

מדינת ישראל
משרד ראש הממשלה



נספח תחזיות כח אדם בהוראה במגזר הערבי
מצורף לדוח מתודולוגי של מודלים סטוכסטיים לחיזוי היקף כח אדם
בהוראה במערכת חינוך בישראל

אלון שפירא
תחום ניתוח סטטיסטי
אגף מדען ראשי

ד"ר אלברט וקסלר
תחום ניתוח סטטיסטי
אגף מדען ראשי

1 מבוא ותמצית העבודה

תהליך חיזוי מספר מורים במגזר הערבי בדומה למגזר העברי התבסס על שימוש במודלים לוגיסטיים שהותאמו על סמך נתוני פרט של המורים במערכות החינוך בישראל המתייחסים לתקופה 1991-2001. להבדיל מהמגזר העברי שבו 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. פיתוח המודלים ומתן תחזיות במגזר הערבי נעשה בשלב שאחרי פיתוח דומה במגזר העברי, ולכן הוכנסו מספר שינויים ועדכונים לתהליך הפיתוח וקביעת המודלים המועדפים במגזר הערבי. השינויים נבעו בחלקם מהבדלי נתונים במגזר הערבי וחלקם היו תוצאה של שיפור תהליך הפיתוח על סמך ניסיון העבר במגזר העברי. השינוי הראשון קשור לתהליך זקיפת ערכים חסרים במאפייני המורים. בהשוואה לתהליך זקיפות על בסיס ממוצע, הורצו מודלים של רגרסיה ליניארית לזקיפת ערכים חסרים של מאפייני המורים, וזאת לכל שנה בנפרד. תהליך זה הקטין באופן משמעותי את טעות התקן באמידת פרופורציות כניסה/פרישה ממערכת החינוך באמצעות מודלים לוגיסטיים. תהליך הזקיפה המשופר חסך גם את שלב השוואת התחזיות שהופקו על בסיס קבצי נתונים לפני ואחרי זקיפות, כל המודלים הורצו הפעם על קבצי נתונים לאחר תהליך זקיפת נתונים חסרים. השינוי השני מתייחס לאופן החיזוי של פרופורציות מורים שהצטרפו/פרשו ממערכת החינוך. עבור שלושת שלבי החינוך של המגזר הערבי הותאמו שני סוגים של מודלים:

- א. המודלים הלוגיסטיים הרב השנתיים לאמידת פרופורציות מעבר של המורים. מודלים אלו כוללים את משתנה השנה הקלנדרית בתור משתנה פוטנציאלי מסביר. (ראה (X.2.1)).
- ב. המודלים הלוגיסטיים הדו-שנתיים שבאמצעותם נאמדו שיעורי גידול במספר מורים בין שנה לשנה ועל בסיס האומדנים הללו הותאמו מודלים אגרסיביים (*Logit*) של שיעורי הגידול במספר המורים, בעזרתם ניתנו תחזיות מורים לשנים 2002-2007 הקריטריון לבחירת המודל המתאים מבין שתי האפשרויות בדומה לזה שהוסבר ברוח מתודולוגי על המגזר העברי התבסס על טעות החיזוי המינימלית בשנים 2002 ו-2003, השנים שבהן נתונים אמיתיים היו זמינים לצורך השוואה עם תחזיות.

2 טיפול בזקיפת נתונים חסרים

בקבצי מורים בהם נמצאו לפחות 5% 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 השנייה והשלישית של משתנה גיל לצורך זיהוי קשרים לא ליניאריים. המשתנים פוטנציאליים המסבירים מאפיין בעל ערכים חסרים הוכנסו לתהליך התאמת מודל רגרסיה ליניארית שהתבצעה בשלבים (stepwise). בטבלה 1 מובאת רשימת מאפיינים בהם נמצאו ערכים חסרים ורשימת משתנים פוטנציאליים שהשתתפו בתהליך התאמת המודלים.

טבלה 1: 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	
בסיס משרה			X	X	X	X			
מס' ילדים		X	X	X					
דרגה		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.

