מדינת ישראל

משרד ראש הממשלה

הלשכה המרכזית
לסטטיסטיקה

מודלים סטוכסטיים לחיזוי היקף כח אדם
בהוראה במערכת החינוך בישראל
דו"ח מתודולוגי

ד"ר אלברט וקסלר

תחום ניתוח סטטיסטי

אגף מדען ראשי

אלון שפירא

תחום ניתוח סטטיסטי

אגף מדען ראשי

ירושלים, אפריל 2003

תוכן עניינים

I. מבוא

בדו"ח מתודולוגי זה מתוארים שלבי תכנון ופיתוח מודלים סטוכסטיים שבאמצעותם ניתן לחזות את היקף כח אדם בהוראה במערכת החינוך בישראל במגזר העברי הלא חרדי. בשיטה שפותחה נעשה שימוש במודלים לוגיסטיים המבוססים על אמידת הסתברויות מעבר של המורים משנה לשנה, דהיינו, הסתברויות נשירת המורים מהמערכת והסתברויות כניסתם למערכת Vexler and Maagan (2002). בעזרת הסתברויות המעבר שנאמדו באמצעות משתנים מסבירים של המודלים הלוגיסטיים נחזתה מצבת כח אדם בהוראה בשנה מסוימת. נציין כי המודלים הלוגיסטיים לחיזוי היקף כח אדם בהוראה ניבנו בנפרד לכל שלב חינוך (יסודי, חטיבות ביניים וחטיבות עליונות) ולפי שישה מחוזות גיאוגרפיים כפי שהוגדרו על ידי משרד החינוך (ירושלים, צפון, מרכז, תל אביב, חיפה ודרום). המשתנים המסבירים שנכללו במודלים נבחרו מקבוצת המאפיינים האישיים והמקצועיים של המורים (גיל, ותק, משרה, תפקיד וכד') וכללו גם את מספר התלמידים המתאים לכל שלב חינוך ומחוז גיאוגרפי. כמו כן, הותאמו מודלים לחיזוי מספר התלמידים בכל שלב חינוך ובכל מחוז. תחזיות התלמידים שנתקבלו מתוך המודלים שולבו בתהליך חיזוי מספר המורים בכל שלב ומחוז מתאים. המודלים לחיזוי מורים התבססו על מאגרי מידע אודות המורים בתקופה 1991-2001 תוך הנחה כי התפלגות מאפייני המורים בכל תקופה של שנתיים עוקבות הינה שווה. בנוסף, במסגרת המחקר, במטרה לבחון את יציבות הנתונים נעשו סימולציות מונטה-קרלו וכמו כן נעשתה בדיקת אמינות הנתונים. הדו"ח כולל את הפרקים הבאים: מבנה המחקר ועיצובו, סטטיסטיקה תיאורית של המורים במערכת החינוך, בדיקת שלמות הנתונים והתאמת זקיפות, בדיקת יציבות, פיתוח ויישום שיטה לבדיקת אמינות הנתונים, מתודולוגית החיזוי הכוללת את התאמת מודלים לחיזוי תלמידים והתאמת מודלים לוגיסטיים לחיזוי מורים ונספח הכולל את המודלים לחיזוי ותרשימי תחזיות כח אדם בהוראה ב-18 תאים נפרדים (בשלושה שלבי חינוך ולפי שישה מחוזות גיאוגרפיים). הדו"ח אינו כולל דיון במשמעויות החברתיות של הממצאים שנתקבלו, ומטרותיו העיקריות הן להסביר את שלבי הניתוח של הנתונים, לתאר את תהליך בניית המודלים לחיזוי היקף כח אדם בהוראה ולספק תחזיות סבירות.

II. מבנה המחקר

המחקר מתחלק למספר שלבים עיקריים:

1. בבחינת שלמות קבצי הנתונים כולל ניתוח תיאורי של נתוני המורים
2. בדיקת יציבות הנתונים הכולל הרצת סימולציות מונטה קרלו
3. בדיקת אמינות הנתונים, הכוללת פיתוח שיטה לבדיקת הומגניות המדגמים ובדיקה יישומית
4. פיתוח מתודולוגיה לחיזוי היקף כח אדם בהוראה
5. התאמת מודלים לחיזוי תלמידים
6. התאמת מודלים לוגיסטיים לחיזוי מספר מורים ובדיקת יכולת החיזוי.

III. קבצי נתונים ומאפייניהם

1. כללי

קבצי הנתונים שהועמדו לרשותנו על ידי גף חינוך כוללים רשומות מורים ומאפייניהם בשנים 1991-2001. נתוני המורים מסווגים לשלבי החינוך: חינוך יסודי, חטיבת ביניים וחטיבות עליונות, ולמחוזות גיאוגרפיים כפי שהוגדרו על ידי משרד החינוך. כל רשומת מורה כוללת אינפורמציה על שייכותו למערכת ההוראה במשך התקופה הנחקרת, ועל מאפייניו האישיים (מין, גיל, ארץ מוצא וכד') והמקצועיים (ותק, בסיס משרה, שעות הוראה, תפקיד וכד').

2. טיפול בנתונים חסרים

במסגרת בדיקות מקדימות של קבצי הנתונים נתגלה מספר ערכים חסרים בחלק מהמאפיינים העולה על 10%. תהליך השלמת הנתונים התבסס על שני עקרונות: זקיפה מהתקופה הסמוכה וזקיפה על פי שיטת הממוצע. הערכים החסרים של המאפיינים הכרונולוגיים (כגון גיל, וותק וכד') הושלמו במידת האפשר מהערכים בשנה הקרובה ביותר בה ערך המאפיין היה זמין. פרוצדורת הזקיפה הוגדרה באופן הבא:

יהי $X_{ijkl} = \left[x_{1ijkl}, x_{2ijkl}, K\ x_{pijkl}\right]^T$ וקטור מאפיינים של מורה i בשנה j בשלב חינוך k במחוז l

יהי x_{hijkl} - ערך חסר של מאפיין כרונולוגי h של מורה i בשנה j וכו', $1 \le h \le p$, אזי

(III - 2.1) $$1 \le c \le 10\ ,\ x_{hijkl} = x_{hi(j \pm c)kl}\ \mu\ c$$

כאשר $j \pm c$ הינה השנה הסמוכה ביותר לשנת j ובה $x_{hi(j \pm c)kl}$ שונה מערך חסר.

במידה ולא קיים c, כך ש $x_{hi(j \pm c)kl}$ הינו ערך קיים (לא חסר) אזי x_{hijkl} יועד לתהליך זקיפה על פי שיטת הממוצע (ראה **(III - 2.3)** בהמשך).

תהליך של השלמת ערכי מאפיינים על פי הם התבצע סיווג המורים (כגון מחוז, שלב חינוך, סמל מוסד וכד') התבסס על עקרון היציבות, וזאת בהנחה כי המורה נוטה להישאר במצב בו היה נתון בשנה הקודמת. הערכים החסרים במקרים אלו הושלמו מהשנה הראשונה $j \pm c$ בה הערך המתאים למאפיין h היה שונה מערך חסר. כלומר:

(III - 2.2) $$x_{hijkl} = x_{hi(j \pm c)kl}.$$

ערכי מאפיינים כרונולוגיים, שלא ניתן היה להשלימם על פי זקיפה מהתקופה הקרובה, ובלבד שלא השתייכו למאפייני סיווג המורים או למאפיינים בלתי כרונולוגיים כגון: בסיס משרה, תפקיד, שעות הוראה וכד', נזקפו על פי שיטת הממוצע באופן הבא:

$$G = \{1 \le q \le N_{jkl}, x_{hqjkl} \notin \{\emptyset\}\}, x_{hijkl} = \frac{1}{N^*_{jkl}} \sum_{g \in G} x_{hgjkl} \qquad \textbf{(III - 2.3)}$$

כאשר N_{jkl} מציין את מספר המורים ב $\{jkl\}$, N^*_{jkl} - מציין את מספר המורים ב $\{jkl\}$ בהם x_{hijkl} הוא אינו ערך חסר.

IV. נתוני מורים – סטטיסטיקה תיאורית

במסגרת ניתוח נתונים מקדים, רוכזו נתונים על היקף כח אדם בהוראה בתקופה 1991-2001 וחושבו שיעורי גידול שנתיים של המורים בכל שלב חינוך ובכל מחוז גיאוגרפי (ראה טבלאות 1-3).

טבלה 1: כח אדם בהוראה בחינוך יסודי לפי מחוזות

	דרום		תל אביב		מרכז		חיפה		צפון		ירושלים	
	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי
1991	**5280**	**..**	**8056**	..	**8350**	..	**4128**	..	**4615**	..	**3547**	..
1992	**5613**	**1.063**	**8003**	**0.993**	**8715**	**1.044**	**4270**	**1.034**	**4983**	**1.080**	**3918**	**1.105**
1993	**5947**	**1.060**	**8341**	**1.042**	**8984**	**1.031**	**4535**	**1.062**	**5270**	**1.058**	**4160**	**1.062**
1994	**6336**	**1.065**	**8577**	**1.028**	**9515**	**1.059**	**4699**	**1.036**	**5499**	**1.043**	**4399**	**1.057**
1995	**6942**	**1.096**	**9061**	**1.056**	**10063**	**1.058**	**4939**	**1.051**	**5850**	**1.064**	**4832**	**1.098**
1996	**7311**	**1.053**	**9405**	**1.038**	**10456**	**1.039**	**5203**	**1.053**	**6069**	**1.037**	**5286**	**1.094**
1997	**7277**	**0.995**	**9111**	**0.969**	**10630**	**1.017**	**5334**	**1.025**	**5960**	**0.982**	**5539**	**1.048**
1998	**7537**	**1.036**	**9012**	**0.989**	**10789**	**1.015**	**5502**	**1.031**	**6045**	**1.014**	**5760**	**1.040**
1999	**7192**	**0.954**	**8797**	**0.976**	**10826**	**1.003**	**5329**	**0.969**	**6031**	**0.998**	**5804**	**1.008**
2000	**7233**	**1.006**	**8556**	**0.973**	**10931**	**1.010**	**5298**	**0.994**	**6010**	**0.997**	**5787**	**0.997**
2001	**7115**	**0.984**	**8292**	**0.969**	**11093**	**1.015**	**5204**	**0.982**	**5907**	**0.983**	**5581**	**0.964**

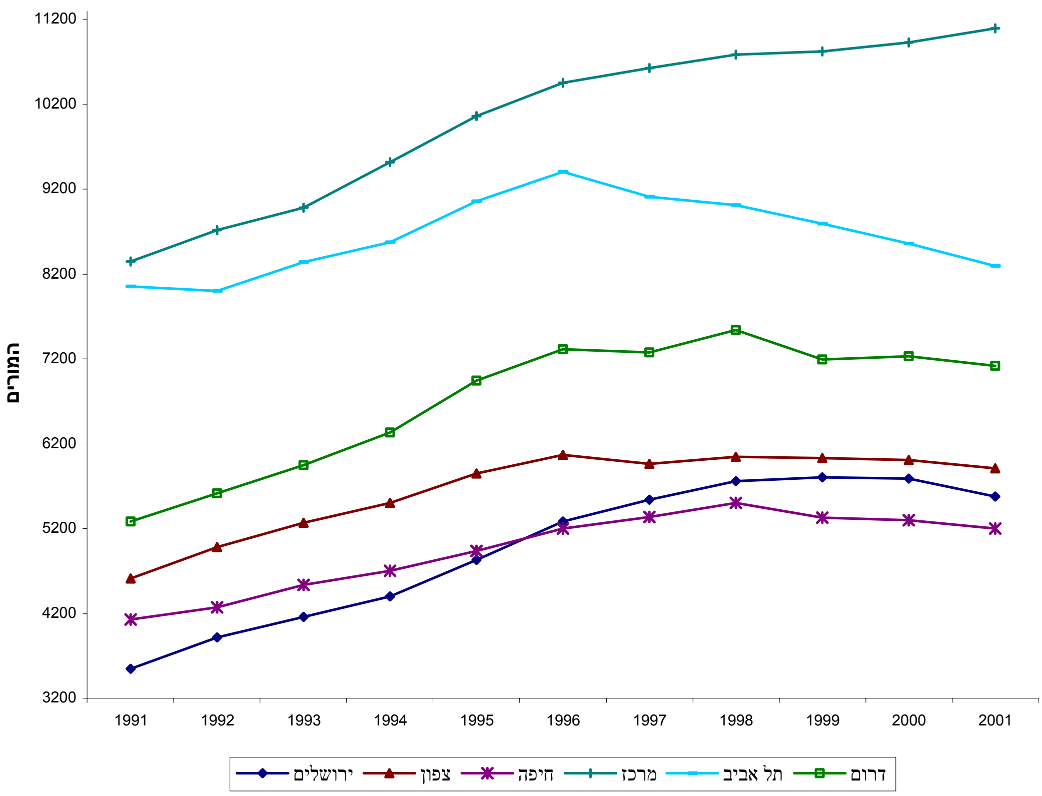
טבלה 2: : כח אדם בהוראה בשלב חטיבות ביניים לפי מחוזות

	דרום		תל אביב		מרכז		חיפה		צפון		ירושלים	
	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי
1991	**1746**	**..**	**1568**	**..**	**3295**	**..**	**1673**	**..**	**1946**	**..**	**1230**	**..**
1992	**1950**	**1.117**	**1599**	**1.020**	**3435**	**1.042**	**1712**	**1.023**	**2083**	**1.070**	**1266**	**1.029**
1993	**2082**	**1.068**	**1687**	**1.055**	**3544**	**1.032**	**1910**	**1.116**	**2196**	**1.054**	**1302**	**1.028**
1994	**2326**	**1.117**	**1686**	**0.999**	**3637**	**1.026**	**2073**	**1.085**	**2316**	**1.055**	**1296**	**0.995**
1995	**2513**	**1.080**	**1864**	**1.106**	**3911**	**1.075**	**2280**	**1.100**	**2349**	**1.014**	**1365**	**1.053**
1996	**2853**	**1.135**	**2144**	**1.150**	**4568**	**1.168**	**2518**	**1.104**	**2483**	**1.057**	**1409**	**1.032**
1997	**3288**	**1.152**	**2419**	**1.128**	**4907**	**1.074**	**2664**	**1.058**	**2587**	**1.042**	**1529**	**1.085**
1998	**3545**	**1.078**	**2400**	**0.992**	**5149**	**1.049**	**2682**	**1.007**	**2612**	**1.010**	**1638**	**1.071**
1999	**3850**	**1.086**	**2458**	**1.024**	**5340**	**1.037**	**2714**	**1.012**	**2723**	**1.042**	**1790**	**1.093**
2000	**3808**	**0.989**	**2412**	**0.981**	**5457**	**1.022**	**2626**	**0.968**	**2674**	**0.982**	**1886**	**1.054**
2001	**3861**	**1.014**	**2326**	**0.964**	**5198**	**0.953**	**2462**	**0.938**	**2560**	**0.957**	**2130**	**1.129**

טבלה 3 : : כח אדם בהוראה בשלב חטיבות עליונות לפי מחוזות

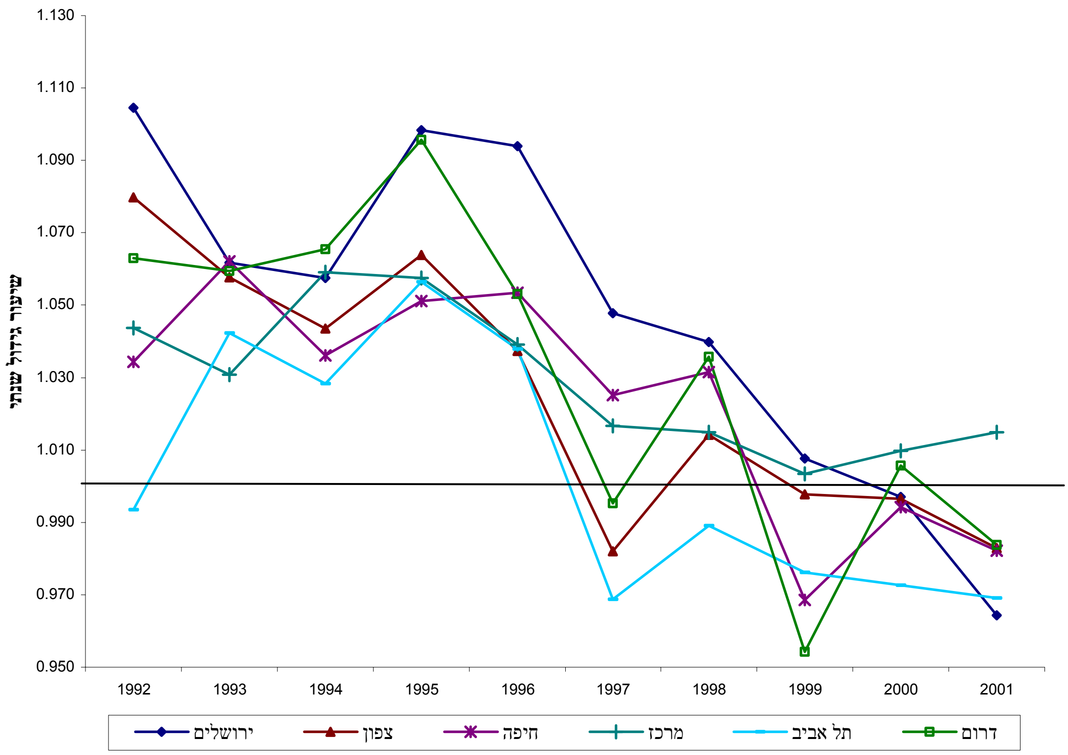
	ירושלים		צפון		חיפה		מרכז		תל אביב		דרום	
	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי	סך מורים	שיעור גידול שנתי
1991	**2788**	**..**	**2560**	**..**	**3064**	**..**	**5392**	**..**	**4906**	**..**	**3235**	**..**
1992	**2928**	**1.051**	**2778**	**1.051**	**3212**	**1.048**	**5701**	**1.058**	**5450**	**1.111**	**3437**	**1.062**
1993	**3183**	**1.087**	**2905**	**1.046**	**3300**	**1.027**	**6025**	**1.056**	**5782**	**1.061**	**3534**	**1.028**
1994	**3478**	**1.093**	**3074**	**1.058**	**3462**	**1.049**	**6426**	**1.066**	**5840**	**1.010**	**3672**	**1.039**
1995	**3644**	**1.047**	**3177**	**1.033**	**3485**	**1.006**	**6644**	**1.034**	**5998**	**1.027**	**3859**	**1.051**
1996	**3721**	**1.021**	**3153**	**0.992**	**3446**	**0.989**	**6644**	**1.000**	**6140**	**1.024**	**3861**	**1.001**
1997	**3528**	**0.948**	**3299**	**1.046**	**3378**	**0.980**	**6927**	**1.043**	**5825**	**0.949**	**4018**	**1.041**
1998	**3929**	**1.113**	**3228**	**0.978**	**3486**	**1.032**	**6907**	**0.997**	**5744**	**0.986**	**4214**	**1.049**
1999	**4599**	**1.171**	**3722**	**1.153**	**3834**	**1.100**	**7842**	**1.135**	**6239**	**1.086**	**4757**	**1.129**
2000	**4801**	**1.044**	**3761**	**1.010**	**3758**	**0.980**	**8045**	**1.026**	**6015**	**0.964**	**4860**	**1.022**
2001	**5485**	**1.142**	**4170**	**1.109**	**4012**	**1.068**	**8440**	**1.049**	**6199**	**1.031**	**5443**	**1.120**

