6 מחוזות גיאוגרפיים). במגזר הערבי חושבו התחזיות עבור שלושה שלבי חינוך, ללא חלוקה למחוזות גיאוגרפיים, וזאת עקב ריכוז מוסדות החינוך הערביים במחוז צפון. ניתוח פרופורציות מורים חדשים ופיתוח מודלים לאמידה וחיזוי מספר מורים חדשים התבצעו על בסיס קובץ נתוני מורים לשנים 1991-2001 שהועמד לרשותנו על ידי גף כח אדם בהוראה של הלשכה המרכזית לסטטיסטיקה. התחזיות ניתנו בפרופורציות (אחוזים) מתוך סך המורים בשנים 2002-2007 ובנפרד במספרי מורים החדשים שחושבו על סמך תחזיות סך כח אדם בהוראה (ראה דוחות מעגן ווקסלר 2002, וקסלר ושפירא 2003). בחלק המתודולוגי של פיתוח המודלים אימצנו הגדרת המושג "מורה חדש" המתאימה למצב בו המורה לא הופיע במשך חמש שנים במערכת החינוך. באופן כזה נמנענו מלהכליל בקבוצה זו את המורים הותיקים שעזבו לתקופה של מספר שנים וחזרו אליה. הגדרה זו סייעה לסנן את המורים העוזבים והשבים שאינם נחשבים חדשים מחד, אך מאידך, גרמה להקטנת מספר התצפיות האגרטיביות אודות המורים החדשים בשנים הראשונות. באופן כזה עמדו לרשותנו שבע תצפיות אגרטיביות של פרופורציות המורים החדשים המתייחסים לשנים 1995-2001 ועל מנת להשלים את התצפיות המתייחסות לשנים 1991-1995 השתמשנו במידע אודות מאפייני המורים בשנים הללו כדי לזקוף את נתוני מורים חדשים באמצעות מודלים לוגיסטיים ושיטת החיזוי לאחור (*backcasting*).

לאחר השלמת הנתונים החסרים אודות המורים החדשים הותאמו מודלים אגרטיביים לאמידה וחיזוי פרופורציות מורים חדשים כפונקציה של שנה קלנדרית. הדוח חולק למספר פרקים המתייחסים לשלבי הפיתוח בעבודת המחקר: סטטיסטיקה תיאורית של מורים חדשים במערכות החינוך, מתודולוגיה של אמידת פרופורציות מורים חדשים, זקיפת נתונים חסרים באמצעות מודלים לוגיסטיים, התאמת מודלים אגרטיביים לחיזוי פרופורציות מורים חדשים במערכות חינוך וחיזוי אחוז ומספר מורים חדשים לשנים 2002-2007.

2 הגדרת משתנה מורה חדש

בשלב הראשון לניתוח פרופורציות מורים חדשים במערכות חינוך הוגדר משתנה דיכוטומי המתאר את המצב של כל מורה על פני השנים הרלוונטיות מההיבט האם הוא חדש במערכת החינוך או לא. בקובץ כלל ארצי של המורים סומנו הופעותיהם של מורים במערכות חינוך בתקופה הנחקרת של 1991-2001 והוגדר משתנה מעקב אחר הצטרפותו של המורה n_{jkl} , $i = 1, K$, למערכת בשנה j , כאשר $j = 1991, K, 2001$, לתא $\{kl\}$, כאשר k מציין את שלב החינוך במערכת הכוללת, l מציין את המחוז הגיאוגרפי למגזר העברי בלבד¹, ו- n_{jkl} מציין את סך המורים בתא $\{kl\}$ בשנה j באופן כזה:

$$I_{kl} = \{C_{kl}\} \text{ אוסף של כל התאים במערכת החינוך } k = 1, K, 3, l = 1, K, 6,$$

יהי $(I_{new})_{ijkl}$ אינדיקטור מצבים "מורה חדש/מורה לא חדש" עבור מורה i , $i = 1, K, n_{jkl}$, בשנה j , לתא $\{kl\}$, $j = 1996, K, 2001, j$, המקבל ערכים כדלהלן:

1 - כאשר המורה i הופיע לראשונה במערכת לפחות בתא אחד $\{kl\}$ בשנה $j = 1995, K, 2001, j$ כלומר לא הופיע במערכת בתא $\{kl\}$ ולא באחד מהתאים $\{k'l'\}$, $\{k'l'\} \in \{C_{kl}\}$, $\{k'l'\} \neq \{kl\}$, בשנה j' , לכל $j' < j$, $j' = 1991, K, 2000$, כלומר המורה הוא חדש בתא $\{kl\}$ בשנה j , $j = 1996, K, 2001, j$.

0 - כאשר המורה i נמצא בתא $\{kl\}$ בשנה $j = 1995, K, 2001, j$ ונמצא בתא $\{kl\}$ או בכל תא אחר $\{k'l'\} \in \{C_{kl}\}$, $\{k'l'\} \neq \{kl\}$ לפחות פעם אחת בשנים $j' < j$, $j' = 1991, K, 2000$.

הערה: לצורך המחקר נקבע כי על מנת שמורה מסוים ייחשב כחדש בשנה מסוימת, עליו לא להופיע במערכת בשום תא לפחות בחמש שנים הקודמות לשנה שבה הופיע לראשונה. אי לכך, $(I_{new})_{ijkl}$ הוגדר רק החל משנת 1995, וזאת עקב זמינות הנתונים אודות המורים החל משנת 1991. נתוני מורים חדשים בשנים 1994-1991 הושלמו באמצעות תהליך זקיפה (ראה בהמשך).

¹ ראה הגדרות בדוח מודלים סטוכסטיים לחיזוי היקף כוח אדם בהוראה במערכת החינוך בישראל – וקסלר ושפירא, אפריל 2003.