3 המודל האגרטיבי מול המודל הרב שנתי

כתוצאה מהרצת מודלים לוגיסטיים דו-שנתיים נתקבלו שני וקטורים של הסתברויות מעבר ממוצעות,

$$\pi_{(j-1+m)klm} \quad j = 1992, K, 2000 \text{ מתוך}$$

$$\pi_{(j-1+m)klm} = n_j^{-1} \sum_{i=1}^{n_j} \left(1 + \exp(-X_{i(j-1+m)kl}^T \beta_{(j-1+m)klm}) \right)^{-1},$$

כאשר $m = 0, 1$, המציינים כניסה או יציאת מורים מהמערכת בהתאמה.

שיעור גידול שנתי של מספר המורים במעבר בין שנה $j-1$ לשנת j נתקבלו מתוך

$$u_{jkl} = \frac{1 - \pi_{j-1.kl1}}{1 - \pi_{jkl0}}$$

על סמך הנחת התלות בזמן, בחרנו את השנה הקלנדרית בתור משתנה פוטנציאלי המסביר את שיעור גידול במספר מורים. מספר התלמידים במערכות חינוך לא נלקח בחשבון כמשתנה פוטנציאלי מחמת מחסור בנתוני מספר התלמידים בשנים 1991-1994. על מנת להתאים מודל לאמידה ולחיזוי שיעור גידול במספר מורים היה צורך לבטל תלות בין התצפיות ולבצע התמרה מתאימה כדי להביא את המשתנה המוסבר עד כדי התפלגות הקרובה לכל היותר להתפלגות נורמלית. לשם כך השתמשנו בהתמרה לוגיסטית ובנינו לכל תא $\{kl\}$ מודל אגרטיבי לאמידת שיעור גידול במספר מורים:

$$j = 1991, K, 2001, \log it(v_{j,kl}) = \lambda_{0kl} + \lambda_{1kl}j + \lambda_{2kl}\tau_{jkl} + \varepsilon_{kl}$$

כאשר τ_{jkl} - הינו שיעור גידול במספר תלמידים משנה $j-1$ לשנה j המחושב

$$\tau_{jkl} = \frac{S_{jkl} - S_{j-1,kl}}{S_{j-1,kl}}$$

המודל האלטרנטיבי לאמידה וחיזוי שיעור גידול במספר מורים התקבל מתוך שימוש במודל הלוגיסטי הרב שנתי על בסיס נתוני מורים בשנים 1992, K, 2001, כאשר השנה הקלנדרית נכללת בקרב משתנים מסבירים של המודל. לפי כך שיעור גידול במספר מורים חזוי חושב כדלקמן:

$$\tilde{j} = 2001, K, 2006, \hat{v}_{\tilde{j}kl} = \frac{1 - \pi_{(\tilde{j}-1)kl}}{1 - \pi_{\tilde{j}kl0}}$$

כאשר $j = 1992, K, 2001, j \in X_{ikl}^T, \pi_{klm} = n_j^{-1} \sum_{i=1}^{n_j} (1 + \exp(-X_{ikl}^T \beta_{klm}))^{-1}$
 ו- $\hat{\beta}_{klm}, m = 0, 1$ וקטורי מקדמים של הרגרסיה הלוגיסטית.

תחזיות מספר מורים במערכת בשנה $\tilde{j} = 2001, K, 2006$ חושבו באופן הבא:

$$\hat{N}_{\tilde{j}kl} = \hat{v}_{\tilde{j}kl} N_{(\tilde{j}-1)kl}$$

התחזיות ניתנו על פי שני המודלים שתוארו לעיל: המודל האגרסיבי והמודל הרב שנתי, הקריטריון לבחירת המודל המתאים מביניהם התבסס על טעות החיזוי שחושבה על סמך נתונים אמיתיים למספר מורים בשנים 2002 ו-2003. כך לדוגמה, המודל האגרסיבי (מודל 1) נמצא עדיף על מודל הרב שנתי (מודל 2) אם לכל 2002, 2003 \tilde{j}

$$e_{\tilde{j}kl}(\hat{N}_{\tilde{j}kl1}) > e_{\tilde{j}kl}(\hat{N}_{\tilde{j}kl2})$$

$$e_{\tilde{j}kl} = \frac{|N_{\tilde{j}kl} - \hat{N}_{\tilde{j}kl}|}{N_{\tilde{j}kl}} \text{ כאשר}$$

תוצאות החיזוי על פי המודל שנבחר על פי הקריטריון להשוואת המודלים מופיעות בתרשימים המצורפים לנספח זה.