תרשים 1 : נתוני מורים – חינוך יסודי 1991-2001

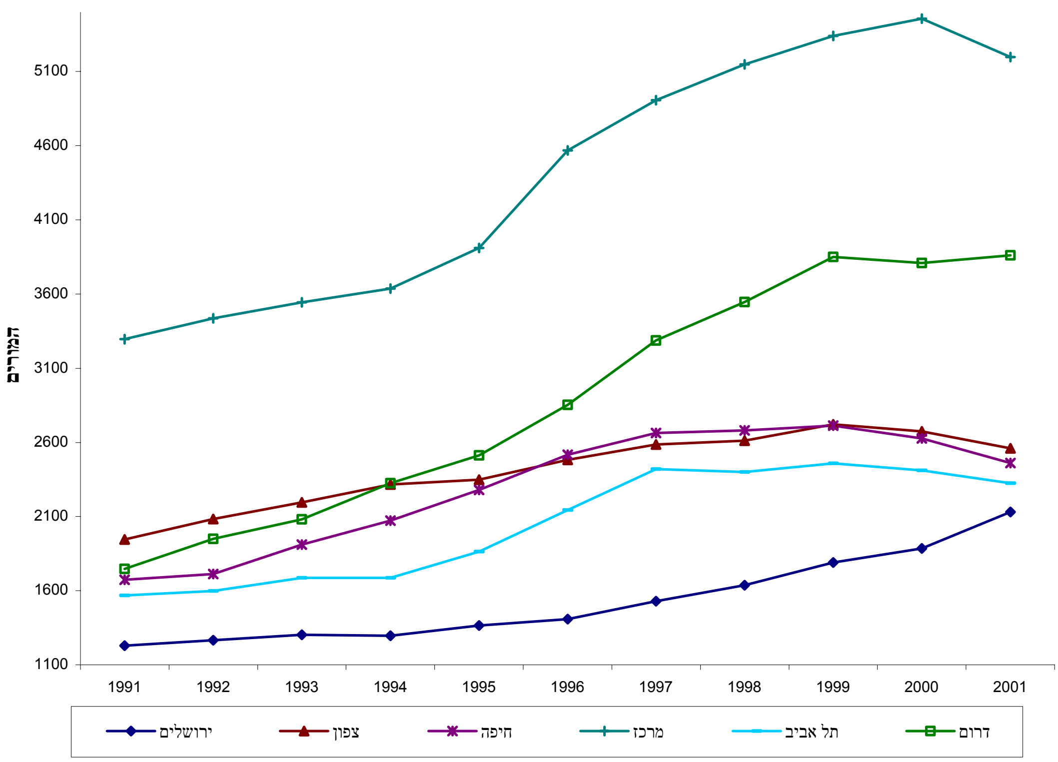


ניתן לראות כי עד שנת 1996 בחינוך היסודי בכל המחוזות ישנה עלייה במספר המורים (ראה תרשים 1). החל משנת 1997 התמונה שונה בין מחוז למחוז. במחוז מרכז, למשל, העלייה נמשכת גם לאחר 1997, אך היא מתונה יותר, במחוז תל אביב חלה ירידה החל מ-1997, בחיפה העלייה נמשכת עד 1998 ולאחר מכן מסתמנת ירידה. בשאר המחוזות לאחר סיום העלייה בשנת 1997 קשה לתאר תמונה מדויקת ולתת הערכה מהי מגמת ההמשך של מספר המורים.

תרשים 2: שיעורי גידול מורים על פי מחוזות – חינוך יסודי 1992-2001

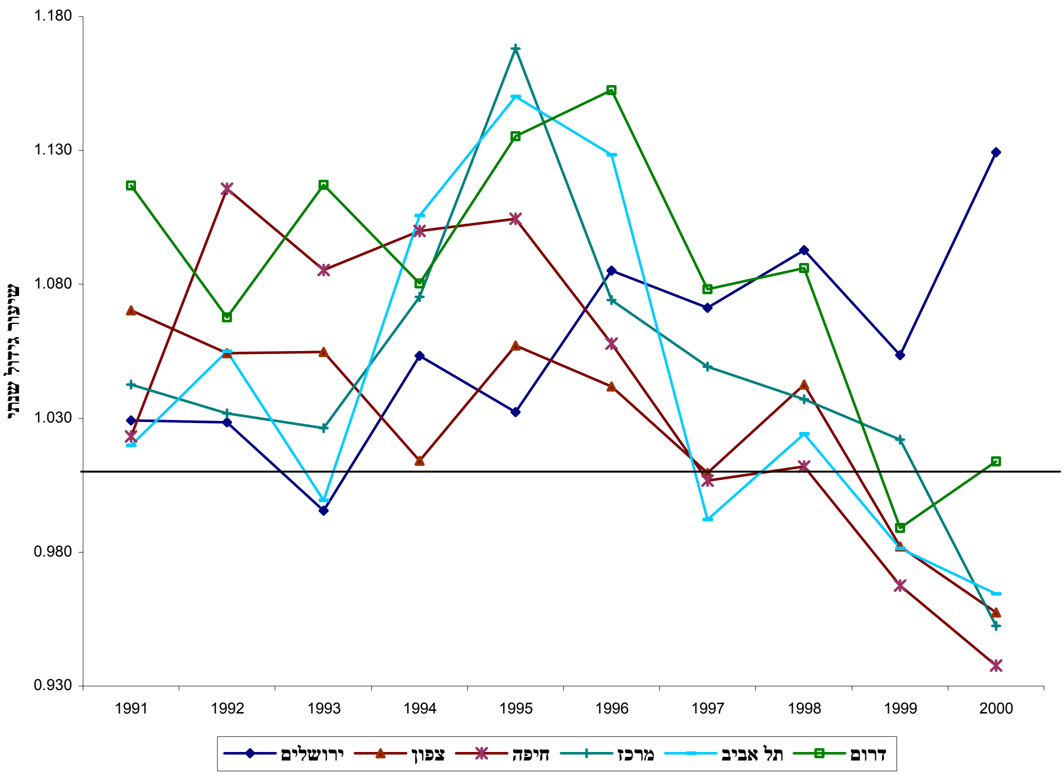


תרשים 3: נתוני מורים – חטיבת ביניים 1991-2001



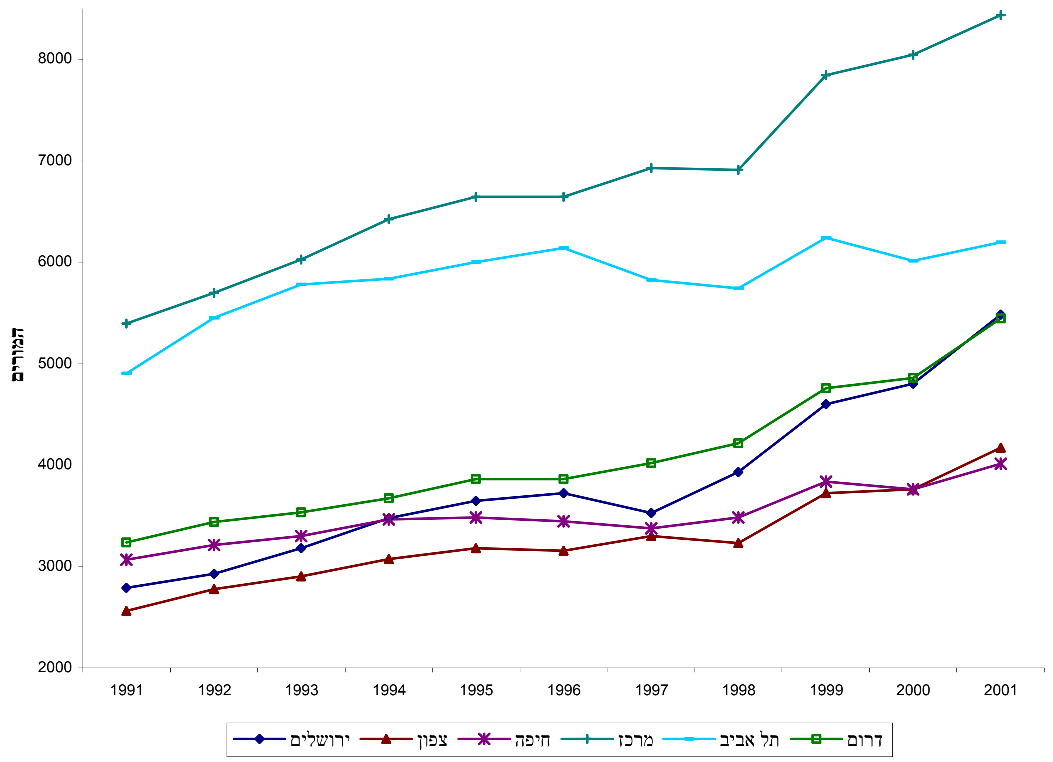
למרות תנודתיות רבה, בהסתכלות כוללת נראים שיעורי גידול שנתיים של המורים בחינוך היסודי (תרשים 2) במגמת ירידה, ונראה כי בשנת 1999 שיעורי גידול מורים במחוזות מרכז וירושלים מגיעים ל-1, ושאר המחוזות חצו את גבול ה-1 והגיעו לערכים שליליים (קטנים מ-1). בתרשים 3 ניתן לראות כי גם בחטיבות הביניים יש גידול בכל המחוזות במספר המורים עד 1997, ולאחר מכן התמונה בכל מחוז היא שונה, כאשר בחלק מהמחוזות (חיפה ותל אביב) ישנה ירידה במספר מורים.

תרשים 4: שיעורי גידול מורים – על פי מחוזות – חטיבת ביניים



כאשר מתבוננים בשיעור גידול המורים בחטיבות ביניים (תרשים 4) ניתן לזהות תנודתיות שיעור גידול המורים. עד שנת 1999 נצפים ברוב המחוזות שיעורי גידול חיוביים. החל משנת 2000 שיעורי הגידול במחוזות תל אביב, צפון וחיפה מגיעים אל מתחת לקו ה-1, כלומר לשיעורי גידול שליליים.

תרשים 5: נתוני מורים – חטיבה עליונה 1991-2001



תרשים 6: שיעורי גידול מורים על פי מחוזות שונים- חטיבה עליונה

גם בתרשים 5 המתאר את מספר המורים בשלב חטיבות עליונות ניתן לראות כי עד 1996 ישנה עליה במספר המורים בכל המחוזות למעט מחוז חיפה, אך החל משנת 1997 במחוז דרום, מרכז וירושלים מסתמנת עליה תלולה יותר ובשאר המחוזות העלייה, כאשר במחוז תל אביב ישנה ירידה בין השנים 1996-1998. בתרשים 6 ניתן לראות כי עד 1997 ישנה ירידה בשיעור גידול המורים ברוב המחוזות, והחל מ-1997 ישנה תנודתיות רבה שאינה מצביעה על מגמה ברורה.

V. בעיית חיזוי על סמך נתונים אגרגטיביים

אחד הכלים החשובים המשרתים את מערכת החינוך ומסייעים לה בתכנון יעיל ובהערכות נכונה בעוד מועד לקראת השינויים הצפויים הינו כלי חיזוי היקף כח אדם בהוראה. כלי זה אמור לספק מידע מדויק במידת האפשר אודות שינויים בהיקף כח אדם בהוראה תוך ניצול מידע מרבי על שינויים בגורמים המשפיעים עליו. בעייתיות החיזוי קשורה בעיקר בכמות המידע הקיים, בזמינותו ובמידת אמינותו. שיטות רבות לבניית מודלים לחיזוי תופעה כלשהי ידועות בספרות מדעית ומקצועית. מודלים ליניאריים לחיזוי המגמה, Croxton, Cowden and Klein (1968) מודלים לחיזוי על בסיס רגרסיה ליניארית Spencer, Clark and Hoguet (1961), Robinson (1971) מודלים סטוכסטיים לניתוח סדרות עתיות Harvey (1989), Box and Jenkins (1970) ורבים אחרים עוסקים בהתאמת מודלים לתחזיות על סמך נתונים אגרגטיביים. נשים לב כי שימוש במודלים ליניאריים אינו מהווה פתרון עבורנו ממספר סיבות. ראשית, ברשותנו אחד עשר תצפיות בלבד, דבר שפוגע במובהקות המודל. בנוסף, תוצאות מודל רגרסיה ליניארית הבודק קשר בין מספר המורים לבין משתנים מסבירים, כגון מספר התלמידים, גודל אוכלוסייה ומשתנים כמותיים רלוונטיים אחרים[1] אינן אמינות, מאחר וקיימת תלות חזקה בין התצפיות (מספר מורים בכל שנה). יתרה מזאת, הנחת הליניאריות גוררת הנחת שיעורי גידול קבועים במספר המורים, וכפי שראינו בפרק של סטטיסטיקה תיאורית, שיעורי הגידול הנצפים בתקופה 1991-2001 נעים במגמת ירידה. גישה נוספת של התאמת מודל אוטורגרסיבי האומד את שיעור הגידול במספר מורים כפונקציה של משתנים מסבירים[2] גם היא בעייתית, היות ועל סמך כמות מוגבלת של תצפיות נקבל אומדנים בלתי עמידים ומובהקות המודל בלתי מספקת. לאור הבעיות שהוזכרו לעיל ננקטה כאן גישה חלופית של ניתוח נתוני פרט והתאמת מודלים לוגיסטיים המתבססים על ניתוח הסתברויות מעבר של המורים משנה לשנה.

[1] מודלים ליניאריים אלה בוצעו במסגרת הפרוייקט "חיזוי צרכי הוראה "(1996) ע"י המכון לחיזוי טכנולוגי ליד אוניברסיטת ת"א

[2] Projections of Education Statistics to 2005, NATIONAL CENTER FOR EDUCATION STATISTICS(1995), U.S. Department of Education, Office of Educational Research and Improvement.

VI. מעברי מורים במערכת החינוך – בדיקת יציבות הנתונים

1. חישוב פרופורציות מעבר

על מנת לנתח את מעברי המורים במערכת החינוך נשתמש במספר הגדרות שילוו אותנו בהמשך המחקר. נתייחס למורה i בשלב חינוך k , $k = 1,2,3$, במחוז l , $l = 1,\ldots,6$ בשתי שנים עוקבות j ו- $j+1$.

נסמן תא $\{kl\}$ כתת מערכת חינוך בשלב חינוך k ובמחוז l.

כעת נתאר את כל המצבים האפשריים של המורה בתקופה דו שנתית זו.

בהתייחס לאפשרויות נשירה מהמערכת בשנה j , ייתכנו שני מצבים משלימים זה את זה:

- מצב א-1, כאשר מורה מסוים היה שייך לתא $\{kl\}$ בשנה j ולא הופיע ברשימות מורים בתא זה בשנה $j+1$.
- מצב א-2, כאשר מורה זה ימשיך להשתייך לתא $\{kl\}$ בשנה $j+1$.

בהתייחס לאפשרויות הכניסה של מורה לתא $\{kl\}$ בשנה $j+1$ ייתכנו שני מצבים משלימים זה את זה:

- מצב ב-1, כאשר מורה מסוים השייך לתא $\{kl\}$ בשנה $j+1$ לא השתייך אליה בשנה j , כלומר הצטרף למערכת בשנה $j+1$.
- מצב ב-2, הזהה למצב א-2, כאשר מורה מסוים שייך לתא $\{kl\}$ בשנה $j+1$ והיה בתא זה גם בשנה הקודמת j .

המצב שבו המורה לא הופיע בתא $\{kl\}$ בשתי השנים העוקבות אינו מעניינו, כי איננו שייך לקבוצת מצבי מעבר אפשריים בתקופה הדו שנתית. נגדיר כעת שני משתנים דיכוטומיים שיתארו את כל מצבי המעבר האפשריים עבור מורה מסוים.

נגדיר אינדיקטורים המתארים את המצבים שהוסברו לעיל:

(VI-1.1) יהי $\left(I_{out}\right)_{ijkl}$ אינדיקטור מצבים א-1 וא-2 עבור מורה i לתא $\{kl\}$ השווה ל-

1 - כאשר המורה i נמצא בתא $\{kl\}$ בשנה j ואינו נמצא בתא זה בשנה $j+1$.

0 - כאשר המורה i נמצא בתא $\{kl\}$ בשנה j ונמצא בתא זה גם בשנה $j+1$

(VI-1.2) ויהי $\left(I_{in}\right)_{i(j+1)kl}$ אינדיקטור מצבים ב-1 וב-2 עבור מורה i לתא $\{kl\}$ השווה ל-

1 - כאשר המורה i אינו נמצא בתא $\{kl\}$ בשנה j ומצטרף לתא זה בשנה $j+1$

0 - כאשר המורה i נמצא בתא $\{kl\}$ בשנה j ונמצא בתא זה גם בשנה $j+1$

כעת, בהינתן כי מספר המורים בשנה j בתא $\{kl\}$ הינו N_{jkl}, ומספר המורים באותו תא בשנה $j+1$ הוא N_{j+1}, הרי כי פרופורציית המורים מתוך סך מורים השייכים לתא $\{kl\}$ בשנה j שלא עברו לשנה $j+1$, כלומר נשרו תא $\{kl\}$ היא:

(VI-1.3) $$\pi^{out}_{j,(j+1)kl} = \frac{\sum_{i=1}^{N_{jkl}} \left(I_{out}\right)_{ijkl}}{N_{jkl}}.$$

ובאופן דומה פרופורציית המורים שהצטרפו למערכת בשנה $j+1$ היא

(VI-1.4) $$\pi^{in}_{j,(j+1)kl} = \frac{\sum_{i=1}^{N_{(j+1)kl}} \left(I_{in}\right)_{i(j+1)kl}}{N_{(j+1)kl}}.$$

לכן, עבור כל תא $\{kl\}$ ולכל תקופה של שתי שנים עוקבות $j, j+1$ מתקיים:

(VI-1.5) $$N_{(j+1)kl} = \left(1-\pi^{out}_{j,(j+1)kl}\right)N_{jkl} + N_{(j+1)kl}\pi^{in}_{j,(j+1)kl}$$

מכאן, כי מספר מורים בתא $\{kl\}$ בשנה $j+1$ הוא:

(VI-1.6) $$N_{(j+1)kl} = N_j \upsilon_{j,(j+1)kl}, \quad \text{כאשר} \quad \upsilon_{j,(j+1)kl} = \frac{1-\pi^{out}_{j,(j+1)kl}}{1-\pi^{in}_{j,(j+1)kl}}$$

כעת, בהינתן פרופורציות מעבר של המורים בתא $\{kl\}$ משנת 2001 לשנת 2002 $\upsilon_{2001,2002kl}$, ובהינתן מספר המורים בתא $\{kl\}$ בשנת 2001 N_{2001kl} נוכל בקלות לחזות את מספר המורים בשנת 2002

(VI-1.7) $$\hat{N}_{2002kl} = N_{2001}\upsilon_{2001,2002kl}$$

2. ההנחה הבסיסית של יציבות פרופורציות מעבר

מטרת שלב זה של המחקר היא לבדוק את יציבות מעברי המורים בתקופה 1991-2001 וזאת מתוך ההנחה הבסיסית כי מעברי המורים בין שנה j לבין שנה $j+1$, $j = 1991,\ldots,2001$ בכל תא $\{kl\}$ ניתנים להסבר בעזרת במאפיינים האישיים והמקצועיים של המורים בלבד (נסמן הנחה זו כקבוצת השערות $H_o^{j,(j+1)kl}$). על מנת להשתמש בהנחה זו בהמשך ולהתאים מודל שיסביר את מעברי המורים עלינו לבדוק האם היא מתקיימת. לצורך כך, נאמוד את שונות פרופורציות המעבר של המורים באמצעות דגימות מקריות פשוטות (עם החזר) שוות הסתברות של מורים מתוך הקבצים הקיימים וננתח את תוצאותיהן. מאחר ודגימות המורים נעשות על בסיס הסתברות שווה לכל מורה להיכנס למדגם, שונות גבוהה של תצפיות הדגימות תביא לדחיית השערת ותצביע על אפשרות של קיום השפעה חיצונית כלשהי

על פרופורציות מעבר של המורים (נסמן אפשרות זו כקבוצת השערות אלטרנטיביות $H_1^{j,(j+1)kl}$). לחילופין, במידה ונקבל שונות נמוכה של תצפיות הדגימה לא נדחה את השערת הבסיס ונוכל לנתח את פרופורציות מעבר של המורים באמצעות מאפייניהם.