3 בניית רווחי סמך לפרופורציית מורים חדשים

בשלב של ניתוח ראשוני של נתוני מורים חדשים היה מענייננו לבחון את הנתונים הקיימים ולבנות רווחי סמך לפרופורציית מורים חדשים הנצפות בשנים 1995-2001 על מנת להעריך באופן ראשוני את מגמת השינוי בהצטרפות מורים חדשים למערכות חינוך בישראל. על סמך הגדרת מצבים חדש/לא חדש חישבנו את פרופורציית המורים החדשים \hat{p}_{jkl}^{new} הנצפות בכל שנה $j = 1995, K, 2001$ בכל תא $\{kl\}$ באופן

הבא:

$$\hat{p}_{jkl}^{new} = \frac{1}{n_{jkl}} \sum_{i=1}^{n_{jkl}} (I_{new})_{ijkl} \quad (3.1)$$

בהנחה כי התצפיות של $(I_{new})_{ijkl}$ הן ב"ת שווי התפלגות ומתקיים:

$$(I_{new})_{ijkl} \sim \text{Bernoulli}(1, p_{newijkl}) \quad (3.2)$$

אומד את ההסתברות הממוצעת עבור מורה להיות חדש במערכת.

נבנה רווח סמך לפרופורציית מורים חדשים, כאשר האומד לשונות של הפרופורציה:

$$\widehat{Var}(\hat{p}_{jkl}^{new}) = \frac{\hat{p}_{jkl}^{new}(1 - \hat{p}_{jkl}^{new})}{n_{jkl}} \quad (3.3)$$

על סמך רווחי הסמך שניבנו לכל תא בשני המגזרים בחנו את השינויים על פני זמן בפרופורציית המורים החדשים. הקריטריון לשינוי של ממש בפרופורציית מורים חדשים התבסס על השוואת טווחים של רווחי סמך לאורך השנים והוגדר כדלהלן: השינוי (עליה/ירידה) בפרופורציית מורים חדשים במעבר בין שנה j לשנה \tilde{j} , $j, \tilde{j} = 1995, K, 2001$ נחשב כשינוי של ממש אם:

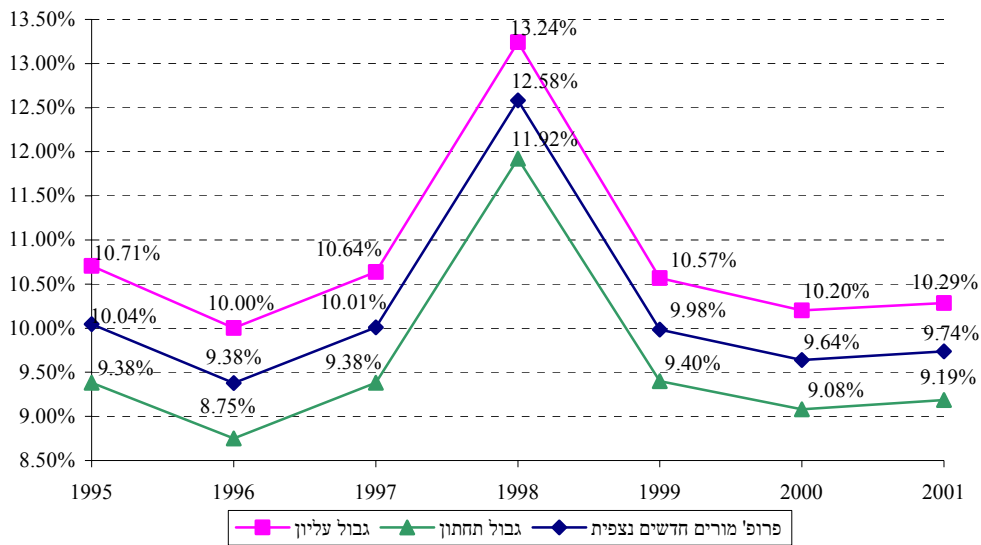
$$\left[\hat{p}_{jkl}^{new} - Z_{1-\alpha/2} \hat{s}_{jkl}, \hat{p}_{jkl}^{new} + Z_{1-\alpha/2} \hat{s}_{jkl} \right] \cap \left[\hat{p}_{\tilde{j}kl}^{new} - Z_{1-\alpha/2} \hat{s}_{\tilde{j}kl}, \hat{p}_{\tilde{j}kl}^{new} + Z_{1-\alpha/2} \hat{s}_{\tilde{j}kl} \right] = \emptyset \quad (3.4)$$

$$\alpha = 0.05, \hat{s}_{jkl} = \sqrt{\frac{\hat{p}_{jkl}^{new}(1 - \hat{p}_{jkl}^{new})}{n_{jkl}}} \quad \text{כאשר}$$

כלומר, אם לשני רווחי סמך לפרופורציית מורים חדשים בשנים j ו- \tilde{j} אין טווח משותף, אזי נאמר כי חל שינוי בפרופורציית מורים חדשים במעבר משנה j לשנת \tilde{j} . ניתוח זה מוגבל מבחינה סטטיסטית וודאי אינו מספיק לבדיקת השערות בדבר שינויים מובהקים על פני זמן, אך מספק הערכה ראשונית של סדרת נתונים לפני התאמת מודלים סטטיסטיים לחיזוי פרופורציית מורים חדשים.

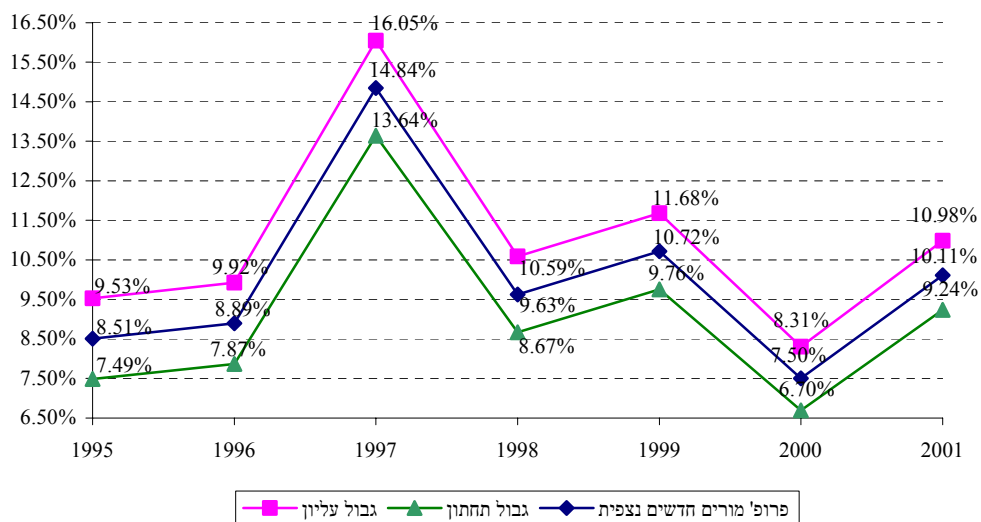
להלן מובאים מספר תרשימים לדוגמה המתארים רווחי סמך של פרופורציית מורים חדשים בתקופת 1995-2001.

תרשים 1: רווח סמך לפרופורציות מורים חדשים - מגזר ערבי - הינוך יסודי



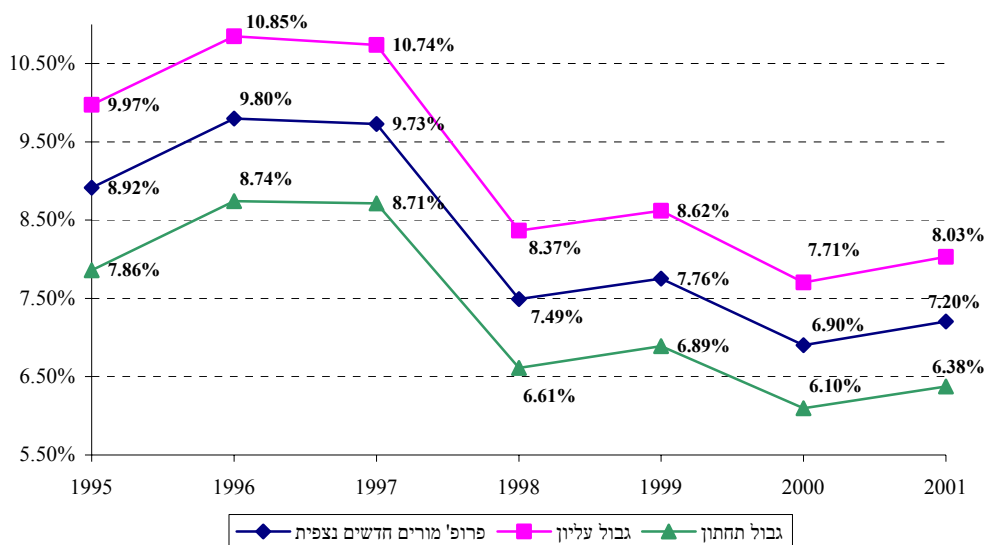
בתרשים 1 המתאר את פרופורציות המורים החדשים במגזר הערבי בחינוך היסודי לא ניתן לזהות מגמה או שינוי על פני זמן בפרופורציות מורים חדשים. כל רווחי הסמך, פרט לזה המתייחס לשנת 1998 הם בעלי טווח משותף. למרות הקפיצה בשנת 1998 אם מתבוננים על טווחי הפרופורציות לאורך השנים, נראה כי לכולם טווח משותף וגדול, כך שלא ניתן להצביע על מגמת עליה או ירידה בפרופורציות מורים חדשים על סמך נתונים אגרגטיביים.

תרשים 2: רווח סמך לפרופורציות מורים חדשים - מגזר ערבי - חטיבת ביניים



בתרשים 2 המתאר את חטיבות הביניים במגזר הערבי נראית תנודתיות לאורך השנים ולא ניתן לזהות מגמת עליה או ירידה ברורה. שנת 1997 נראית כחריגה מבין השנים הנחקרות, אך בהסתכלות לאורך התקופה כולה אין שינוי של ממש בפרופרציית המורים החדשים.

תרשים 3: רווח סמך לפרופורציות מורים חדשים - מגזר ערבי - חטיבה עליונה



בתרשים 3 ניתן לזהות שתי תקופות נפרדות : 1997-1995 ו- 2001-1998. רווחי הסמך בשתי התקופות אינם בעלי טווח משותף, ונראית ירידה משמעותית וחדה במעבר בין 1997 ו-1998.