3. בדיקת יציבות מעברי מורים: שלב 1 - דגימת מורים בחתך דו שנתי

עבור כל קובץ נתוני מורים דו שנתית (למשל 1991-1992) חושבו ערכי $\pi^{out}_{j,(j+1)kl}$ ו- $\pi^{in}_{j,(j+1)kl}$ בעזרת $(I_{out})_{ijkl}$ ו- $(I_{in})_{i(j+1)kl}$ כמוגדר ב- **(VI-1.1,1.2)** ו- $\upsilon_{j,(j+1)}$, $j = 1991,..,2001$ (פרופורציות המעבר של המורים משנה לשנה בכל תקופה דו-שנתית). כעת בכל קובץ דו שנתי בוצעה דגימת מורים מקרית שוות הסתברות (עם החזר) בעזרת סימולציית מונטה-קרלו, כך שמכל קובץ דו שנתי מקורי נתקבלו שנים עשר מדגמי מורים בחתך דו-שנתי בתקופה של 1991-2001. לכל תקופה דו-שנתית, עבור כל המדגמים שנתקבלו מסימולציה חושבו פרופורציות מעבר של המורים משנה לשנה וחושבו ממוצעים וסטיות תקן. בדיקת השערת היציבות התבססה על עקרון מספר תצפיות חריגות שיצאו מחוץ למרווח של שתי סטיות תקן מהממוצע.

דגימת המורים מתוך קבצים דו-שנתיים התבצעה בהסתברות דגימה השווה ל- $\frac{1}{N_{j,(j+1)kl}}$, כאשר $N_{j,(j+1)kl}$ הוא גודל המדגם בתא $\{kl\}$ והוא מציין את מספר המורים שנמצאים לפחות באחת מהשנים $j, j+1$ בתא $\{kl\}$. עבור נתוני המורים המקוריים ושנים עשר מדגמי מורים שנתקבלו מסימולציה בחתך הדו-שנתי $j, j+1$ חושבו $\pi^{out(n)}_{j,(j+1)kl}$, $\pi^{in(n)}_{j,(j+1)kl}$, $\upsilon^{(n)}{}_{j,(j+1)kl}$ על פי הגדרות **(VI-1.3, 1.4, 1.6)** לכל תא $\{kl\}$ ומדגם n-י, $n = 1,2,\ldots,13$, (עבור המדגם המקורי $n = 1$).

כעת, חושבו ממוצעים ושונויות עבור כל $\upsilon^{(n)}{}_{j,(j+1)kl}$ ונעשה מבחן לבדיקת היציבות פרופורציות המעבר של המורים. על מנת לדחות את $H_o^{j,(j+1)kl}$, נדרוש עבור $\upsilon^{(n)}{}_{j,(j+1)kl}$

$$\sum_{n=1}^{13} I\left\{\upsilon^{(n)}{}_{j,(j+1)kl} \notin \left(\overline{\upsilon}_{j,(j+1)kl} - Z_{0.975}\hat{s}_{\upsilon}, \overline{\upsilon}_{j,(j+1)kl} + Z_{0.975}\hat{s}_{\upsilon}\right)\right\} \geq n_0 , \qquad \textbf{(VI-3. 1)}$$

כאשר

$I\{A\}$ -אינדיקטור של קבוצה A ,

$Z_{0.975}$ הוא ערך התפלגות נורמלית המתאים לאחוזון ה-97.5% , $Z_{0.975} \approx 1.96$,

n_0 נקבע מראש ככלל הכרעה טהור (לצורך המחקר הנוכחי בחרנו $n_0 = 3$)

ו- $\overline{\upsilon} = \frac{1}{13}\sum_{n=1}^{13}\upsilon^{(n)}{}_{j,(j+1)kl}$, $\hat{s}_{\nu}^2 = \frac{1}{12}\sum_{n=1}^{13}\left(\upsilon^{(n)}{}_{j,(j+1)kl} - \overline{\upsilon}_{j,(j+1)kl}\right)^2$.

4. בדיקת יציבות מספר מורים: שלב 2 - דגימת מורים מתוך קובץ רב שנתי

בשלב זה נבדקה יציבות נתונים אודות מספר המורים בכל שנה בכל שלב חינוך ובכל מחוז. קבצים של נתוני כל המורים בכל שלבי החינוך ובכל המחוזות אוחדו לתוך קובץ אחד רב שנתי, ובו N מורים שהשתייכו למערכת החינוך הכללית לפחות בשנה אחת בתקופה 1991-2001 וממנו נעשתה דגימת מורים על בסיס הסתברות שווה של $1/N$. כתוצאה מכך נתקבלו שנים עשר מדגמים הכוללים נתוני מורים רב שנתיים דמויי הקובץ המקורי. עבור כל מדגם חושב מספר המורים בכל שנה j, $j=1991,\ldots,2001$ עבור כל תא $\{kl\}$. כעת, נבדוק את השערת הבסיס ליציבות נתוני המורים, כאשר המבחן מבוצע על מספר המורים בתאי $\{kl\}$ באופן הבא:

יהי $N_{jkl}^{(n)}$ מספר המורים הנצפה בשנה j בתא $\{kl\}$ במדגם n

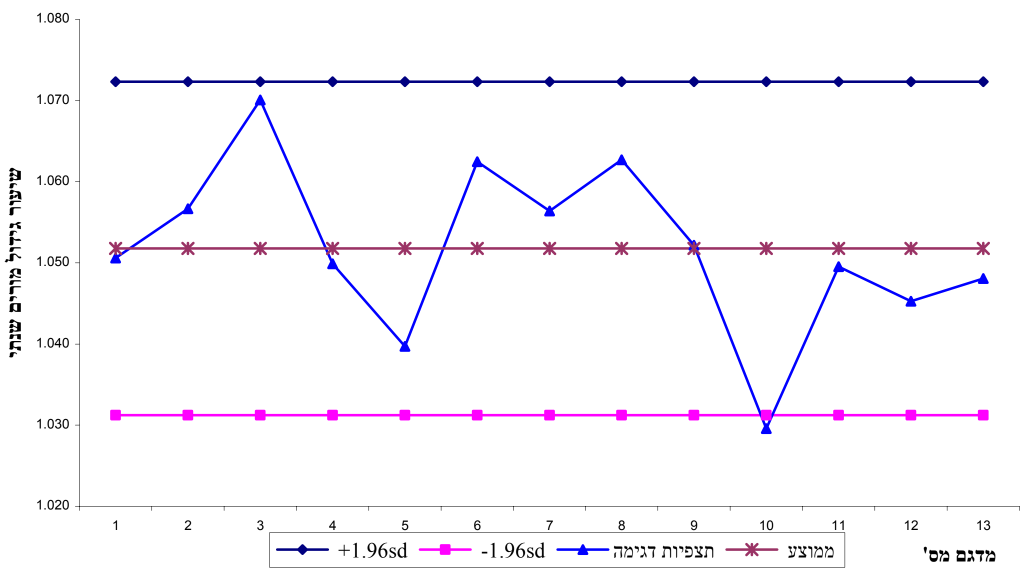
הפעם, על מנת לדחות את השערת $H_o^{j,(j+1)kl}$ לגבי מספר המורים בכל תא$\{kl\}$, בדומה ל- **(VI-3.1),** נדרש כי :

$$\sum_{n=1}^{13} I\left\{\widetilde{N}_{jkl}^{(n)} \notin \left(\overline{N}_{jkl} - Z_{0.975}\hat{s}_{\widetilde{N}_{kl}}, \overline{N}_{jkl} + Z_{0.975}\hat{s}_{\widetilde{N}_{kl}}\right)\right\} \geq n_0, \qquad \textbf{(VI-4.1)}$$

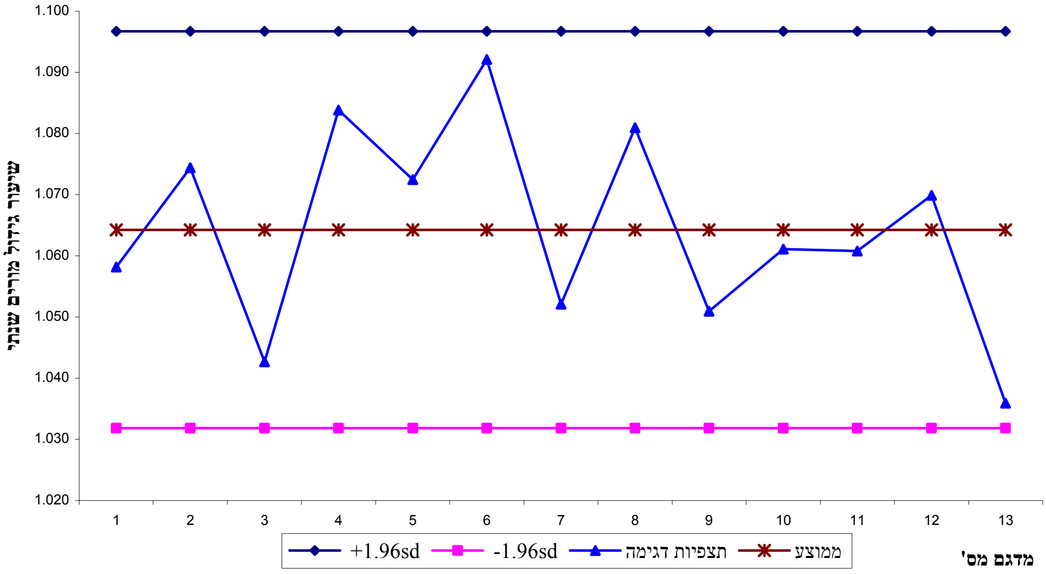
כאשר $\overline{N}_{jkl} = \frac{1}{13}\sum_{n=1}^{13} N^{(n)}{}_{jkl}$, $\hat{s}_{\widetilde{N}_{jkl}}^2 = \frac{1}{12}\sum_{n=1}^{13}\left(\widetilde{N}_{jkl}{}^{(n)} - \overline{N}_{jkl}\right)^2$ ו- $n_0 = 3$.

להלן מובאים תרשימים המדגימים את תוצאות סימולציות מונטה-קרלו שנעשו . דוגמאות אלה ממחישות היטב את מבחן יציבות פרופורציות מעבר ומספר המורים בתקופה 1991-2001.

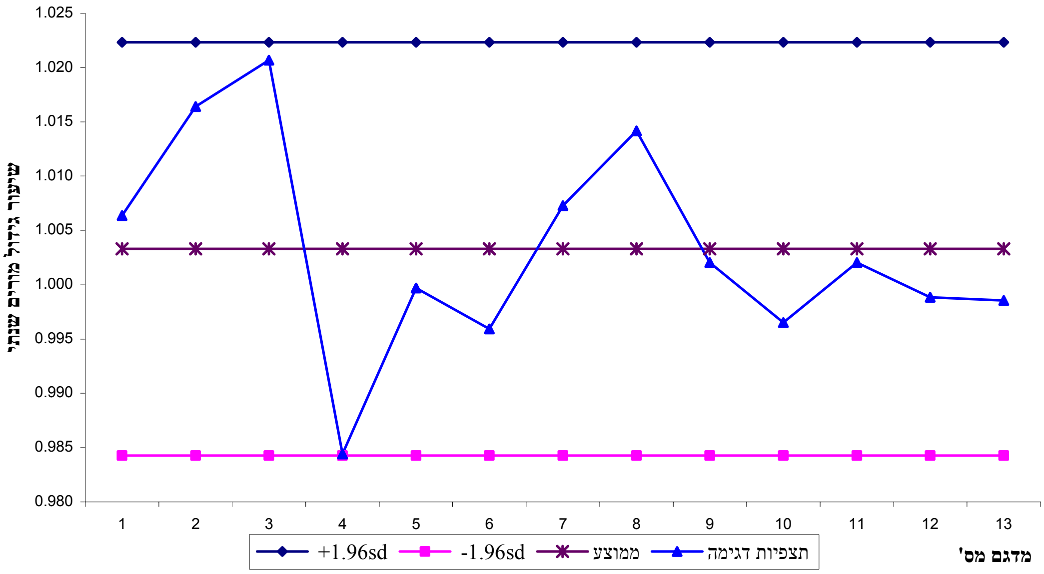
תרשים 7: תוצאות סימולציות מונטה-קרלו –שיעור גידול מורים שנתי-חטיבה עליונה 91-92 – מחוז ירושלים



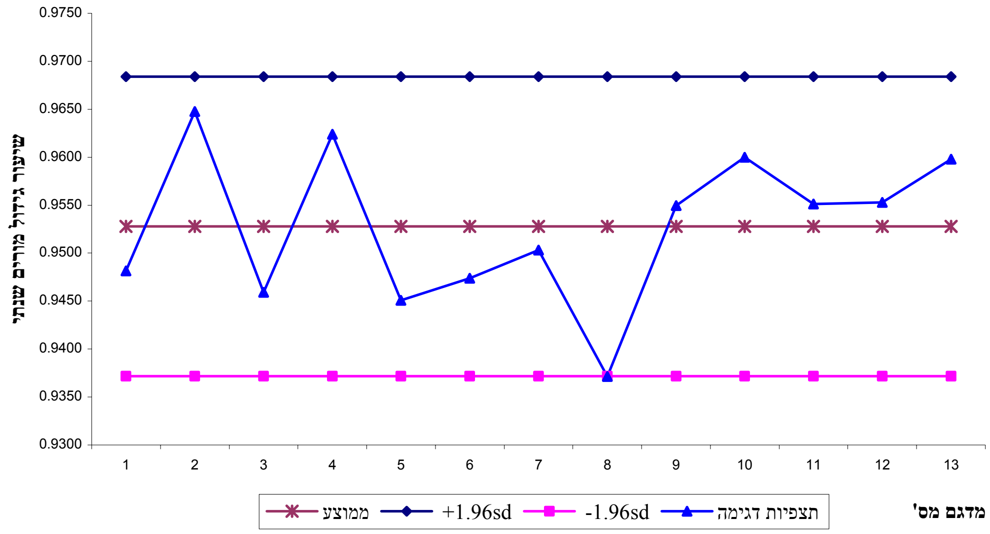
תרשים 8: תוצאות סימולציות מונטה-קרלו –שיעור גידול מורים שנתי-חטיבה עליונה 93-94 – מחוז צפון



תרשים 9: תוצאות סימולציות מונטה-קרלו –שיעור גידול מורים שנתי-חטיבה עליונה 94-95 – מחוז חיפה

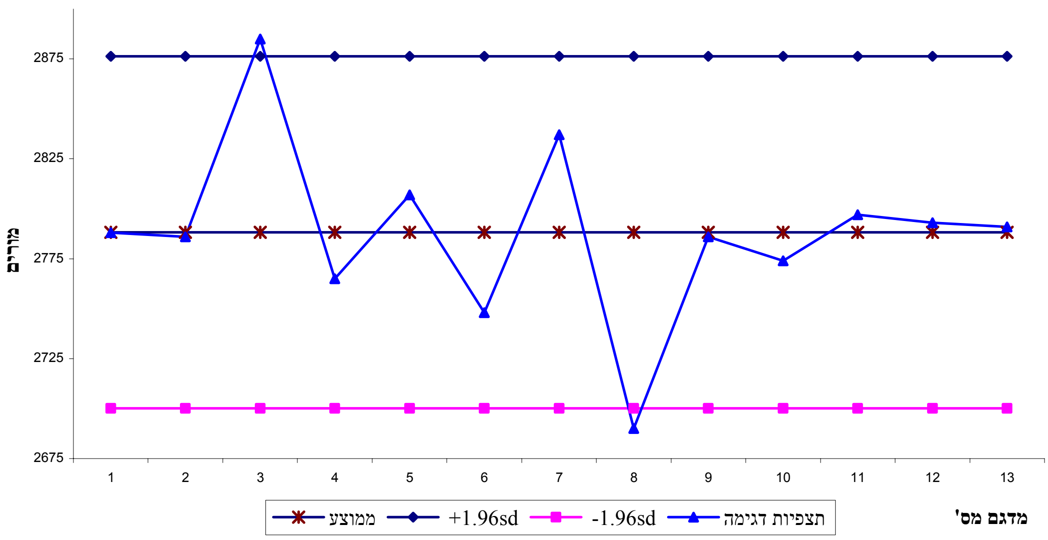


תרשים 10 : תוצאות סימולציות מונטה-קרלו –שיעור גידול מורים שנתי-חטיבה עליונה 96-97 – מחוז ירושלים

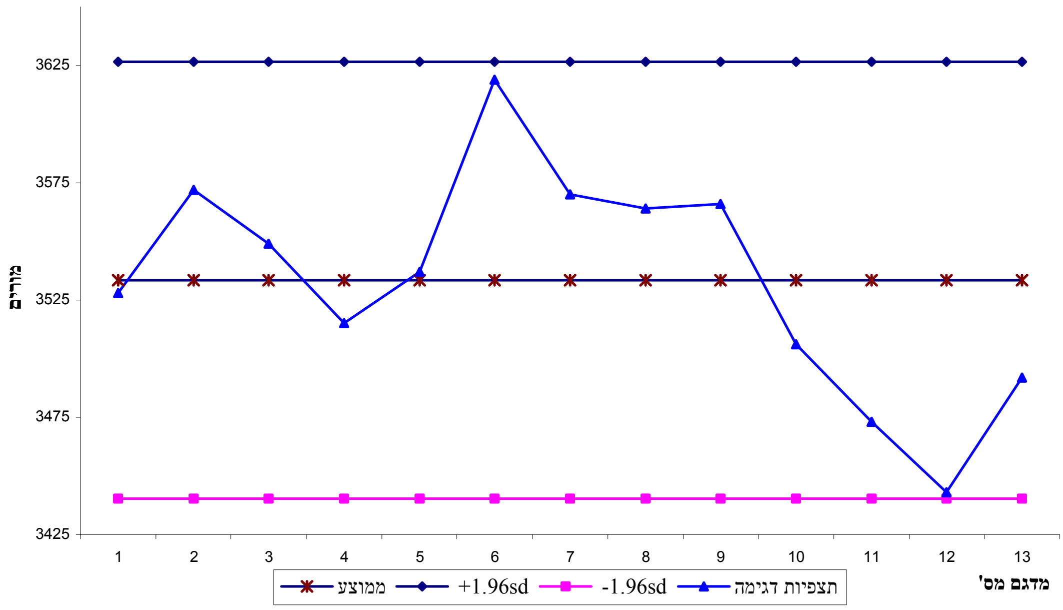


נשים לב כי נתגלו מספר תצפיות קריטיות של פרופורציות מעבר שחושבו מתוך מדגמים אחדים, ואף אלה החורגות מטווח רווח הסמך לממוצע תצפיות (ראה דוגמאות של תצפיות חורגות: תרשים 7- תצפית 10, תרשים 10- תצפית 8, תצפיות קריטיות: תרשים 7 - תצפית 3, תרשים 9- תצפיות 3 ו- 4). עדיין, תוצאות מבחני יציבות לפרופורציות מעבר של המורים תומכות בהשערת היציבות של פרופורציות מעבר על פי קריטריון **(VI-3.1)**.