טבלה 1: ניתוח ראשוני של מגמות שינוי בפרופורציית מורים חדשים במגזר הערבי

תרשים בנספח	מגמות שינוי	מחוז	שלב חינוך
3.1.1	ירידה משמעותית לאחר 1998	ירושלים	יסודי
3.2.1	ירידה משמעותית לאורך 1997-1995, התייצבות לאחר 2000	צפון	יסודי
3.3.1	ירידה משמעותית לאחר 1998 והתייצבות לאחר 1999	חיפה	יסודי
3.4.1.	ירידה הדרגתית לאורך השנים והתייצבות לאחר 1999	מרכז	יסודי
3.5.1	ירידה הדרגתית לאורך השנים והתייצבות לאחר 1999	תל אביב	יסודי
3.6.1	ירידה הדרגתית לאורך השנים והתייצבות לאחר 1999	דרום	יסודי
3.1.2	ירידה לאחר 1997 והתייצבות לאחר 1998	ירושלים	חט"ב
3.2.2	ירידה מתונה והדרגתית לאורך השנים	צפון	חט"ב
3.3.2	ירידה מתונה והדרגתית לאורך השנים	חיפה	חט"ב
3.4.2	תנודתיות חזקה ב-1997-1995, התייצבות לאחר 1997	מרכז	חט"ב
3.5.2	ירידה הדרגתית עם תנודתיות קלה	תל אביב	חט"ב
3.6.2	ירידה מתונה לאורך השנים	דרום	חט"ב
3.1.3	תנודתיות חזקה ללא מגמה ברורה	ירושלים	חט"ע
3.2.3	תנודתיות חזקה ללא מגמה ברורה	צפון	חט"ע
3.3.3	אין מגמה ברורה, יציבות לאורך השנים	חיפה	חט"ע
3.4.3	תנודתיות ללא מגמה ברורה	מרכז	חט"ע
3.5.3	תנודתיות ללא מגמה ברורה	תל אביב	חט"ע
3.6.3	תנודתיות ללא מגמה ברורה	דרום	חט"ע

בטבלה 1 מובאת תמצית ניתוח ראשוני של מגמת שינוי בפרופורציית מורים חדשים במגזר העברי. רוב התאים במגזר העברי המתייחסים לחינוך יסודי וחטיבות ביניים מאופיינים בירידה בפרופורציית מורים חדשים ובחטיבות עליונות לא ניתן להצביע על מגמה כלשהי של הנתונים לאורך השנים. בחלק מהתאים

נמצאו רווחי הסמך לאורך השנים בעלי טווח משותף, בשאר תאי חטיבות עליונות מסתמנת תנודתיות גבוהה. התרשימים המתאימים לנתוני פרופורציות מורים חדשים בכל התאים מופעים בנספח מודלים תרשימים (ראה פרק 8 לדוח זה).

4 טיפול בזקיפת נתונים חסרים

בקבצי מורים בהם נמצאו לפחות 10% ערכים חסרים במאפייני מורים נבנו מודלים של זקיפות. המאפיינים עבורם נבנו מודלים של זקיפות : בסיס משרה של המורה, מס' ילדים, דרגה, תפקיד, שעות הוראה וותק. תהליך זקיפת ערכים חסרים בכל מאפיין נעשה בכל שנת נתונים בנפרד על ידי התאמת מודל רגרסיה ליניארית. השיקולים לבחירת משתנים מסבירים פוטנציאליים במודל התבססו על הקשרים הטבעיים בין מאפייני המורים, כמו לדוגמה: גיל וותק, גיל ומספר ילדים, וכד'. כמו כן הוספנו את החזקה השניה והשלישית של משתנה גיל לצורך זיהוי קשרים לא ליניאריים. המשתנים הפוטנציאליים המסבירים מאפיין בעל ערכים חסרים הוכנסו לתהליך התאמת מודל רגרסיה ליניארית שהתבצעה בשלבים (stepwise). בטבלה 2 מובאת רשימת מאפיינים בהם נמצאו ערכים חסרים ורשימת משתנים פוטנציאליים שהשתתפו בתהליך התאמת המודלים.

טבלה 2: משתנים פוטנציאליים של המודלים הליניאריים לזקיפת ערכים חסרים של מאפייני מורים

משתנה מסביר	שערכיו נזקפים/ משתנה פוטנציאלי	גיל	גיל בריבוע	גיל מעוקב	מין	משרה	נשוי/ לא נשוי	ותק	תפקיד
ותק	X	X	X	X	X	X	X		
תפקיד	X	X	X	X	X	X		X	
בסיס משרה					X	X			
מס' ילדים		X	X	X					
דרגה	X	X	X	X	X	X		X	X

זקיפות ערכים חסרים נעשו כאמור לכל תא $\{kl\}$ לכל שנה $j = 1991, K, 2001, j$. המודלים הותאמו באמצעות רגרסיה ליניארית ע"פ קריטריון בחירת המשתנים מתוך וקטור משתנים פוטנציאליים המסבירים ברמת מובהקות לכל הפחות $\alpha = 0.05$ את המאפיין בו נמצאו ערכים חסרים.

5 התאמת מודלים לוגיסטיים לאמידת פרופורציות מורים חדשים בשנים

2001-1995

כאמור בפרק 2, ערכי אינדיקטור מצבים מורה חדש/לא חדש מתייחסים לשנים 2001-1995, לכן מספר התצפיות האגרטיביות של פרופורציות מורים חדשים הסתכם לשבע בלבד, דבר שהקשה על בניית מודל אגרטיבי והשגה של רמת מובהקות מספקת. אי לכך, היה עלינו לחפש דרך להשלים

סדרה של פרופורציות מורים חדשים בתקופת שנים 1994-1991 על מנת לבנות מודל אגרגטיבי ובאמצעותו לחזות פרופורציות מורים חדשים לשנים 2002-2007. על מנת לאמוד את פרופורציות המורים החדשים בכל תא הותאמו מודלים לוגיסטיים לאמידת ההסתברות של מורה מסוים להיות חדש במערכת באמצעות מאפייניו המקצועיים והאישיים והאינטראקציות (ראה טבלאות 3 ו-4) בין מאפיין x_m ומאפיין x_n , $m \neq n$, $m = 1, K, N$, $n = 1, K, N$, כאשר N הינו המספר הכולל של מאפייני מורים בקובץ המורים. האינטראקציות הוכנסו למודלים כמשתנים מסבירים פוטנציאליים על פי קריטריון קיום מתאם משמעותי בין המאפיינים $\rho_{x_m x_n} \geq 0.2$ ברמת מובהקות $\alpha = 0.05$. אם כן, לכל שנה $j = 1995, K, 2001, j$ ההסתברות לכל מורה i להיות חדש בתא $\{kl\}$ בשנה j הנה:

$$\hat{p}_{ijkl}^{new} = \left(1 + \exp\left(-X_{ijkl}^T \beta_{jkl}\right)\right)^{-1} \quad (5.1)$$