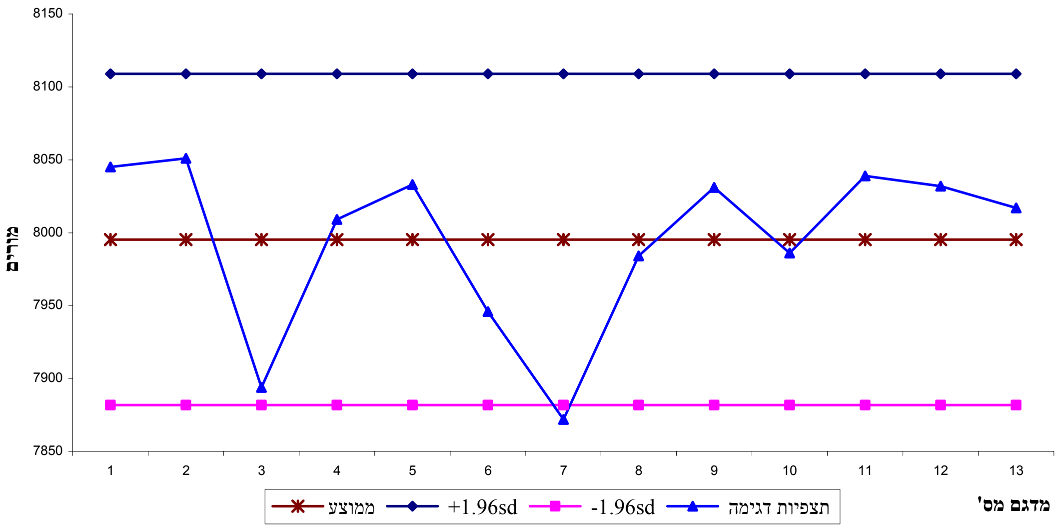
תרשים 11: תוצאות סימולציות מונטה-קרלו –דגימת מורים-חטיבה עליונה 1991 – מחוז ירושלים



תרשים 12: : תוצאות סימולציות מונטה-קרלו –דגימה חוזרת של מורים-חטיבה עליונה 1997 – מחוז ירושלים



תרשים 13: : תוצאות סימולציות מונטה-קרלו –דגימה חוזרת של מורים-חטיבה עליונה 2000 – מחוז מרכז



בתרשימים 11-13 ניתן לראות את תוצאות דגימות מורים מתוך הקובץ הרב שנתי הכולל את כל הנתונים ההיסטוריים בין 1991-2001. באופן כללי, על פי קריטריון **(VI-4.1)** , תוצאות הדגימה תומכות בהשערות היציבות $H_o^{j,(j+1)kl}$, כאשר באופן פרטני ניתן לזהות במדגמים אחדים תצפיות חריגות וקריטיות המספקות עדות לבעיית היציבות. ראה לדוגמא תצפיות חורגות: תרשים 11- תצפיות 3 ו-8, תרשים 13- תצפית 7, תצפיות קריטיות: תרשים 12- תצפיות 6 ו-12, תרשים 13- תצפית 3. במסגרת בדיקת יציבות כוללת לא צפינו לקבל מספר גבוה של תופעות קריטיות והריגות ועל פי הקריטריונים

שקבענו אין במספר קטן של תצפיות חריגות וקריטיות כדי לדחות את השערות $H_o^{j,(j+1)kl}$. חשוב לציין כי כאשר $H_o^{j,(j+1)kl}$ אינה נדחית, אין להסיק מכך על אי קיום של השפעה חיצונית כלשהי על מעבר המורים משנה לשנה, אלא במקרה זה אימצנו את ההנחה של חוסר השפעה חיצונית מובהקת על פרופורציות מעבר המורים לאורך התקופה הנחקרת.

VII. בדיקת אמינות נתוני קבצים

1. שיעור גידול שלילי במספר המורים ובעיית אי דיווח

בשלב זה של עבודת המחקר היה מעניינו לבחון את אמינות נתוני המורים ולאתר פרופורציות מעבר בעייתיות. על פי **(VI-1.6)** $\nu_{j(j+1)kl}$ מציין את פרופרציית מעבר של המורים משנה j לשנה $j+1$ או במילים אחרות את שיעור הגידול במורים בין השנים הללו. ברור, כי אם $\nu_{j(j+1)kl} < 1$, מספר המורים בתא$\{kl\}$ בשנה $j+1$ קטן מזה שבשנה j ולכן מקובל לומר כי שיעור הגידול במורים בשנים אלה בתא זה הינו שלילי. שיעורי גידול שליליים במספר המורים שנצפו בתאים מסוימים ובשנים מסוימות לאורך תקופת המחקר מחייבים בדיקת אמינות הנתונים, היות ובתנאים של גידול באוכלוסייה ובמספר התלמידים ולאור קיום ההנחה כי אין השפעה חיצונית מובהקת על מעברי המורים, לא היינו מצפים לשיעור גידול שלילי במספר מורים. אי לכך, כל שיעור גידול שלילי הנצפה בתא כלשהו במעבר משנה לשנה לאורך תקופת המחקר מחייב בדיקת אמינות הנתונים בשנה הספציפית בה מתרחשת ירידה משמעותית במספר המורים. בדיקת האמינות קשורה למעשה לבדיקת ליקויים בדיווח מורים. דיווח לקוי בשנים מסוימות, הנובע מטעויות אקראיות אינו מהווה בעיה סטטיסטית הקשורה להשפעה על התפלגות המשותפת של מאפייני המורים באותן השנים. לעומת זאת, ליקוי מגמתי או מכוון בדיווח, כמו, למשל, אי דיווח מכוון על מורים בעלי השכלה מסוימת, ותק מסוים או מאפיינים ספציפיים אחרים עלול ליצור הבדל משמעותי בין ההתפלגויות המשותפות של המאפיינים בשנה הבעייתית לבין אלה בשנה הקודמת. שוני מובהק בין ההתפלגויות המשותפות מאפייני המורים עלול לגרום להטיות משמעותיות באומדני הסתברויות מעבר של המורים ולכן יש לבדוק האם הירידה הבלתי הסבירה במספר מורים מקורה בליקוי מגמתי.

2. מבחן לאיתור שיעורי גידול שליליים מובהקים

במסגרת בדיקת אמינות הנתונים היה עלינו להחליט אילו מהתאים מהווים בעיה פוטנציאלית הקשורה לירידה משמעותית של מספר המורים, כלומר אילו משיעורי הגידול הנצפים הינם שליליים באופן מובהק. השתמשנו ברווחי הסמך לפרופורציות המורים שהתקבלו מדגימת מורים בפרק בדיקת יציבות הנתונים (ראה תת פרק VI-3) על מנת לזהות תאים (מחוזות ושלבי חינוך) בהם שיעור הגידול הינו שלילי באופן מובהק. ביצענו מבחן סטטיסטי לפיו דחינו את ההשערה (ברמת מובהקות של 0.05) כי שיעור הגידול

בתא $\{kl\}$ הינו שלילי (קטן מ-1) אם החסם העליון של רווח סמך לממוצע שיעורי גידול מורים שנתיים היה גדול ממש מ-1, דהיינו:

$$\overline{\upsilon}_{j,(j+1)kl} + Z_{0.975}\hat{s}_{\upsilon} > 1 \qquad \textbf{(VII-2.1)}$$

3. בעיית בדיקת אמינות נתוני הקבצים

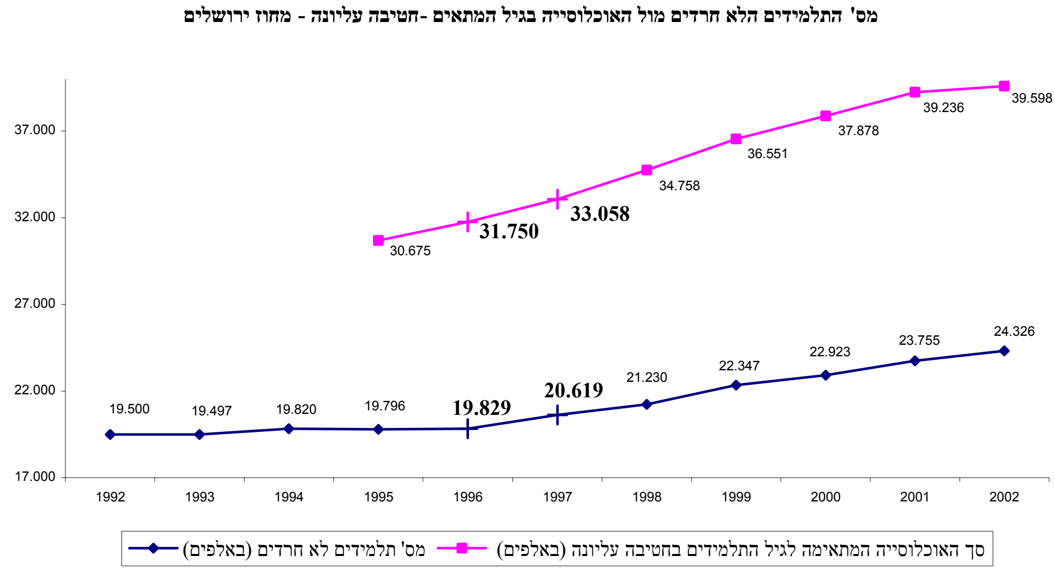
בטבלה 4 ניתן לראות את רווחי הסמך של שיעורי גידול בחטיבה עליונה כדוגמה הממחישה את בדיקת המובהקות של שיעור גידול שלילי. רווחי הסמך לשיעורי גידול שנתיים נבנו עבור כל תא על סמך הנתונים שנתקבלו מדגימות המורים (ראה תתי פרקים VI-3,4). התאים המוצללים מצביעים על שיעורי גידול שליליים מובהקים (קטנים מ-1) על פי מבחן **(VII-2.1)**. ניקח לדוגמא מחוז ירושלים בחטיבות עליונות. רווח הסמך לפרופורציית מעבר בתא זה בשנים 1996-1997 מצביע על שיעור הגידול שלילי מובהק (0.9372 0.9684), בעוד שבמעבר בין השנים 1996-1997 חל גידול באוכלוסייה ובמספר התלמידים תוספת המורים במעבר מ-1996 ל-1997 הייתה שלילית (ראה תרשים 14).

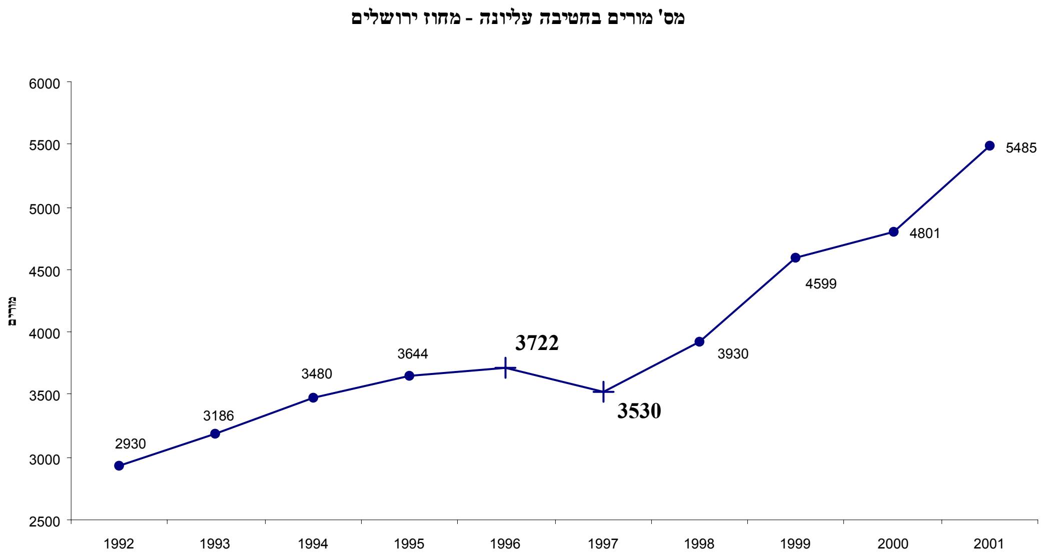
טבלה 4: רווחי סמך לפרופורציות מעבר בקרב מורים – חטיבה עליונה

(התאים המוצללים מציינים שיעורי גידול מורים שליליים מובהקים, רמת מובהקות $\alpha = 0.05$)

מחוז		1995-96	1996-97	1997-98	1998-99	1999-2000	2000-01
מחוז 1 - ירושלים	$\nu + 1.96\,\sigma$	1.0472	0.9684	1.1390	1.2097	1.0638	1.1587
	ν	1.0218	0.9528	1.1107	1.1738	1.0441	1.1420
	$\nu - 1.96\,\sigma$	0.9964	0.9372	1.0823	1.1380	1.0244	1.1254
מחוז 2 - צפון	$\nu + 1.96\,\sigma$	1.0160	1.0714	0.9922	1.1770	1.0249	1.1290
	ν	0.9951	1.0494	0.9781	1.1560	1.0064	1.1083
	$\nu - 1.96\,\sigma$	0.9743	1.0274	0.9639	1.1350	0.9879	1.0875
מחוז 3 - חיפה	$\nu + 1.96\,\sigma$	1.0001	1.0016	1.0570	1.1152	1.0003	1.0903
	ν	0.9848	0.9770	1.0358	1.0968	0.9756	1.0679
	$\nu - 1.96\,\sigma$	0.9694	0.9524	1.0145	1.0783	0.9509	1.0456
מחוז 4 - מרכז	$\nu + 1.96\,\sigma$	1.0100	1.0602	1.0155	1.1506	1.0434	1.0669
	ν	1.0018	1.0413	0.9991	1.1355	1.0262	1.0528
	$\nu - 1.96\,\sigma$	0.9935	1.0224	0.9826	1.1203	1.0090	1.0387
מחוז 5 - ת"א	$\nu + 1.96\,\sigma$	1.0362	0.9685	1.0045	1.0953	0.9775	1.0433
	ν	1.0217	0.9521	0.9875	1.0842	0.9629	1.0321
	$\nu - 1.96\,\sigma$	1.0071	0.9357	0.9704	1.0732	0.9483	1.0208
מחוז 6 - דרום	$\nu + 1.96\,\sigma$	1.0215	1.0591	1.0654	1.1375	1.0377	1.1326
	ν	1.0005	1.0421	1.0503	1.1217	1.0203	1.1201
	$\nu - 1.96\,\sigma$	0.9795	1.0251	1.0353	1.1059	1.0028	1.1075

תרשים 14: גידול במספר התלמידים ובאוכלוסייה לעומת הגידול במורים – חטיבה עליונה – מחוז ירושלים





כעת נותר לבדוק האם בתאים בהם נצפה שיעור גידול שלילי בשנים מסוימות קיים חוסר דיווח מגמתי אודות מורים בעלי תכונות מסוימות המשפיע באופן מובהק על התפלגות מאפייניהם באותה שנה. בהתבסס על הנחת שוויון ההתפלגויות המשותפות של מאפייני המורים בכל שנתיים עוקבות (ראה תת פרק VII-1), היה עלינו לבדוק קיום הנחה זו עבור כל תא בו נצפה שיעור גידול מורים שלילי בתקופה הנחקרת. לשם כך נבדקה ההשערה הסטטיסטית האם התפלגות משותפת של מאפייני המורים בשנה שקדמה לגידול שלילי זהה לזו של השנה העוקבת. במקרה של דחיית ההשערה בדבר שוויון התפלגויות המלצנו על גריעת נתוני השנה בה חל קיטון במספר מורים בתא מסוים. לצורך השוואת התפלגויות אלה נבחרו מאפייני מפתח של המורים: גיל, ותק, השכלה, ושעות לימוד שבועיות. הקושי העיקרי בביצוע

תהליך השוואה היה נתון בכך שמאפייני המפתח נתגלו כמתואמים ביניהם, אי לכך, לא ניתן היה להפעיל את המבחן הסטנדרטי להשוואת התפלגויות משותפות (χ^2) והיה צורך למצוא מבחן מיוחד המתאים להשוואת התפלגויות משותפות המורכבות מווקטורים מתואמים. לשם כך אנו מביאים פרק העוסק במציאת פתרון תיאורטי פורמלי לבעיה הספציפית העומדת במחקרנו.

VIII. היבטים תיאורטיים בבדיקת הומוגניות של מדגמים סטטיסטיים

1. הגדרת הבעיה

הבעיה העומדת לפנינו קשורה למעשה להשוואת התפלגויות והיא מחייבת לבחון את הסוגייה הסטטיסטית הבאה: נניח מרחבים סטטיסטיים $A_r, r = 1,\ldots ,R$. בכל אחד מהמרחבים הנ"ל מתפלגים N משתנים X_i , $i = 1,2,\ldots ,N$. מעונינים לבדוק ההשערה הבאה:

$$H_0 :\ P(X_1, X_2,\ldots ,X_N|A_r) = P(X_1, X_2,\ldots ,X_N|A_l), \forall\ r \neq l,\ \ r,l = 1,\ldots ,R \qquad \textbf{(VIII-1.1)}$$

כנגד השערה אלטרנטיבית פשוטה H_1,

$$H_1 :\ P(X_1, X_2,\ldots ,X_N|A_r) \neq P(X_1, X_2,\ldots ,X_N|A_l), \forall\ r \neq l,\ \ r,l = 1,\ldots ,R \qquad \textbf{(VIII-1.2)}$$

כאשר $P(X_1, X_2,\ldots , X_N|A_r)$ הינה התפלגות מותנית בבחירת מדגם $A_r, r = 1,\ldots ,R$.

נשים לב כי בדיקת השערה סטטיסטית זו כרוכה במספר קשיים:

1. קיים קושי בלבדוק השערה זו עבור N גדול
2. X_i -ים יכולים להיות תלויים ביניהם.

אנו מציעים שיטה לבדיקת השערה זו תוך שימוש במספר פרוצדורות סטטיסטיות פשוטות.

תחת השערת H_0 מתקיים "$A_r = A_l$" בהקשר להתפלגות של X_i . אם נאחד את המרחבים A_r לתוך מרחב אחד $A = \bigcup_{r=1}^{R} A_r$. במטרה להקטין את מספר המשתנים במרחב A נבנה קומבינציות ליניאריות

$$W_j = \sum_{i=1}^{N} \alpha_{ji} X_i\ ,\quad j = 1,\ldots ,K,\quad K \leq N \qquad \textbf{(VIII-1.3)}$$

בלתי תלויות (בקירוב).

באמצעות שיטת ניתוח מרכיבים ראשיים (Principle Component Analysis).

נניח כי $W_j(A_r)$ השלכת גורם W_j במרחב A_r , $r = 1,\ldots ,R,\quad j = 1,\ldots ,N$.

כעת נבדוק את מערכת ההשערות:

$$H_{0jrl} :\quad P\big(W_j(A_r)\big) = P\big(W_j(A_l)\big),\quad j = 1,\ldots ,K,\ r \neq l = 1,\ldots ,R \qquad \textbf{(VIII-1.4)}$$

באמצעות המבחנים הבאים:

1. T-test בהנחה כי W_j מתפלג נורמלית

2. מבחנים אי פרמטריים כדוגמת Wilcoxon, Kolmogorov-Smirnov

נדחה את H_0 אם"ם נדחה את אחת מההשערות H_{0jrl}.

נשים לב כי שימוש בניתוח מרכיבים ראשיים (Principle Component Analysis) מאפשר מציאת מקדמים $\alpha_i, i = 1,\ldots,N$

ב- **(VIII-1.3)** , כך ש- $E(W_i W_j) = EW_i EW_j$. בהמשך נראה כי שימוש בשיטה זו הינו נכון בהנחה שההתפלגות P מוגדרת על ידי 2 מומנטים.

2. שימוש בשיטת ניתוח מרכיבים ראשיים

נגדיר:

$$a_{m_1 \ldots m_N} = \int_{R^N} x_1^{m_1} \ldots x_N^{m_N} P(dx_1,\ldots,dx_N),$$
$$A_n = a_{n0\ldots 0} + a_{0n0\ldots 0} + \ldots + a_{0\ldots 0n}, \qquad \textbf{(VIII-2.1)}$$

כאשר $\Re^N$ - מרחב מספרים ממשיים ממדי N - $\Re^N = \underbrace{(-\infty,\infty)\times(-\infty,\infty)\times\ldots\times(-\infty,\infty)}_{N}$

<u>טענה 1</u>. נניח כי

$$\sum_0^\infty \left(A_{2n}\right)^{-\frac{1}{2n}} = \infty, \qquad \textbf{(VIII-2.2)}$$

וקיימת סדרה $\{\alpha_{ji}\}$ $\quad j = 1,\ldots,N, \quad i = 1,\ldots,N,$

כך שמתקיים

$$EW_k^\gamma W_l^\rho = EW_k^\gamma EW_l^\rho, \quad k \neq l = 1,\ldots,N,$$
$$\gamma = 1,\ldots,N-1, \qquad p = 1,\ldots,N-\gamma, \qquad \textbf{(VIII-2.3)}$$

כאשר $W_k, k = 1,\ldots,N$ מוגדר ע"י **(VIII-1.3)**,

ואז קיים פונקציונל חד חד ערכי F , כך ש-

$$P(X_1 < x_1,\ldots,X_N < x_N) =$$
$$F(P(W_1 < z_1(x_1,\ldots,x_N)),\ldots,P(W_N < z_N(x_1,\ldots,x_N))), \qquad \textbf{(VIII-2.4)}$$

כאשר $z_m, m = 1,\ldots,N$ פונקציות מוגדרות על מרחב $\Re^N$

הוכחת טענה זו מופיעה בפרק XI, בנספח א'.

על פי טענה 1, בתנאי שהתפלגות P מוגדרת על ידי $N > 2$ מומנטים, אין זה מספיק להשתמש ב-Principle Component Analysis, אלא יש צורך למצוא $\alpha_i, i = 1,\ldots n$ מתוך **(VIII-1.3)** , כך ש- **(VIII-2. 3)** יתקיים.

שימוש במבחן זה מאפשר בקלות לקבל אפיונים הסתברותיים לבדיקת מערכת מורכבת של השערות.

לדוגמא: נחשב את רמת המובהקות של דחיית ההשערה. נדגים את נכונות השיטה בהנחה ש- $R = 2$

לכן ניתן לנסח את תת ההשערות H_{0jrl} באופן הבא:

$$H_{0j}: \quad P(W_j(A_1)) = P(W_j(A_2)), \quad j = 1,\ldots, K. \qquad \textbf{(VIII-2.5)}$$

ברור שרמת המובהקות מתקבלת מתוך:

$$\begin{aligned} \alpha &= P_{H_0}(reject \quad H_0) = 1 - P_{H_0}(no\ reject \quad H_0) \\ &= 1 - \left(1 - P_{H_{01}}(reject \quad H_{01})\right)\left(1 - P_{H_{02}}(reject \quad H_{02})\right)\ldots\left(1 - P_{H_{oK1}}(reject \quad H_{0K})\right) \\ &= 1 - (1 - \kappa_1)(1 - \kappa_2)\ldots(1 - \kappa_k), \end{aligned} \qquad \textbf{(VIII-2.6)}$$

כאשר $\kappa_j, j = 1,\ldots K$ הינה רמת מובהקות לבדיקת H_{0j}.

כעת, נשתמש בשיטת Bonferroni בהנחה כי $\kappa_1 = \kappa_2 = \ldots = \kappa_K \equiv \delta$, אזי מתוך **(VIII-2.6)** נקבל

$$\alpha = 1 - (1 - \delta)^K. \qquad \textbf{(VIII-2.7)}$$

דוגמא: נבחר $\alpha = 0.05$, אזי בהתאם **(VIII-2.7)** רמת המובהקות לדחיית תת ההשערות תהיה $\delta = 1 - 0.95^{1/K}$. נחשב p-value למבחן. נניח ש- $p_j, j = 1,\ldots, K$ הוא ערך p מתוך בדיקת תת ההשערות, $p_m = \min\{p_1, p_2, \ldots, p_K\}$. אם $\delta > p_m$ אזי לפחות באחד מתת המבחנים נדחה את השערת H_{0l} ולכן נדחה את H_0. אי לכך, מתוך **(VIII-2.7)** נקבל כי $p_m = 1 - (1 - "p_{value}")^{1/K}$ לכן

$$"p_{value}" = 1 - (1 - p_m)^K. \qquad \textbf{(VIII-2.8)}$$

הערה: נשים לב כי המבחן הסטנדרטי של נראות מרבית הבודק שוויון מטריצות שונויות משותפות ב- k מדגמים, כשהסטטיסטי שלו נתון על ידי ((Morisson (1967, p.153):

$$M = \left(\sum_{i=1}^{k}(N_i - 1)\right) ln|S| - \sum_{i=1}^{k}(N_i - 1) ln|S_i|, \qquad \textbf{(VIII-2.9)}$$

כאשר N_i מספר התצפיות של משתנים מקריים במדגם $i = 1,\ldots, k$,

S_i היא מטריצת שונויות משותפות של מדגם ו- $S = \sum_{i=1}^{k} \frac{n_i S_i}{\sum_{i=1}^{k} n_i}, \quad i = 1,\ldots, k$

מתיר תלות בין המשתנים המקריים, כי הרי במקרים של אי תלות אין בשימושו שום צורך. חשוב לציין גם כי מבחן זה מאבד כל משמעות עבור משתנים מקריים בעלי תלות חזקה, כמו, למשל, כאשר $X_p = \gamma X_q + c, p \neq q$, היות ואז $|S_i| = 0$. מגבלה זו של תלות חזקה מפנה אותנו במקרים אלו להשתמש בשיטת הקומבינציות הלינאריות באמצעות ניתוח מרכיבים ראשיים, המתמודדת גם עם מצבים של תלות חזקה בין המשתנים.

3. יישום השיטה לבדיקת אמינות נתוני הקבצים

להלן מספר ממצאים מבדיקת התפלגויות משותפות של מאפייני מורים לשנים 1996, 1997 עבור תא המתייחס לחטיבה עליונה במחוז ירושלים (ראה תת פרק VII-3). על מנת להשוות את התפלגויות מאפייני גיל וותק מורים שהם משתנים רציפים, ערכיהם קובצו למספר קטגוריות. תוצאות שני מבחני χ^2 מראים כי עבור מאפיין הגיל אין אנו דוחים את ההשערה כי התפלגויות משתנה הגיל בשתי השנים שוות (p-value = 0.2036), לעומת זאת עבור משתנה ותק אנו דוחים את ההשערה כי התפלגויות משתנה הוותק שוות (p-value=0.0067). נשים לב כי המתאם בין ותק המורים וגילם גדול (כ-0.65), יחד עם זאת משתנה אחד אינו נגזר ישירות ממשנהו. נשתמש בשיטה שהוצעה לעיל על מנת לבדוק השערה כוללת של שיווין התפלגויות משותפות של מאפייני מפתח: גיל, ותק, שעות שבועיות ודרגה (השכלת המורה). תחת השערת H_0 שכאמור מניחה שוויון התפלגויות משותפות של המאפיינים, אוחדו באופן לגיטימי שני המדגמים 1996 ו-1997 למדגם אחד A, כך שמקובץ 1997 הוצאו כל המורים ששייכים גם לקובץ 1996 על מנת לבטל תלות בין המדגמים. עבור מדגם זה נתקבלו שלושת הגורמים הבאים[3]:

$$W_1 = 0.92 * seniority + 0.93 * age + 0.3 * weekly\_hours - 0.2 * education,$$

$$W_2 = 0.2 * seniority + 0.12 * age - 0.4 * weekly\_hours + 0.88 * education,$$

$$W_3 = -0.08 * seniority - 0.1 * age + 0.86 * weekly\_hours + 0.43 * education.$$

כעת נשווה את התפלגות W_1, W_2, W_3 במדגם מורים בשנת 1996 עם התפלגויותיהם במדגם מורים בשנת 1997 על ידי בדיקת מערכת תת ההשערות:

$$H_{01} = P(W_1(1996)) = P(W_1(1997))$$

$$H_{02} = P(W_2(1996)) = P(W_2(1997))$$

$$H_{03} = P(W_3(1996)) = P(W_3(1997))$$

באמצעות קבלת $p_{H_{01}}, p_{H_{02}}, p_{H_{03}}$ עבור המבחנים הבאים: Wilcoxon Two-Sample Test, **Kolmogorov-Smirnov Two-Sample Test (Asymptotic)** לכל אחת מתת ההשערות הללו נבדוק את ההשערה הכללית על ידי מציאת $"p - value"$ לפי **(VIII-2.6)** (ראה טבלה 5).

---

[3] מספר הגורמים נבחר על ידי קריטריון, לפיו תרומתו של כל גורם בקומבינציות הליניאריות לא פחותה מ- 15%, ובלבד שהתרומה המצטברת של סך הגורמים שנתקבלו אינה פחותה מ-75%.

טבלה 5: תוצאות בדיקת השערת שוויון התפלגויות משותפות של מאפייני מפתח (p-values)

חטיבה עליונה – מחוז ירושלים – מעבר מ-1996 ל-1997

מבחנים סטטיסטיים	H_{01}	H_{02}	H_{03}	השערה כללית
Wilcoxon 2-Sample Test	0.2355	0.3811	0.1976	0.483
Kolmogorov-Smirnov 2-Sample Test (Asymptotic)	0.0457	0.552	0.0933	0.13

כפי שניתן לראות מתוך טבלה 5 ה- p-value לדחיית השערה כללית לפי מבחן **Wilcoxon** הינו 0.483 ולפי **Kolmogorov-Smirnov** הינו 0.13, לכן לא דחינו את ההשערה הכללית, ולפיכך לא המלצנו על גריעת תת קובץ של מורי חטיבות עליונות במחוז ירושלים בשנת 1997. נציין כי השתמשנו במבחנים אי פרמטריים בלבד, משום שהבדיקה הראשונית על קיום נורמליות ההתפלגות (למשל עבור גורם W_1) נדחתה (ראה טבלה 6).

טבלה 6: מבחנים לנורמליות ההתפלגות של גורם 1 מתוך מאפייני מפתח של מורים

(חטיבה עליונה – מחוז ירושלים – מעבר מ-1996 ל-1997)

Test for Normality		
Test Statistic	**Value**	**p-value**
Kolmogorov- Smirnov	0.033	< .01
Cramer - von Mises	2.09	< .005
Anderson -Darling	16.92	< .005

באופן זהה, על כל התאים הבעייתיים בהם נמצא שיעור גידול שלילי מובהק נבדקה השערת שוויון התפלגויות משותפות של מאפייני המורים תוך שימוש בשיטת ניתוח מרכיבים ראשיים שתוארה לעיל ובאמצעות הפעלת סדרת מבחנים אי פרמטריים. נציין כי המלצנו לגרוע נתוני שנה מסוימת, רק אם לפי שני המבחנים נדחתה ההשערה של שוויון התפלגויות. בדוגמא של תא המתייחס לחינוך יסודי במחוז ירושלים במעבר מ-2000 ל-2001 (ראה טבלה 7) נראה כי ניתן לדחות את השערת שוויון התפלגויות משותפות לפי מבחן **Wilcoxon** ברמת מובהקות של $\alpha = 0.05$, אך לא ניתן לדחותה ברמת מובהקות זו לפי מבחן **Kolmogorov-Smirnov**, אי לכך ההשערה לא נדחתה. לעומת זאת, בדוגמא של תא המתייחס לחינוך יסודי במחוז דרום במעבר בין 1998 ל-1999 נדחתה השערה זו לפי שני המבחנים האי-פרמטריים (ראה טבלה 8), ולפי כך המלצנו על גריעת נתוני מורים בשנת 1999 בחינוך יסודי במחוז דרום. נציין כי פרט לתא זה, בשאר התאים שנמצאו בהם בשנים "בעייתיות" על פי מבחן **(VII-2.1)** לא נדחתה השערה שוויון התפלגויות משותפות של מאפייני המפתח של המורים, ולכן לא המלצנו עוד על גריעות נתוני מורים בשנים הללו.

טבלה 7: תוצאות בדיקת השערת שוויון התפלגויות משותפות של מאפייני מפתח (p-values)

(חינוך יסודי – מחוז ירושלים – מעבר מ-2000 ל-2001)

מבחנים סטטיסטיים	H_{01}	H_{02}	H_{03}	השערה כללית
Wilcoxon 2-Sample Test	0.8414	0.0034	0.174	0.0102
Kolmogorov-Smirnov 2-Sample Test (Asymptotic)	0.6444	0.0289	0.2859	0.0842

טבלה 8: תוצאות בדיקת השערת שוויון התפלגויות משותפות של מאפייני מפתח

(חינוך יסודי – מחוז דרום – מעבר מ-1998 ל-1999)

מבחנים סטטיסטיים	H_{01}	H_{02}	H_{03}	השערה כללית
Wilcoxon 2-Sample Test	0.0818	0.0002	0.1494	0.0006
Kolmogorov-Smirnov 2-Sample Test (Asymp.)	0.0439	0.0001	0.2198	0.0003

IX. התאמת מודלים לחיזוי מספר תלמידים

1. מבוא

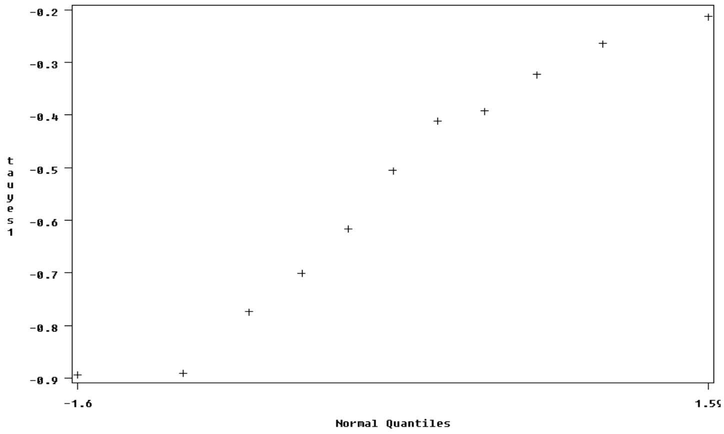
שלב זה של העבודה עוסק בבניית מודלים לחיזוי מספר התלמידים במערכת חינוך בישראל על סמך נתוני אוכלוסייה. משתנה מספר התלמידים נבחר כמשתנה חיצוני (מאקרו) על סמך ההנחה כי מספר התלמידים במערכת החינוך הינו גורם חיוני, בעל השפעה מרובה על הסיכויים של המורה להצטרף למערכת או לנשור ממנה. מתוך משתני המאקרו הפוטנציאליים השונים בחרנו במשתנה מספר התלמידים בלבד כדי לשלבו במודל לחיזוי מורים, וזאת עקב המגבלה של מספר מצומצם של שנים נצפות. תחזיות של מספר התלמידים נועדו לשילוב במודלים לחיזוי מספר המורים במערכת החינוך. המודלים לחיזוי מספר תלמידים הותאמו לכל תא (מחוז ושלב חינוך) והסתמכו על הנתונים והתחזיות של האוכלוסייה בישראל שנתקבלו מגף אוכלוסיה בלשכה המרכזית לסטטיסטיקה. עקב העובדה כי תחזיות כח אדם בהוראה והתלמידים מתייחסות למגזר העברי שאינו כולל תלמידים חרדים, השתמשנו בנתונים של אחוז התלמידים החרדים מתוך סך התלמידים על מנת לחזות את התלמידים הלא חרדים[4]. הבעיה הראשונה שהתעוררה בהתאמת המודל לתחזיות התלמידים היא תלות בין התצפיות, דבר שלא אפשר להתאים מודל ליניארי פשוט. יתרה מזאת, נתוני האוכלוסייה המתייחסות לטווח גילאים של התלמידים בכל שלב חינוך היו זמינים רק בין השנים 1995 ואילך. לאור המגבלות הללו התאמנו מראש מודל תחזית אחוז התלמידים החרדים בכל תא.

2. שלב 1 – חיזוי אחוז החרדים בקרב התלמידים

נגדיר משתנה $\tau_{jkl} = log\frac{t_{jkl}}{1-t_{ijkl}}$, כאשר $t_{jkl} = \frac{T^h_{jkl}}{T_{jkl}}$, T^h_{jkl} מספר התלמידים החרדים בשנה j בתא $\{kl\}$ ו- T_{jkl} הינו סך התלמידים בשנה j בתא $\{kl\}$. לפני התאמת המודל ליניארי ל- τ_{jkl} נבצע מבחנים לנורמליות עבור ההתפלגות של τ_{jkl} - ולהלן מספר דוגמאות עבור תאים נבחרים.

[4] פרט לתא המתייחס לחטיבה עליונה במחוז ירושלים, בו נתוני המורים כללו בין היתר גם מורים השייכים למגזר החרדי, ולכן במודל לחיזוי מספר המורים בתא זה השתמשנו במספר התלמידים החרדים (ראה טבלה 19 בהמשך).