לאחר הרצת הרגרסיות הלוגיסטיות נתקבלו וקטורי המקדמים $\hat{\beta}_{jkl}$, מובהקים ברמה של $\alpha = 0.05$, עבור וקטור המאפיינים X_{ijkl}^T , המסבירים את ההסתברויות \hat{p}_{ijkl}^{new} .

טבלה 3: רשימת מאפיינים ששימשו משתנים מסבירים פוטנציאליים במודלים הלוגיסטיים לאמידה וחיזוי פרופורציות מורים חדשים – מגזר ערבי

שם המשתנה בקובץ	תאור המשתנה
BasMis	חלקיות משרה של מורה ביחס לשעות הוראה
Child	מספר הילדים במשפחה
Darga	דרגת שכר של המורה
Kafull	האם המורה מלמד ביותר משלב חינוך אחד בו זמנית
Gil	גיל המורה
Misra	מס' בי"ס בהם מלמד המורה
Nasui	משתנה עזר נשוי/אחר
New	מורה בעל ותק 1
Nihul	משתנה עזר תפקיד ניהולי/אחר
Sex	מין המורה
Shaot	סך שעות הוראה של המורה
Tafkd	תפקיד המורה (מרכז, מורה בכיר וכד')
Vetek	ותק המורה בשנים
MozaAv	מוצא אב של המורה (למגזר עברי בלבד)

טבלה 4: משתני אינטראקציה פוטנציאליים במודלים לאמידה וחיזוי פרופורציות מורים חדשים – מגזר ערבי

שם המשתנה	האינטראקציה	שם המשתנה	האינטראקציה
<i>bagi</i>	<i>Basmis*gil</i>	<i>shata</i>	<i>Shaot*tafk</i>
<i>baska</i>	<i>kafulI*Basmis</i>	<i>shave</i>	<i>vetek * shaot</i>
<i>baschi</i>	<i>basmis*child</i>	<i>veta</i>	<i>vetek*tafk</i>
<i>gich</i>	<i>gil*child</i>	<i>dag</i>	<i>darga*gil</i>
<i>basnew</i>	<i>basmis*new</i>	<i>nach</i>	<i>nasui*child</i>
<i>shachi</i>	<i>shaot*child</i>	<i>basni</i>	<i>basmis*nihul</i>
<i>chita</i>	<i>child*tafk</i>	<i>gini</i>	<i>gil*nihul;</i>
<i>vetchi</i>	<i>vetek*child</i>	<i>vetni</i>	<i>vetek*nihul</i>
<i>dataf</i>	<i>darga*tafk</i>	<i>nenas</i>	<i>new*nasui</i>
<i>gin</i>	<i>gil*new</i>	<i>china</i>	<i>child*nasui</i>
<i>shagi</i>	<i>shaot*gil</i>	<i>ginas</i>	<i>gil*nasui</i>
<i>gita</i>	<i>gil*tafk</i>	<i>basex</i>	<i>basmis*sex</i>
<i>newsha</i>	<i>new*shaot</i>	<i>gisex</i>	<i>gil*sex</i>
<i>nevet</i>	<i>new*vetek</i>	<i>shkaf</i>	<i>shaot*kafulI</i>
<i>sesha</i>	<i>sex*shaot</i>	<i>nisha</i>	<i>nihul*shaot</i>
<i>setaf</i>	<i>sex*tafk</i>	<i>basta</i>	<i>basmis*tafk</i>
<i>sevet</i>	<i>sex*vetek</i>	<i>basve</i>	<i>basmis*vetek</i>
<i>dav</i>	<i>darga*vetek</i>	<i>vetgi</i>	<i>vetek*gil</i>
<i>Pichi[#]</i>	<i>pikuah*child</i>	<i>mozg*</i>	<i>mozaav*gil</i>
<i>Pidar[#]</i>	<i>pikuah*darga</i>	<i>pikaf*</i>	<i>pikuah*kafulI</i>

- אינטראקציות במגזר העברי בלבד

6 זקיפת פרופורציות מורים חדשים בשנים 1991-1994

זקיפת פרופורציות המורים החדשים בשנים 1991-1995 נעשתה באמצעות אומדנים שחושבו מתוך המודלים הלוגיסטיים שהותאמו לשנים 1995-2001. ברור כי כל אחד מהמודלים שהותאמו לשנים 1995-2001 סיפק אומדן משלו לפרופורציות המורים החדשים בשנים 1991-1994. על מנת להשתמש במלוא האינפורמציה אודות המורים החדשים בשנים 1995-2001 ולספק אומדנים טובים יותר, היה עלינו לבדוק את האם תרומתו של כל מודל לחיזוי לאחור (*backcasting*) של פרופורציות מורים חדשים בשנים הראשונות הינה שווה, או במילים אחרות האם המודלים בשנים המתקדמות שקולים זה לזה מבחינת יכולתם לאמוד נתון חסר מהשנים הראשונות.