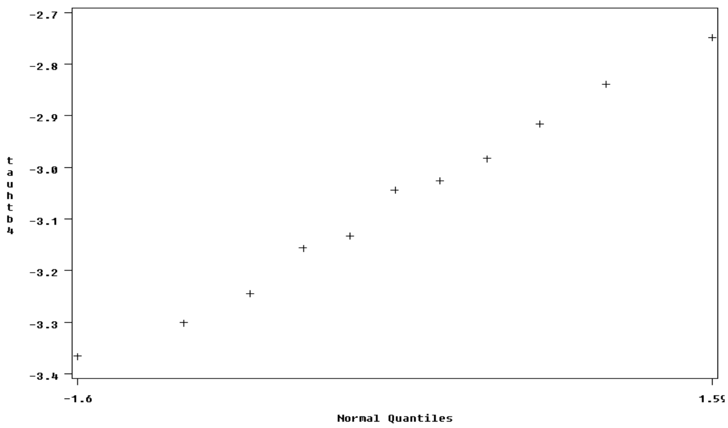
תרשים 15: QQ-plot – τ_{jkl} שלב חינוך יסודי – מחוז ירושלים



טבלה 9: מבחנים לנורמליות – τ_{jkl} שלב חינוך יסודי – מחוז ירושלים

Tests for Normality		
Test Statistic	Value	p-value
Shapiro-Wilk	0.93016	0.4124
Kolmogorov-Smirnov	0.15956	>.1500
Cramer-von Mises	0.04146	>.2500
Anderson-Darling	0.28707	>.2500

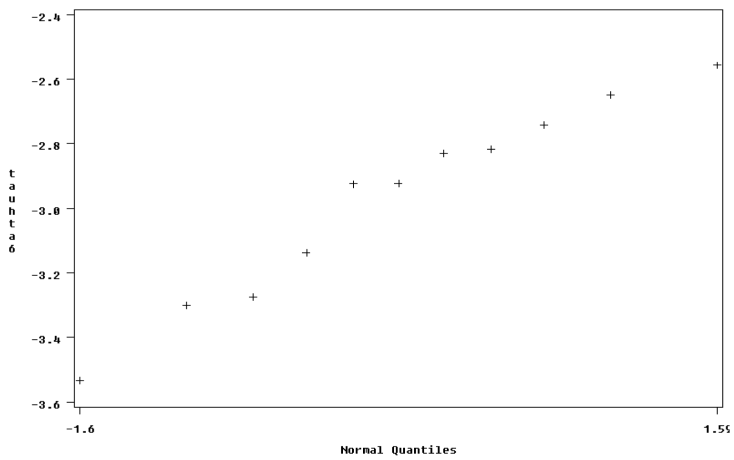
תרשים 16: QQ-plot τ_{jkl} - שלב חינוך חט"ב – מחוז מרכז



טבלה 10: מבחנים לנורמליות - τ שלב חינוך חט"ב – מחוז מרכז

Tests for Normality		
Test Statistic	Value	p-value
Shapiro-Wilk	0.98164	0.9746
Kolmogorov-Smirnov	0.0972	>.1500
Cramer-von Mises	0.01533	>.2500
Anderson-Darling	0.11657	>.2500

תרשים 17 QQ-plot τ_{jkl} - שלב חינוך חטיבה עליונה – מחוז דרום



טבלה 11 : מבחנים לנורמליות – τ_{jkl} חטיבה עליונה – מחוז דרום

Tests for Normality		
Test Statistic	***Value***	***p-value***
Shapiro-Wilk	***0.95132***	***0.6608***
Kolmogorov-Smirnov	***0.19799***	***>.1500***
Cramer-von Mises	***0.05017***	***>.2500***
Anderson-Darling	***0.28162***	***>.2500***

מתוך תרשימים 17-15 וטבלאות 11-9 ניתן לראות כי על פי המבחנים האי פרמטריים אין אנו דוחים את השערה כי τ_{jkl} מתפלג נורמלית.

כעת הותאמו מודלים של רגרסיה ליניארית ל τ_{jkl} כפונקציה של שנה לכל תא $\{kl\}$

$$\varepsilon_{jkl} \sim N(0,\sigma_{kl}) \ ,\tau_{jkl} = \beta_0^{kl} + \beta_1^{kl} j + \varepsilon_{jkl} \qquad \textbf{(IX-2.1)}$$

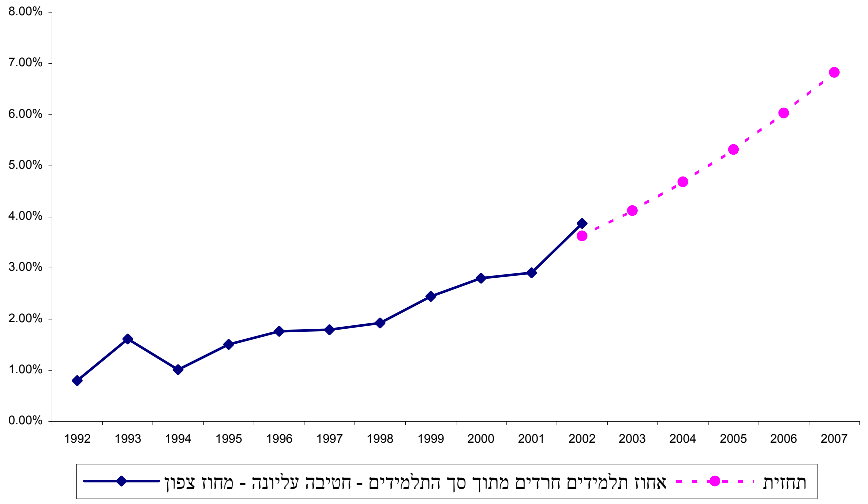
כאשר j - היא השנה המתייחסת ליחס נצפה של תלמידים חרדיים, $j = 1992,\mathrm{K}\ ,2002$.

על סמך המקדמים שנתקבלו במודל (ראה טבלה 12) ניתנו תחזיות לאחוז התלמידים החרדים $\hat{\tau}_{jkl}$, $j = 2003,\mathrm{K}\ ,2007$ (ראה דוגמאות בתרשימים 20-18).

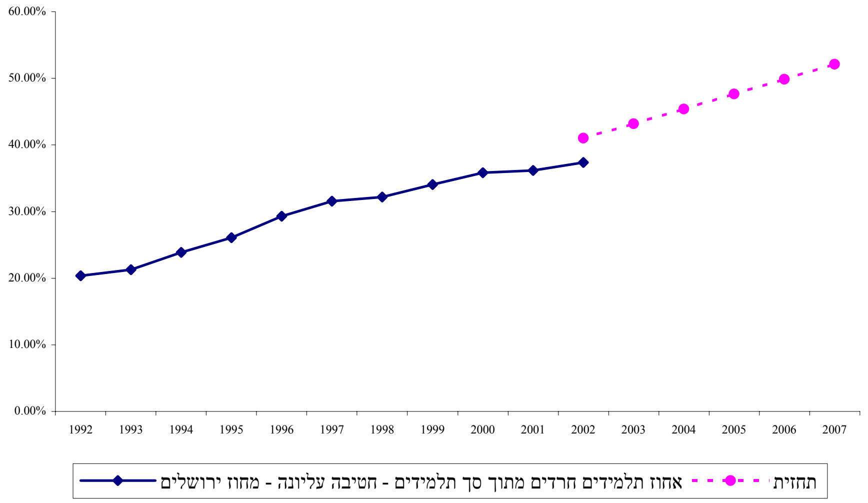
טבלה 12: תוצאות רגרסיה ליניארית בתאים נבחרים (כל המקדמים מובהקים p-value<.05)

תא	intercept	Coeff. of year	תא	intercept	Coeff. of year
חינוך יסודי–מחוז ירושלים	-147.334	0.074	חטיבת ביניים– מחוז מרכז	-118.357	0.058
חינוך יסודי–מחוז צפון	-201.999	0.100	חטיבת ביניים– מחוז תל אביב	-107.074	0.053
חינוך יסודי–מחוז חיפה	-140.893	0.069	חטיבת ביניים– מחוז דרום	-143.022	0.070
חינוך יסודי–מחוז מרכז	-119.942	0.059	חטיבה עליונה– מחוז ירושלים	-179.542	0.090
חינוך יסודי–מחוז תל אביב	-114.441	0.057	חטיבה עליונה– מחוז צפון	-269.946	0.133
מחוז דרום-חינוך יסודי	-181.686	0.090	חטיבה עליונה– מחוז חיפה	-77.353	0.037
חטיבת ביניים–מחוז ירושלים	-151.800	0.076	חטיבה עליונה– מחוז מרכז	-144.680	0.071
חטיבת ביניים–מחוז צפון	-180.301	0.089	חטיבה עליונה–מחוז תל אביב	-163.581	0.081
חטיבת ביניים–מחוז חיפה	-34.748	0.016	מחוז דרום -חטיבה עליונה	-179.565	0.088

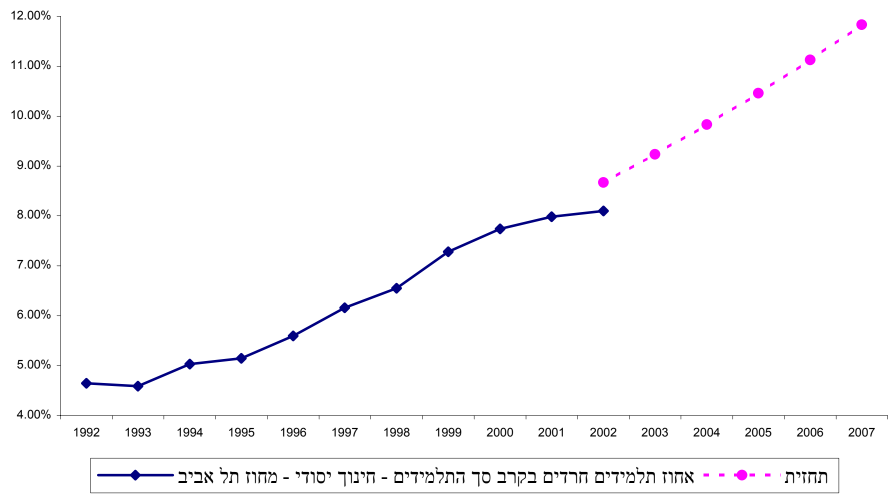
תרשים 18 : תחזית אחוז התלמידים החרדים מתוך סך התלמידים : חטיבה עליונה - מחוז צפון



תרשים 19: : תחזית אחוז התלמידים החרדים מתוך סך התלמידים – חטיבה עליונה - מחוז ירושלים



תרשים 20: : תחזית אחוז התלמידים החרדים מתוך סך התלמידים –חינוך יסודי - מחוז תל אביב



3. שלב 2 – התאמת מודל לחיזוי סך התלמידים

בשלב זה נעשתה התאמת מודל לחיזוי סך התלמידים בכל תא $\{kl\}$ כפונקציה של סך האוכלוסייה בטווח גילאים מתאים (נתוני האוכלוסייה קיימים רק משנת 1995).

יהי S_{jkl} סך התלמידים (כולל חרדים) בתא $\{kl\}$ בשנה j, אזי ניתן לכתוב את שיעור גידול הסטודנטים משנה $j-1$ לשנה j באופן הבא:

$$\eta_{jkl} = \frac{S_{jkl} - S_{(j-1)kl}}{S_{jkl}} \qquad \textbf{(IX-3.1)}$$

יהי P_{jkl} הינו גודל האוכלוסייה (כולל חרדים) בטווח הגילאים המתאים לתא $\{kl\}$ בשנה j, אזי, באופן דומה, שיעור גידול אוכלוסיית הגיל המתאים משנה $j-1$ לשנה j יתקבל כ-

$$\lambda_{jkl} = \frac{P_{jkl} - P_{(j-1)kl}}{P_{(j-1)kl}} \qquad \textbf{(IX-3.2)}$$

לפני התאמת מודל רגרסיה ליניארית נעשתה בדיקת נורמליות ל- η_{jkl}. תוצאות בדיקה נבחרות מובאות בטבלאות 13-15.

טבלה 13: מבחני נורמליות – שיעור גידול מספר התלמידים (η) - חינוך יסודי – מחוז דרום (yes6)

Tests for Normality		
Test Statistic	***Value***	***p-value***
Shapiro-Wilk	***0.915361***	***0.4342***
Kolmogorov-Smirnov	***0.22268***	***>.1500***
Cramer-von Mises	***0.051596***	***>.2500***
Anderson-Darling	***0.317826***	***>.2500***

טבלה 14: מבחני נורמליות – שיעור גידול מספר התלמידים (η) - חטיבת ביניים – מחוז מרכז (htb4)

Tests for Normality		
Test Statistic	***Value***	***p-value***
Shapiro-Wilk	***0.88908***	***0.2699***
Kolmogorov-Smirnov	***0.242557***	***>.1500***
Cramer-von Mises	***0.078824***	***0.1885***
Anderson-Darling	***0.434698***	***0.2157***

טבלה 15: מבחני נורמליות – שיעור גידול מספר התלמידים (η) - חטיבת ביניים – מחוז מרכז (hta1)

Tests for Normality		
Test Statistic	***Value***	***p-value***
Shapiro-Wilk	***0.845568***	***0.1119***
Kolmogorov-Smirnov	***0.22974***	***>.1500***
Cramer-von Mises	***0.091089***	***0.1246***
Anderson-Darling	***0.522172***	***0.1206***

כפי שניתן לראות בטבלאות 15-13, המבחנים האי פרמטריים מצביעים על כך כי השערת הנורמליות אינה נדחית. כעת, לאחר בדיקת הנורמליות נעשתה התאמת מודלים של רגרסיה ליניארית עבור η_{jkl} באופן הבא:

$$\tilde{\varepsilon}_{jkl} \sim N(0,\tilde{\sigma}_{kl}) \ ,\eta_{jkl} = \tilde{\beta}_0^{kl} + \tilde{\beta}_1^{kl}\lambda_{jkl} + \tilde{\varepsilon}_{jkl} \qquad \textbf{IX-5.1}$$

תהליך התאמת המודל נעשה בהתחשב ב-p-values שנתקבלו בכל מודל. במקרים בהם חותך המודל נתקבל כלא מובהק, ומקדם משתנה המסביר נתקבל כמובהק, הותאם מודל ללא חותך. במקרים בהם שני המקדמים נתקבלו כבלתי מובהקים נבחר מודל עם חותך בלבד.

להלן רשימת מקדמים כפי נתקבלו כתוצאה מהרצת המודלים **(IX-5.1)**, (ראה טבלה 16).

טבלה 16 : טבלת מקדמי רגרסיה – מודלים לשיעור גידול תלמידים

	יסודי י-ם	***יסודי צפון***	***יסודי חיפה***	***יסודי מרכז***	***יסודי ת"א***	***יסודי דרום***	***חט"ב י-ם***	***חט"ב צפון***	***חט"ב חיפה***
intercept	***0.0370***	***-0.0074***	***-0.0084***	***0.0101***	***-0.0193***	***0.0088***	***---***	***0.0096***	***0.0032***
population coefficient	***---***	***---***	***---***	***---***	***---***	***---***	***1.3076***	***---***	***---***
	חט"ב מרכז	***חט"ב ת"א***	***חט"ב דרום***	***חט"ע י-ם***	***חט"ע צפון***	***חט"ע חיפה***	***חט"ע מרכז***	***חט"ע ת"א***	***חט"ע דרום***
intercept	***---***	***0.0048***	***0.0238***	***---***	***0.0374***	***0.0112***	***0.0320***	***0.0029***	***0.0438***
population coefficient	***1.1499***	***---***	***---***	***0.0547***	***---***	***---***	***---***	***---***	***---***

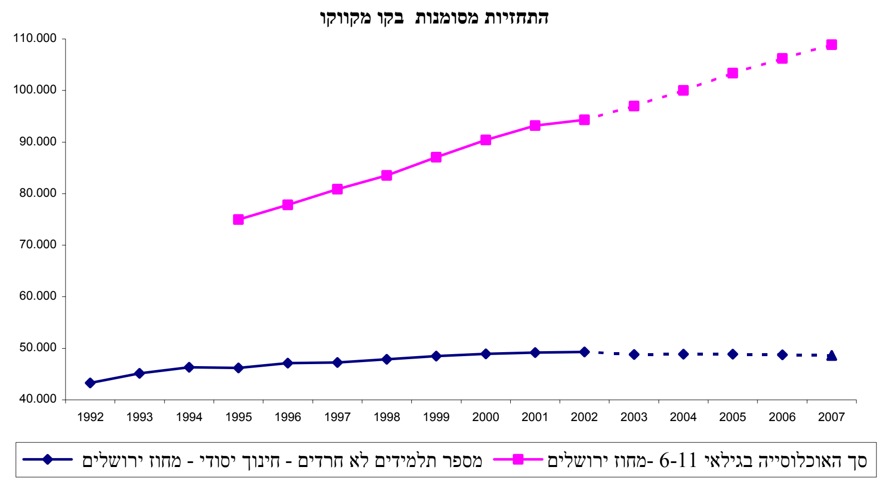
4. שלב 3 - חיזוי מספר תלמידים לא חרדיים

כעת, כשיש בידנו תחזיות של יחס תלמידים חרדים לעומת סך התלמידים, וכמו כן תחזיות סך התלמידים כפונקציה של סך האוכלוסייה בטווח גילאים מתאים, נותר לנו לבצע תחזית תלמידים לא חרדים $\hat{S}_{jkl}^{nh}$ בתא $\{kl\}$ לשנה j , $j = 2003,\ldots 2007$, באופן הבא:

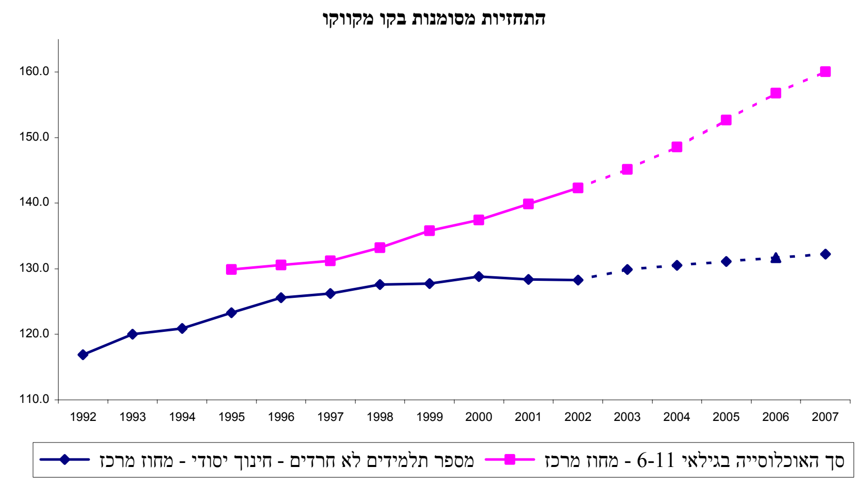
$$\hat{S}_{jkl}^{nh} = \hat{S}_{jkl}\left(1-\hat{t}_{jkl}\right) \qquad \textbf{(IX-4.1)}$$

בתרשימים 27-21 מובאות דוגמאות נבחרות מתחזיות מספר תלמידים לא חרדיים מול תחזיות אוכלוסייה בגילאים ובמחוז גיאוגרפי מתאימים.