6.1 בדיקת שקילות המודלים

על מנת לבצע בדיקת "שקילות" המודלים שנתקבלו בפרק 5 בחרנו לאמוד את פרופורציית מורים חדשים בשנה הראשונה, בה נתון זה היה זמין, כלומר בשנת 1995. אמדנו את ההסתברות של מורה *i* להיות חדש בשנת 1995, פעם בעזרת המודל הלוגיסטי שהותאם עבור שנת 2001, ופעם בעזרת מודל שהותאם עבור נתוני שנת 1996. בדקנו האם ישנו הבדל בין שונות של טעות האמידה של

המתקבלת כתוצאה משימוש במודלים משנים שונות עבור האמידה בשנת 1995. לשם כך הוגדר המשתנה הבא:

$$\delta_{i,1995,kl}^{(j)} = (I_{new})_{i,1995,kl} - \hat{p}_{i,1995,kl}^{new(j)}, \quad j = 1996, K, 2001 \quad (6.1)$$

כאשר

$$\hat{p}_{i,1995,kl}^{new(j)} = \left(1 + \exp\left(-X_{i,1995,kl}^T \beta_{j,kl}\right)\right)^{-1} \quad (6.2)$$

הינה ההסתברות של מורה i להיות חדש בתא $\{kl\}$ בשנת 1995 הנאמדת באמצעות מודל שהותאם עבור תא זה בשנת j .

הערה: נשים לב כי $\hat{p}_{i,1995,kl}^{new(j)}$ מחושבת בעזרת מכפלת וקטור המקדמים של רגרסיה לוגיסטית שנתקבל כתוצאה מהתאמת מודל לשנת j במטריצת ערכי המאפיינים של המורים בשנת 1995.

חישבנו את האומדן הבלתי מוטה לשונות של $\delta_{i,1995,kl}^{(j)}$ באופן הבא:

$$s_{\delta_{i,1995,kl}^{(j)}}^2 = (n_{1995,kl} - 1)^{-1} \sum_{i=1}^{n_{1995,kl}} \left(\delta_{i,1995,kl}^{(j)} - \bar{\delta}_{i,1995,kl}^{(j)}\right)^2 \quad (6.3)$$

$$\bar{\delta}_{i,1995,kl}^{(j)} = (n_{1995,kl})^{-1} \sum_{i=1}^{n_{1995,kl}} \delta_{i,1995,kl}^{(j)} \quad \text{כאשר}$$

כעת בדקנו מערכת ההשערות הבאה:

$$H_0 : \sigma_{\delta_{i,1995,kl}^{(2001)}}^2 = \sigma_{\delta_{i,1995,kl}^{(1996)}}^2 \quad (6.4)$$

$$H_1 : \sigma_{\delta_{i,1995,kl}^{(2001)}}^2 \neq \sigma_{\delta_{i,1995,kl}^{(1996)}}^2$$

לפי המבחן הסטטיסטי להשוואת שוניויות בשני מדגמים (מבחן χ^2) אי דחיית H_0 משמע לטעון ברמת מובהקות α כי המודלים שהותאמו עבור שנת 1996 ושנת 2001 הנם שקולים לצורך אמידת פרופורציות מורים חדשים בשנת 1995.

בטבלה 5 מובאות דוגמאות נבחרות לתוצאות מבחן השוואה של טעיות אמידה.

טבלה 5: השוואת סטיות תקן של טעויות האמידה של הסתברויות – מגזר עברי

	יסודי דרום	יסודי מרכז	יסודי י-ם	חט"ב דרום	חט"ב מרכז	חט"ב י-ם	חט"ע דרום	חט"ע צפון	חט"ע י-ם
אמידה באמצעות מודל 2001	0.20610	0.18779	0.20539	0.17948	0.15140	0.18052	0.19957	0.17361	0.21923
אמידה באמצעות מודל 1996	0.21509	0.19466	0.22799	0.19473	0.17269	0.18213	0.19489	0.18242	0.22849

כל תוצאות מבחני ההשוואה (χ^2) בכל המחוזות ובכל שלבי החינוך הראו כי המודלים שנתקבלו על סמך נתוני שנת 2001 שקולים למודלים שנתקבלו על סמך נתוני שנת 1996 בכל מחוז ושלב חינוך בהתאמה. מכאן ניתן היה להסיק כי תרומתו של כל מודל לאמידת פרופורציות מורים חדשים בשנים בהן מידע זה בלתי זמין הינה שווה.

6.2 חישוב פרופורציות מורים חדשים לשנים 1994-1991

לאחר ש- (6.4) לא נדחתה ניתן היה לחשב פרופורציות מורים חדשים לשנים 1994-1991 באמצעות המודלים הלוגיסטיים שנתקבלו לשנים 2001-1995 באופן הבא:

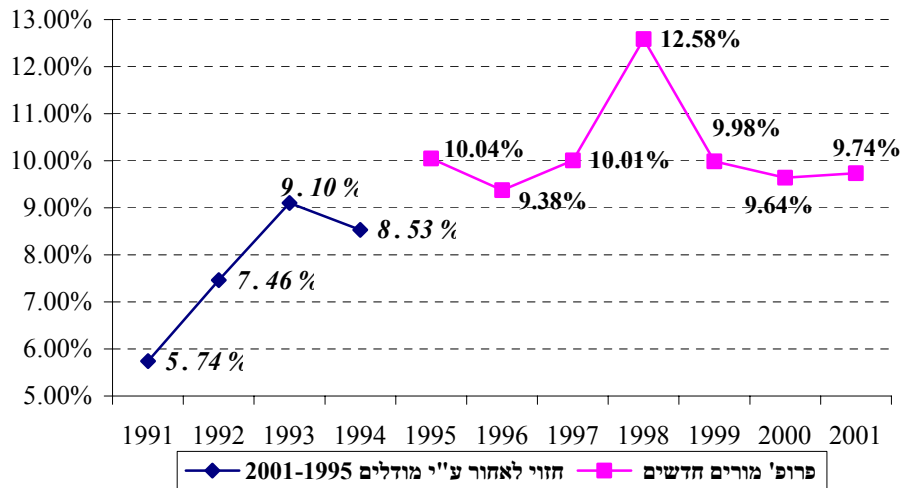
$$\hat{p}_{j'.kl}^{new} = \frac{1}{n_j T} \sum_j \sum_i \left(1 + \exp(-X_{i,j'.kl}^T \beta_{j'.kl}) \right)^{-1} \quad (6.5)$$