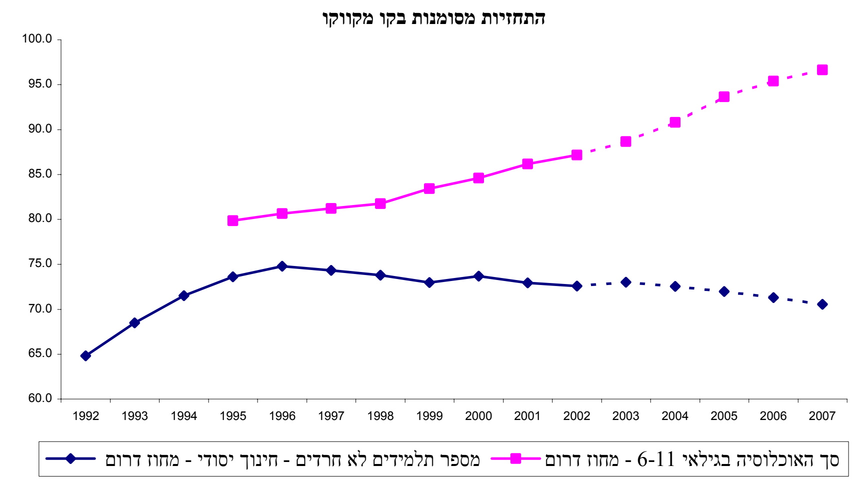
תרשים 21: תחזיות תלמידים לא חרדים מול סך האוכלוסייה בגילאי 6-11 : חינוך יסודי- מחוז ירושלים



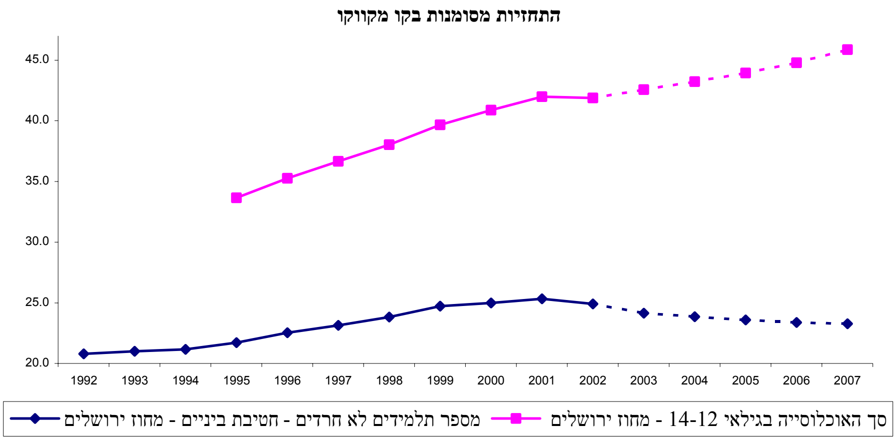
תרשים 22: תחזיות תלמידים לא חרדים מול סך האוכלוסייה בגילאי 6-11 : חינוך יסודי –מחוז מרכז



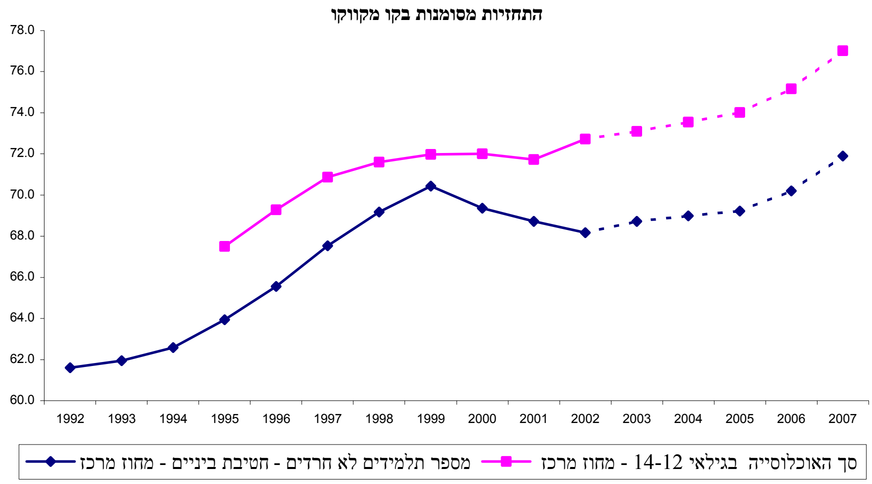
תרשים 23: : תחזיות תלמידים לא חרדים מול סך האוכלוסייה בגילאי 6-11 : חינוך יסודי –מחוז דרום



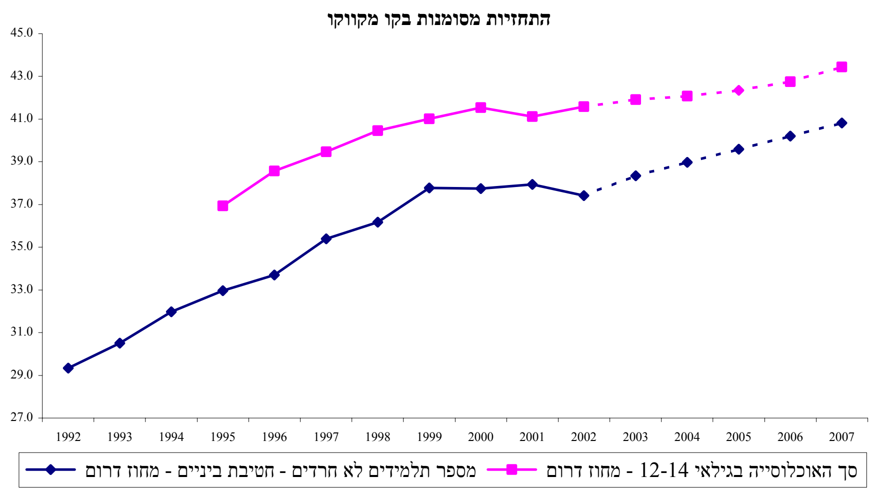
תרשים 24: : תחזיות תלמידים לא חרדים מול סך האוכלוסייה בגילאי 12-14: חטיבת ביניים –מחוז ירושלים



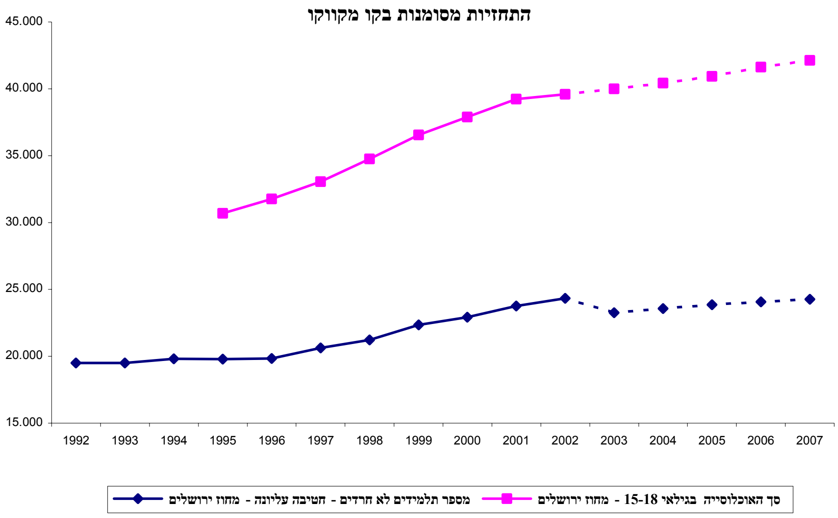
תרשים 25: : תחזיות תלמידים לא חרדים מול סך האוכלוסייה בגילאי 12-14: חטיבת ביניים –מחוז מרכז



תרשים 26: : תחזיות תלמידים לא חרדים מול סך האוכלוסייה בגילאי 12-14 : חטיבות ביניים –מחוז דרום



תרשים 27: תחזית תלמידים לא חרדים מול סך האוכלוסייה בגילאי 15-18 : חטיבה עליונה – מחוז ירושלים



X. חיזוי מספר מורים במערכת החינוך בישראל

1. בניית מודל לוגיסטי תוך שימוש בהסתברויות מעבר של המורים

רוב המודלים לחיזוי מגמות בכח אדם בהוראה מבוססים על ניתוח סדרות עתיות של נתונים אגרגטיביים. מודלים אלה, כידוע, מסתמכים על אוסף של נתונים היסטוריים אמינים המתייחסים לתקופה ארוכה יחסית. מאחר ובישראל קיים מחסור בנתונים היסטוריים ולרשות החוקר עומדות אחד עשר תצפיות בלבד, מוצעת בזאת גישה חלופית לטיפול בבעיית החיזוי של היקף כח אדם בהוראה בעזרת רגרסיה לוגיסטית.

לפנינו עומדת בעיה של חיזוי התנהגות קבוצה חברתית, כאשר ידועים מאפייניה האישיים והמקצועיים. אם נשתמש במידע זה וננתח הסתברויות מעבר של פרט כפונקציה של משתנים מסבירים נוכל לדייק בתהליך החיזוי גם במצב של כמות מוגבלת של נתונים אגרגטיביים.

נגדיר a_i בתור מורה i , A_{jkl} כקבוצת מורים בשנה j בשלב חינוך k,

$k \in$ {יסודי, חטיבות ביניים, חטיבות עליונות} , במחוז l,

$l \in$ {ירושלים, צפון, חיפה, מרכז, תל אביב, דרום}.

יהי N_{jkl} מספר המורים בקבוצה A_{jkl} . כעת ניצור קבוצה מאוחדת $B_{jkl} = \{A_{(j-1)kl} \cup A_{jkl}\}$ בה יהיו מורים השייכים למערכת בשנה j ומורים השייכים אליה בשנה $j-1$. לכל המורים ב- B_{kjl} נגדיר שני וקטורים של המאפיינים $\{X_{i(j-1)kl,} X_{ijkl}\}$, כאשר $X_{ijkl} \in$ {גיל, ותק, בסיס משרה, שעות הוראה וכד'}

כמו כן נניח כי ההתפלגות השולית של וקטור X_{ijkl} תהא $\widetilde{F}_{kjl}$.

כעת נגדיר משתנה מקרי (כניסה/נשירה) באופן הבא:

$$Y_{ik(j-1)l0} = \begin{cases} 1, & a_i \in A_{k(j-1)l} \ \textbf{וגם} \ a_i \notin A_{kjl} \\ 0, & a_i \in A_{k(j-1)l} \ \textbf{וגם} \ a_i \in A_{kjl} \end{cases} \qquad \textbf{(X-1.1)}$$

$$Y_{ijkl} = \begin{cases} 1, & a_i \notin A_{k(j-1)l} \ \textbf{וגם} \ a_i \notin A_{kjl} \\ 0, & a_i \in A_{k(j-1)l} \ \textbf{וגם} \ a_i \in A_{kjl} \end{cases} \qquad \textbf{(X-1.2)}$$

נניח, כי (X,Y) תצפיות שוות התפלגות המקיימות את מודל רגרסיה לוגיסטית, כך שלכל מורה $a_i \in B_{kjl}$, ולכל $m = 0,1$, המציינים כניסה או נשירה בהתאמה יתקיים:

$$p_{i(j-1+m)klm} = P\left\{Y_{i(j-1+m)klm} = 1 \middle| X_{i(j-1+m)kl}\right\} = \left(1 + exp\left(-X^T_{i(j-1+m)kl}\beta_{(j-1+m)klm}\right)\right)^{-1}, \qquad \textbf{(X-1.3)}$$

כאשר $\beta_{(j-1+m)klm}$ מציין את וקטור הפרמטרים הבלתי ידועים של המודל הלוגיסטי.

אם $m = 0$, אזי הביטוי **(X-1.3)** שווה להסתברות שהמורה ינשור מהמערכת בשנה j, אם $m = 1$, אזי ביטוי זה שווה להסתברות עבור מורה להיכנס לראשונה למערכת בשנה j, כאשר ידוע כי בשנה $j-1$ הוא לא השתייך אליה.

נסמן את E_F בתור תוחלת תחת פונקצית התפלגות F.

$$E_F g(\xi) = \int g(z)dF(z) \qquad \textbf{(X-1.4)}$$

מספר המורים הכולל מחושב מתוך:

$$N_{jkl} = \left(1 - p_{(j-1)kl0}\right)N_{(j-1)k} + p_{kjl1}N_{kjl} \qquad \textbf{(X-1.5)}$$

כאשר עבור $m = 0,1$ $\quad p_{(j-1+m)klm} = E_{\tilde{F}_{(j-1+m)kl}} p_{i(j-1+m)klm},$

כתוצאה מכך קיבלנו נוסחה רקורסיבית

$$N_{jkl} = \frac{1 - p_{(j-1)kl0}}{1 - p_{jkl1}} N_{(j-1)kl} \qquad \textbf{(X-1.6)}$$

כעת, אם B_{kjl} ידוע ונניח הומוגניות בזמן אזי התחזית למספר המורים בשנה $j+1$

$$N_{(j+1)kl} = \frac{1 - E_{\tilde{F}_{jkl}} p_{(j-1)k0}}{1 - E_{F_{jk}} p_{jkl1}} N_{jk} \qquad \textbf{(X-1.7)}$$

אם כן, גישה הסתברותית זו מאפשרת לנו לקבל רמת מובהקות מספקת של המודל, גם כאשר מספר השנים בהן קיימות תצפיות הוא מצומצם.

2. התאמת מודלים לוגיסטיים לחיזוי מורים

תהליך התאמת מודלים לוגיסטיים להסתברויות מעבר של המורים כולל מספר שלבים:

i. בחירת משתנים ואינטראקציות רלוונטיות למודל

כל הנתונים שהופיעו בקבצי מורים והיה להם קשר ישיר או עקיף באפיון אישי או מקצועי של המורה שימשו כבסיס לפיתוח המודלים הסופיים. כמו כן הוגדרו משתני עזר וחלק מהמשתנים המקוריים עברו תהליך התאמה וקטגוריזציה למספר רמות סביר הניתן לניתוח בקלות יתר. טבלה 17 מתארת את המשתנים מסבירים פוטנציאליים במודל הלוגיסטי.

לצורך שילוב משתני אינטראקציה רלוונטיים במודל חושבו מקדמי מתאם של כל המשתנים עבור כל קובץ הנתונים הרב שנתיים בכל מחוז ושלב חינוך. מתוך כל המשתנים נבחרו זוגות בעלי מקדם המתאם ביניהם מובהק וגדול בערכו המוחלט מ-0.2 ובהתאם הוגדרו משתני אינטראקציה (ראה טבלה 18).

טבלה 17: רשימת משתנים מתוך מאפייני מורים רלוונטיים (כולל משתנה מספר התלמידים)

תאור המשתנה	*שם המשתנה בקובץ*
חלקיות משרה של מורה ביחס לשעות הוראה	***BasMis***
מספר הילדים במשפחה	***Child***
דרגת שכר של המורה	***Darga***
האם המורה מלמד ביותר משלב חינוך אחד בו זמנית	***Kaful1***
גיל המורה	***Gil***
מצב משפחתי	***Mazav***
מס' בי"ס בהם מלמד המורה	***Misra***
מוצא אב של המורה	***MozaAv***
משתנה עזר נשוי/אחר	***Nasui***
מורה בעל ותק 1	***New***
משתנה עזר תפקיד ניהולי/אחר	***Nihul***
סוג פיקוח של ביה"ס (ממלכתי , אחר)	***Pikuah***
מין המורה	***Sex***
סך שעות הוראה של המורה	***Shaot***
תפקיד המורה (מרכז, מורה בכיר וכד')	***Tafkd***
מספר התלמידים בשנה מסוימת (משתנה חיצוני)	***Tal1***
ותק המורה בשנים	***Vetek***
שנה אליה מתייחסים הנתונים על המורה	***Year***

בנוסף, עבור המשתנים הרציפים הבאים: גיל, דרגה, ותק ושעות הוראה של המורה הוגדרו ריבועי משתנים אלה והוכנסו כמשתנים פוטנציאליים במודלים לחיזוי מספר המורים, וזאת מתוך כוונה לבדוק תלות אפשרית המבוססת על קשר ריבועי.

טבלה 18: משתני אינטראקציות שנבחרו לצורך ניתוח

האינטראקציה	שם המשתנה	האינטראקציה	שם המשתנה
*Pikuah*kaful1*	*pk*	*Shaot*Gil*	*shg*
*Sex*Gil*	*sg*	*Shaot*Nihul*	*shn*
*Nihul*Shaot*	*ns*	*Shaot*Tafkd*	*sht*
*kaful1*Shaot*	*ks*	*BasMis*Kaful1*	*bk*
*Gil*Child*	*gich*	*Basmis*New*	*bnew*
*Vetek*Gil*	*vg*	*Basmis*Sex*	*bs*
*Year*Tal1*	*yt1*	*Basmis*Nihul*	*bn*
*Year*Tal1*	*yt1m*	*Basmis*Tafkd*	*bt*
*Vetek*Year*	*yvm*	*Basmis*Vetek*	*bv*
*Gil*Year*	*ygm*	*Basmis*Gil*	*bg*
*Sex*Year*	*ysm*	*Darga*Gil*	*dag*
*Year*Shaot*	*yshm*	*Darga*Vetek*	*dv*
*Gil*Tal1*	*tgm*	*Mozaav*Gil*	*mozg*
*Sex*Tal1*	*tsm*	*New*Gil*	*neg*
*Tal1*Shaot*	*tshm*	*Nasui*Child*	*nach*
*Tafkd*Tal1*	*tTm*	*Pikuah*Child*	*pch*
*Vetek*Tal1*	*tVm*	*Pikuah*Darga*	*pd*

ii. התאמת מודלים לוגיסטיים לחיזוי מורים

בשלב זה הורצה פרוצדורה של רגרסיה לוגיסטית בשיטת stepwise על כל המשתנים הרלוונטיים מתוך טבלאות 17- 18 כולל משתנה מספר התלמידים והאינטראקציות הרלוונטיות בינו לבין שאר המשתנים. כעת, עבור כל תא $\{kl\}$ הורצו שתי פרוצדורות של רגרסיה לוגיסטית בשיטת stepwise , כאשר משתנה שנת התצפיות נכלל בוקטור מאפיינים של כל מורה. רגרסיה לוגיסטית אחת הורצה למציאת קשר בין המאפיינים לבין ההסתברות לנשור מהמערכת, והשניה למציאת קשר זה עם ההסתברות להיכנס למערכת. בסיום הרצת הרגרסיות נתקבלו וקטורי המקדמים $\hat{\beta}_{klm}, m = 0,1$, מובהקים ברמה של $\alpha = 0.05$, עבור וקטור המאפיינים X^T_{ikl}, כאשר $j \in X^T_{ikl}$, המסבירים את ההסתברויות המעבר p_{iklm} , כך ש:

$$p_{iklm} = \left(1 + exp\left(-X^T_{ikl}\beta_{klm}\right)\right)^{-1}, \qquad \textbf{(X-2.1)}$$

iii. התאמת מודלים ללא משתנה התלמידים (וללא האינטראקציות בינו לבין שאר המשתנים)

מחשש כי הסתברויות מעבר של המורים שנאמדו בעזרת המודל הלוגיסטי יהיו רגישים במידה רבה למספר התלמידים, התאמנו מודלים נוספים של רגרסיה לוגיסטית, הפעם ללא משתנה התלמידים וללא האינטראקציות בינו לבין שאר המשתנים, וזאת על מנת לבדוק את סבירות תחזיות מספר המורים המבוססות על תחזיות מספר התלמידים. מטרת תהליך זה הייתה לבדוק כיצד גריעת משתנה התלמידים

תשפיע על תחזיות המורים בהשוואה לאלה שיתקבלו במודלים המלאים הכוללים את משתנה התלמידים. תמונה השוואתית זו באה לבחון את התרומה של משתנה התלמידים למודל המורים וכמו כן לספק מדד נוסף למהימנות התחזיות שהתקבלו.

iv. בחינת חלופות חיזוי שונות וקביעת מערכת קריטריונים לאימוץ החלופה העדיפה

לאחר התאמת המודלים הלוגיסטיים עבור כל תא התעוררה סוגיה באשר לאפשרויות החיזוי של מספר המורים לשנים 2002 ואילך. הסוגייה הראשונה קשורה להשפעת זקיפות הנתונים על המודלים הלוגיסטיים. כתוצאה מתהליך זקיפת הנתונים (ראה תת פרק III-2) נוצרו שני סוגי קבצי נתונים של המורים : הקובץ המקורי עם ערכים חסרים וקובץ לאחר ביצוע זקיפות. יצוין, כי הקבצים בעלי נתונים חסרים מכילים מספר קטן יותר של רשומות שלמות הנכללות בחישוב בעת התאמת המודל, דבר העלול לגרום להפסד מידע חיוני אודות המורים. מאידך, אין אנו יכולים באופן מיידי להעריך את השפעת הזקיפות על התפלגות המאפיינים וכתוצאה מכך על הפרמטרים של המודל שיותאם בעקבות פעולה זו. אי לכך בחרנו לבצע את התחזיות על סמך הקבצים שעברו תהליך זקיפה על מנת לא להפסיד רשומות מורים. הסוגייה שנותרה היא האם להריץ פרוצדורת רגרסיה לוגיסטית על בסיס הקובץ המקורי או לחילופין להפיק את המודלים הלוגיסטיים על בסיס הקובץ שעבר תהליך זקיפות. הסוגייה השניה שהתעוררה היא האם לאחר הפקת המודל הלוגיסטי לבצע חיזוי על סמך רשומות מורים בשנת 2001 או לחילופין על סמך הרשומות לאורך כל השנים הקיימים בקבצים (1991-2001).

כלומר אם: A_{jkl} היא קבוצת מורים בשנה j , $j = 1991,\ldots,2001$ אזי $B_{\cdot kl} = \left\{ \bigcup_{j=1991}^{2001} A_{jkl} \right\}$

שתי האפשרויות לחזות את הסתברויות המעבר ל-2002 הינן:

$$E_F\, g(\xi) = \int_{A_{2001kl}} g(z)\,dF(z) \qquad \textbf{(X-2.1)}$$

או

$$E'_F\, g(\xi) = \int_{B_{\cdot kl}} g(z)\,dF(z) \qquad \textbf{(X-2.2)}$$

אם נסכם את כל אפשרויות החיזוי, נקבל שמונה קומבינציות אפשריות $\left(2^3\right)$ לחזות את מספר המורים בשנת 2002 שכללו (ראה טבלה 19):

1. חיזוי על סמך המודלים שנתקבלו מהרצה על הקובץ המקורי (כולל ערכים חסרים) או על סמך המודלים שנתקבלו מהרצה על הקובץ שעבר תהליך זקיפות.
2. חיזוי על סמך רשומות מורים בשנת 2001 בלבד או על סמך הרשומות לאורך כל התקופה 1991-2001.
3. חיזוי על סמך סוג המודל עם השפעה של התלמידים (המודל המלא) לעומת סוג המודל ללא השפעתם (המודל המצומצם), (ראה X-2.iii).

מאחר והועמדו לרשותנו נתונים על מספר המורים בכל מחוז ושלב חינוך לשנת 2002 שלא השתתפו בבניית המודלים בדקנו את דיוק החיזוי באמצעות מדידת טעות יחסית (באחוזים) שנתקבלה בעת החיזוי לפי כל מודל. מידע זה שימש אותנו בקבלת החלטה לגבי שימוש במודל זה או אחר להמשך תחזיות לשנים 2003-2007. בשלב זה הוחלט להשוות את יכולת החיזוי של המודלים לשנת 2002 כאשר הם מופרדים לשתי קבוצות – המודלים המלאים והמודלים המצומצמים.

עבור כל תא $\{kl\}$, מבין ארבעת המודלים המלאים נבחר מודל s, $s=1,\ldots,4$ כמתאים לחיזוי ומבין ארבעת המודלים המצומצמים נבחר מודל r $r=1,\ldots,4$, כמתאים לחיזוי , באופן שוקטור הטעויות היחסיות מינימליות בחיזוי מספר המורים בשנת 2002 יהיה מהצורה:

$$\left(\underset{s}{Min}\left\{\frac{\left|\hat{N}_{2002kls}-N_{2002kl}\right|}{N_{2002kl}}100\%\right\},\underset{r}{Min}\left\{\frac{\left|\hat{N}_{2002klr}-N_{2002kl}\right|}{N_{2002kl}}100\%\right\}\right) \qquad \textbf{(X-2.1)}$$

3. חיזוי לשנים 2003-2007

מבחן (**X-2.1**) נבחר כאמור כקריטריון הראשי להמשך חיזוי מספר המורים לשנים 2003-2007. על מנת לבחון האם התחזיות מספר המורים לשנים מתקדמות משקפות באופן הגיוני את התמונה במערכת החינוך, נבחר קריטריון משני נוסף שעל-פיו נבחנו שני המודלים שנבחרו על פי קריטריון **(X-2.3)**. קריטריון זה הינו יחס תלמידים מורים לאורך תקופת החיזוי, הבודק את השתנות פרופרציית תלמידים-מורים בכל שלב חינוך ומחוז בשנות החיזוי לעומת זו שהסתמנה בשנים שנחקרו.

נסמן ב- ρ_{jkl} את יחס תלמידים מורים בשנה j , $j=1991,\ldots,2001$ בתא $\{kl\}$, ונסמן ב- $\hat{\rho}_{j'klt}$ את יחס תלמידים מורים בשנה j' , $j'=2002,\ldots,2007$ בתא $\{kl\}$ הנאמד מתוך מודל t,

$t=\begin{cases}1, & \textbf{מודל מלא}\\ 2, & \textbf{מודל מצומצם}\end{cases}$.

אזי הקריטריון לקבילות המודל הינו לכל j' , $j'=2002,\ldots,2007$ מתקיים

$$\hat{\rho}_{j'klt}\in\left(\bar{\rho}_{kl}-Z_{0.975}\hat{\sigma}_{\rho_{kl}},\bar{\rho}_{kl}+Z_{0.975}\hat{\sigma}_{\rho_{kl}}\right), \qquad \textbf{(X-3.1)}$$

כאשר $\bar{\rho}_{kl}=\frac{1}{11}\sum_{j=1991}^{2001}\rho_{jkl}$, $\hat{\sigma}^2_{\rho_{kl}}=\frac{1}{10}\sum_{j=1991}^{2001}\left(\rho_{jkl}-\bar{\rho}_{kl}\right)^2$.

תוצאות החיזוי לשנים 2002-2007 מופיעים בפרק XI נספח ב' המצורף לדו"ח. המודלים סווגו על פי המפתח המופיע בטבלה 19. כל תרשים המופיע בנספח כולל תחזיות על פי שני מודלים לחיזוי המורים: המודל הטוב ביותר מבין המלאים, והמודל הטוב ביותר מבין המצומצמים (ראה X-2-iv.3) על פי מבחן (**X-2.1**) ולפני הפעלת הקריטריון **(X-3.1)**.

טבלה 19: מפתח לסיווג מודלים שונים לחיזוי מספר המורים במערכת החינוך

סוג המודל	בסיס לחיזוי	בסיס לבניית המודל	סוג המודל
A	חיזוי על סמך נתוני 2001	מבוסס על קבצים מקוריים	כמודל מלא (כולל משתנה מספר התלמידים הלא חרדיים)[1]
B	חיזוי על סמך נתוני 1992-2001		
C	חיזוי על סמך נתוני 2001	מבוסס על קבצים לאחר זקיפות	
D	חיזוי על סמך נתוני 1992-2001		
E	חיזוי על סמך נתוני 2001	מבוסס על קבצים מקוריים	מודל מצומצם (ללא משתנה מספר התלמידים)
F	חיזוי על סמך נתוני 1992-2001		
G	חיזוי על סמך נתוני 2001	מבוסס על קבצים לאחר זקיפות	
H	חיזוי על סמך נתוני 1992-2001		

([1]פרט לחטיבה עליונה -מחוז ירושלים , כשאר נעשה שימוש במספר התלמידים הכולל חרדים)

XI. נספחים

נספח א' (הוכחת טענה 1 בפרק VIII)

נוכיח את טענה 1.

נשים לב כי, **(VIII-2.2)** הינו תנאי מספיק של Karlehman [Prohorov(1967), p.212] על מנת שהתפלגות הסתברויות P תוגדר אך ורק על ידי המומנטים של עצמה.

לצורך פשטות נניח ש- P מוגדר חד משמעית על ידי n מומנטים. (למשל נורמלית, $n=2$) ו- $N=2$.

השאלה המרכזית היא: אם ידועות התפלגויות של כל משתנה המקרי $W_1=\alpha_{11}X_1+\alpha_{12}X_2$ ו- $W_2=\alpha_{21}X_1+\alpha_{22}X_2$ בנפרד האם ניתן לשחזר את ההתפלגות המשותפת היחידה של המשתנים המקריים X_1,X_2.

התפלגות זו מוגדרת אך ורק על ידי המומנטים הבאים:

$$
\begin{aligned}
&EX_1, EX_2, E(X_1X_2), EX_1^{\ 2}, EX_2^{\ 2},\\
&EX_1^2X_2, E(X_1X_2^2), EX_1^3, EX_2^3, \ldots,\\
&E(X_1^{n-1}X_2), E(X_1^{n-2}X_2^2)\ldots, E(X_1X_2^{n-1}), E(X_1^n, EX_2^n)
\end{aligned}
$$

ידועה לנו המערכת של $2n$ משוואות ליניאריות הבאה:

$$
\begin{aligned}
\alpha_{11}EX_1+\alpha_{12}EX_2 &= c_1\\
\alpha_{21}EX_1+\alpha_{22}EX_2 &= c_2\\
\alpha_{11}^2EX_1^2+2\alpha_{11}\alpha_{12}EX_1X_2+\alpha_{12}^2EX_2^2 &= c_3\\
\alpha_{21}^2EX_1^2+2\alpha_{21}\alpha_{22}EX_1X_2+\alpha_{22}^2EX_2^2 &= c_4 \qquad \textbf{(XI-1.1)}\\
&\vdots\\
\sum_{i=0}^{n}C_n^iE\left[(\alpha_{11}X_1)^{n-i}(\alpha_{12}X_2)^i\right] &= c_{2n-1}\\
\sum_{i=0}^{n}C_n^iE\left[(\alpha_{21}X_1)^{n-i}(\alpha_{22}X_2)^i\right] &= c_{2n}
\end{aligned}
$$

כאשר $C_n^m = \dfrac{n!}{(n-m)!m!}$, c_i ידועים, $i = 1,\ldots,2n$.

נשים לב כי לא ידוע אם מתקיים

$$\begin{aligned} &EW_1^{\gamma}W_2^{\rho} = EW_1^{\gamma}EW_2^{\rho}, \\ &\gamma = 1,\ldots,N-1,\ p = 1,\ldots,N-\gamma, \end{aligned} \qquad \textbf{(XI-1.2)}$$

אזי במקרה הכללי ל-**XI-1.1** חסרות n משוואות שבהן מופיע $EW_1^{\gamma}W_2^{\rho}$, $\gamma = 1,\ldots,N-1,\ p = 1,\ldots,N-\gamma$ על מנת שיהיה לה פתרון יחיד. זה נובע מכך שההתפלגות המשותפת של W_1, W_2 אינה ידועה, ובמערכת של $2n$ משוואות עם $3n$ נעלמים אין פתרון כללי יחיד.

ברם אם **(XI-1.2)** מתקיים אזי משפחת המשוואות הנוספות מהצורה

$$\textbf{(XI-1.3)}\quad \gamma = 1,\ldots,N-1;\ p = 1,\ldots,N-\gamma,\ E\left[(\alpha_{11}X_1 + \alpha_{12}X_2)^{\gamma}(\alpha_{21}X_1 + \alpha_{22}X_2)^{\rho}\right] = c_1^{\gamma}c_2^{\rho}$$

מביאה לקיום פתרון כללי יחיד ל- **(XI-1.1)**

נגדיר סדרת מטריצות $\{A_n\}$ באופן הבא:

$$\textbf{(XI-1.4)}\quad A_1 = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix},\ A_2 = \begin{pmatrix} \alpha_{11}^2 & 2\alpha_{11}\alpha_{12} & \alpha_{12}^2 \\ \alpha_{21}^2 & 2\alpha_{21}\alpha_{22} & \alpha_{22}^2 \\ \alpha_{11}\alpha_{21} & \alpha_{11}\alpha_{22}+\alpha_{12}\alpha_{21} & \alpha_{12}\alpha_{22} \end{pmatrix},\ldots,$$

$$\ldots, A_n = \begin{pmatrix} \alpha_{11}^n & & \cdots & C_n^m\alpha_{11}^{n-m}\alpha_{12}^m & \cdots & & \alpha_{12}^n \\ \alpha_{21}^n & & \cdots & C_n^m\alpha_{21}^{n-m}\alpha_{22}^m & \cdots & & \alpha_{22}^n \\ \alpha_{11}^{n-1}\alpha_{21} & & \cdots & \vdots & \cdots & & \alpha_{12}^{n-1}\alpha_{22} \\ \vdots & & \cdots & & & & \vdots \\ \alpha_{11}^{n-m}\alpha_{21}^m & \alpha_{11}^{n-m}\sum\limits_{i=1}^{m} C_m^i\alpha_{21}^{m-i}\alpha_{22}^i & \cdots & C_{n-m}^s\alpha_{11}^{n-m-s}\alpha_{12}^s(\alpha_{21}+\alpha_{22})^m & \cdots & \alpha_{12}^{n-m}\sum\limits_{i=0}^{m-1} C_m^i\alpha_{21}^{m-i}\alpha_{22}^i & \alpha_{12}^{n-m}\alpha_{22}^m \\ \vdots & & & \vdots & & & \vdots \\ \alpha_{11}\alpha_{21}^{n-1} & & \cdots & \cdots & \cdots & & \alpha_{12}\alpha_{22}^{n-1} \end{pmatrix};$$

$$m = 1,\ldots,n,\ s = 1,\ldots,n-m$$

נסמן $\Delta = |A_1|$, נראה כי אם $\Delta \neq 0$, אז, $|A_n| \neq 0$.

נשים לב כי $|A_2| = \begin{vmatrix} \alpha_{11}^2 & 2\alpha_{11}\alpha_{12} & \alpha_{12}^2 \\ \alpha_{21}^2 & 2\alpha_{21}\alpha_{22} & \alpha_{22}^2 \\ \alpha_{11}\alpha_{21} & \alpha_{11}\alpha_{22}+\alpha_{12}\alpha_{21} & \alpha_{12}\alpha_{22} \end{vmatrix} = -\Delta^3$,

ובאופן דומה

$$|A_3| = \begin{vmatrix} \alpha_{11}^3 & 3\alpha_{11}^2\alpha_{12} & 3\alpha_{11}\alpha_{12}^2 & \alpha_{12}^3 \\ \alpha_{21}^3 & 3\alpha_{21}^2\alpha_{22} & 3\alpha_{21}\alpha_{22}^2 & \alpha_{22}^3 \\ \alpha_{11}^2\alpha_{21} & \alpha_{11}(2\alpha_{12}\alpha_{21}+\alpha_{11}\alpha_{22}) & \alpha_{12}(2\alpha_{11}\alpha_{22}+\alpha_{12}\alpha_{21}) & \alpha_{12}^2\alpha_{22} \\ \alpha_{11}\alpha_{21}^2 & \alpha_{21}(2\alpha_{11}\alpha_{21}+\alpha_{12}\alpha_{21}) & \alpha_{22}(2\alpha_{12}\alpha_{21}+\alpha_{11}\alpha_{22}) & \alpha_{12}\alpha_{22}^2 \end{vmatrix} = \Delta^6, \ldots$$

נקל לראות כי $\left|A_n\right|$ הינו פונקציה מעריכית מצורה $(-1)^{n+1}\Delta^{\frac{n(n+1)}{2}}$ ואם נמשיך באופן דומה את סדרת המטריצות נראה את נכונותה של טענה זו. (ראה Muir and Metzler(1960), p. 160, example 3).
אם כן, מתוך האמור לעיל נובע כי למערכת המשוואות

$$\begin{pmatrix} A_1 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 \\ 0 & 0 & A_3 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & A_n \end{pmatrix} d^T = \begin{pmatrix} c_1 \\ \vdots \\ \\ \vdots \\ c_{2n} \end{pmatrix}, \qquad \textbf{(XI-1.5)}$$

קיים פתרון יחיד עבור וקטור נעלמים d.
כאשר

$$d = \left[EX_1, EX_2, E(X_1X_2), \ldots, E(X_1^{n-1}X_2), E(X_1^{n-2}X_2^2), \ldots, E(X_1X_2^{n-1}), EX_1^n, EX_2^n\right]$$

אי לכך, בהינתן התפלגויות משתנים מקריים W_1, W_2 המקיימות את **(XI-1.2)**, ניתן לקבוע חד משמעית את התפלגויות המשתנים המקריים X_1, X_2. ברור כי הכללת האמור לעיל ל- N משתנים מקריים מביאה אותנו להוכחת טענה 1.

נספח ב' (מודלים ותרשימים)

XII. רשימה ביבליוגרפית

Box, G.E.P.(1949). A general distribution theory for a class of likelihood criteria, *Biometrika* **36** 317-346.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis and Forecasting.* Butterworths, London.

Croxton, F.E., Cowden, D.J., and Klein, S. (1968). *Applied General Statistics.* Pitman.

Harvey, Andrew C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

Hawkins D. M. (1974). The Detection of Errors in Multivariate Data Using Principle Components. *J. Amer. Statist. Assoc.* **69** 340-344.

Lauter, J., Glimm, E., Kropf, S. (1998). Multivariate tests based on linear scores. *Ann. Stat*. **26** 1972-1988.

Meier, P. (1975). Statistics and medical experimentation. *Biometrics* **31** 511-529.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40** 1079-1087.

Morisson, D. F. (1967). *Multivariate Statistical Methods*. McGraw-Hill.

Muir T., Metzler H.(1960) *A Treatise on the Theory of Determinants*. Dover. New York.

Prohorov Ur. (1967). *Theory of probabilities. Basic concepts, limit theorems, stochastic processes*. Nauka. Moscow. (In Russian)

Robinson, C.(1971). *Business Forecasting*. Thomas Nelson and Sons, London.

Schott, J. R. (1991). Some Tests for Common Principle Component Subspaces in Several Groups. *Biometrika* **78(4)** 771-777.

Seber, G. A. F.(1984) *Multivariate observations*. John Wiley & Sons, Inc.

Skinner, C. J., Holmes, D. J., and Smith, T. M. F.(1986). The Effect of Sample Design on Principle Component Analysis. *J. Amer. Statist. Assoc.* **81** 789-798.

Spencer, M.H., Clark, C. and Hoguet, P.W. (1961). *Business and Economic Forecasting in Econometric Approach.* Irwin, Homewood, Illinois.

Vexler, A.A. and Maagan, D. (2002). Forecasting the size of the teaching force in Israel using Stochastic Models. *The 22nd International Symposium on Forecasting, 2002*. Trinity College, Dublin, Ireland.