$$j = 1995, K, 2001, j' = 1991, K, 1994$$

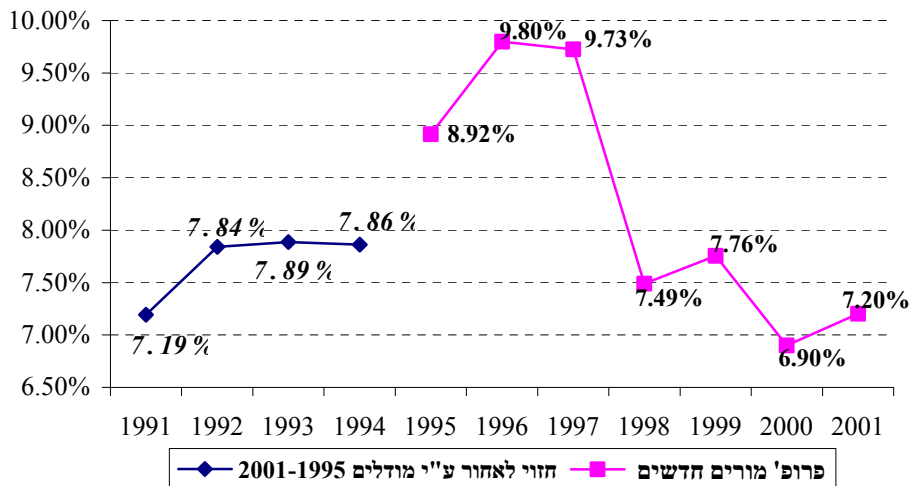
כאשר n_j - הינו סך המורים בשנה j' , $T = 7$, מספר השנים בין 1995 ל-2001.

כלומר, החיזוי של פרופורציות מורים חדשים בכל שנה j' הינו ממוצע אריתמטי של תחזיות שחושבו על סמך המודלים שנבנו לשנים j .

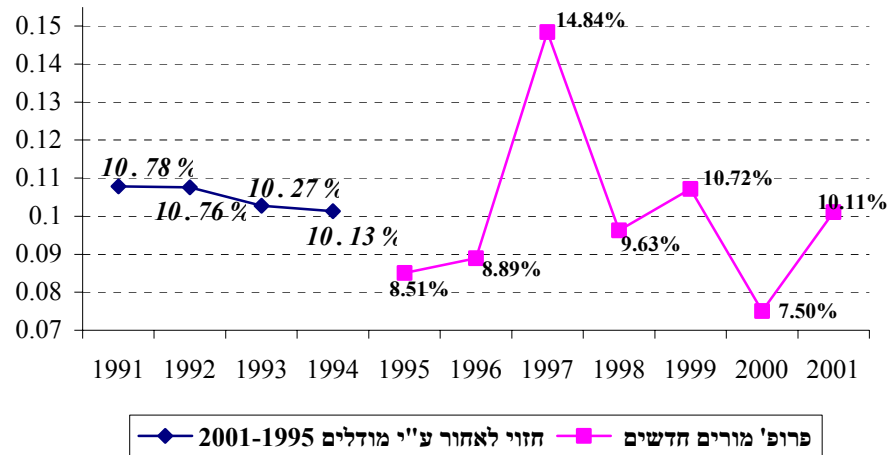
תרשים 4: אמידת פרופ' מורים חדשים ב- 1994-1991 באמצעות חיזוי לאחור – מגזר ערבי, חינוך יסודי (באהוזים)



תרשים 5: אמידת פרופ' מורים חדשים ב- 1994-1991 באמצעות חיזוי לאחור – מגזר ערבי, חטיבת ביניים (באהוזים)



תרשים 6: אמידת פרופ' מורים חדשים 1994-1991 באמצעות חיזוי לאחור – מגזר ערבי, חטיבה עליונה (באחוזים)



7 התאמת מודל אגרגטיבי לחיזוי פרופורציות מורים חדשים לשנים 2009-2003

7.1 התאמת מודל Logit

כעת, לאחר השלמת מידע אודות פרופורציות מורים חדשים בשנים הראשונות למחקר (1994-1991) נתקבל וקטור \hat{p}_{jkl}^{new} , $j = 1991, K, 2001$, הכולל תצפיות אגרגטיביות של פרופורציות מורים חדשים בשנים 2001-1991. וקטור זה מכיל מידע אמפירי על הצטרפות מורים חדשים בשנים 2001-1995 ובנוסף אומדני פרופורציות מורים חדשים בשנים 1994-1991 שנתקבלו מתוך 6.1 בהתבסס על ההנחה כי פרופורציות המורים החדשים היא תופעה על פני זמן, בחרנו את השנה הקלנדרית בתור משתנה פוטנציאלי המסביר את פרופורציית מורים חדשים. מספר התלמידים במערכות חינוך, גם הוא ידוע כמשתנה פוטנציאלי בעל השפעה על הצטרפות מורים חדשים למערכות החינוך, אך מחמת מחסור בנתוני מספר התלמידים בשנים 1994-1991 נבצר מאתנו לנצל להשתמש במספר התלמידים כמשתנה מסביר נוסף במודל האגרגטיבי לפרופורציות מורים חדשים. חשוב לציין כי התצפיות האמיתיות ואומדני פרופורציות מורים חדשים במערכות חינוך הינם תלויים ביניהן, היות וכוללים מידע על אותם מורים העוברים משנה לשנה בתקופה הנחקרת. על מנת להתאים מודל לאמידה ולחיזוי פרופורציות מורים חדשים היה עלינו לבטל תלות בין התצפיות ולבצע התמרה מתאימה כדי להביא את המשתנה המוסבר, שהוא פרופורציות מורים חדשים עד כדי התפלגות הקרובה לכל היותר להתפלגות נורמלית. לשם כך השתמשנו בהתמרה לוגיסטית ונבנה לכל תא $\{kl\}$ מודל אגרגטיבי לאמידת פרופורציות מורים חדשים:

$$j = 1991, K, 2001, \log it(\hat{p}_{jkl}^{new}) = \lambda_{0kl} + \lambda_{1kl}j + \varepsilon_{kl} \quad (7.1)$$

בהתבסס על הנחות רגרסיה, על מנת שיהיה תוקף ל- 7.1 צריך להתקיים:

$$\log it(\hat{p}_{jkl}^{new}) \sim N(\mu_{kl}, \sigma_{kl})$$

7.2 בדיקת נורמליות

נבדוק את ההשערות

$$H_0 : \log it(\hat{p}_{jkl}^{new}) \sim N(\mu_{kl}, \sigma_{kl}) \quad (7.2)$$

אחרת: H_1

מסיבה שקיימים מספר מבחנים אי פרמטריים לבדיקת נורמליות, נקבע כי לא נדחה את השערת הנורמליות ברמת מובהקות של 0.01 אם על פי לפחות אחד המבחנים האי פרמטריים: מבחן שפירו-וילק, מבחן קולמוגורוב-סמירנוב, מבחן קרמר-פון מיסס ומבחן אנדרסון-דרלניג ניתן לא לדחות השערת H_0 .

טבלה 6: תוצאות מבחני נורמליות ל $\log it(\hat{p}_{jkl}^{new})$ – מגזר עברי

שלב חינוך	מחוז	Shapiro-Wilk	Kolmogorov-Smirnov	Cramer -von Mises	Anderson Darling	H_0 (Normal)
יסודי	1	0.5663	>0.1500	>0.2500	>0.2500	Not rejected
יסודי	2	0.5171	>0.1500	>0.2500	>0.2500	Not rejected
יסודי	3	0.5725	>0.1500	>0.2500	>0.2500	Not rejected
יסודי	4	0.2790	0.1197	0.2265	0.2491	Not rejected
יסודי	5	0.7658	>0.1500	>0.2500	>0.2500	Not rejected
יסודי	6	0.3758	>0.1500	0.0976	0.1648	Not rejected
הט"ב	1	0.0450	>0.1500	0.0482	0.0480	Not rejected
הט"ב	2	0.1494	0.0713	0.1464	0.1498	Not rejected
הט"ב	3	0.3277	>0.1500	>0.2500	>0.2500	Not rejected
הט"ב	4	0.1081	>0.1500	>0.2500	0.1826	Not rejected
הט"ב	5	0.2569	>0.1500	>0.2500	>0.2500	Not rejected
הט"ב	6	0.4377	>0.1500	>0.2500	>0.2500	Not rejected
הט"ע	1	0.1730	>0.1500	0.1701	0.1781	Not rejected
הט"ע	2	0.7700	>0.1500	>0.2500	>0.2500	Not rejected
הט"ע	3	0.2655	>0.1500	>0.2500	>0.2500	Not rejected
הט"ע	4	0.0297	0.0426	0.0179	0.0217	Not rejected*
הט"ע	5	0.3597	0.1500	0.2500	0.2500	Not rejected
הט"ע	6	0.2536	0.1337	0.2500	0.2431	Not rejected

*ניתן לדחות רק ברמת מובהקות גדולה מ-0.0426.

בטבלה 6 מובאים מבחנים סטטיסטיים אי פרמטריים לבדיקת נורמליות של $\log it(\hat{p}_{jkl}^{new})$ במגזר העברי. תוצאותיהן מצביעים על קבילות הנחת הנורמליות, דבר המצדיק התאמת מודל אגרגטיבי לאמידה וחיזוי פרופורציות מורים חדשים במערכות חינוך.

7.3 חיזוי מספר מורים לשנים 2002-2007

על פי המודל 7.1 תחזיות לפרופורציית מורים חדשים לשנים 2002, K, 2007 נתקבלו מתוך משוואה הבאה:

$$\hat{p}_{j''kl}^{new} = \frac{e^{\lambda_{0kl} + \lambda_{1kl} j''}}{1 + e^{\lambda_{0kl} + \lambda_{1kl} j''}} \quad (7.3)$$

על מנת לספק נתון אודות היקף המורים החדשים במערכות חינוך, השתמשנו בתחזיות מורים במערכות חינוך שנתקבלו בעבודה להתאמת מודלים סטוכסטיים לחיזוי היקף כח אדם בהוראה במערכות חינוך (ראה דוח וקסלר ושפירא 2003, פרק X-1).

אם $\hat{N}_{j''.kl}$ - הוא מספר מורים בשנה $j'' = 2002, K, 2007, j''$, אזי מספר המורים החדשים בשנה j'' הינו

$$\hat{N}_{j''.kl}^{new} = \hat{p}_{(j''-1).kl}^{new} \hat{N}_{(j''-1).kl} \quad (7.4)$$

נתוני תחזיות מורים חדשים מובאים בנספח 8 לדוח זה.

8 נספח מודלים ותרשימים