

מדינת ישראל
משרד ראש הממשלה



**מודלים סטטיסטיים לחיזוי מספר התלמידים בכיתות א'
במערכת החינוך בישראל
דו"ח מתודולוגי**

אלון שפירא
תחום ניתוח סטטיסטי
אגף מדען ראשי

לואיזה בורק
תחום ניתוח סטטיסטי
אגף מדען ראשי

תוכן עניינים

1	מבוא	3
2	מודל לאמידת שיעורי למידה בגילאים השונים	3
2.1	הגדרת משתנים	3
2.2	חיזוי שיעורי למידה באמצעות הפונקציה האמפירית	4
2.3	חיזוי שיעורי למידה בשיטה של ממוצע נע	5
2.4	חיזוי מספר התלמידים בכיתה א' ברמה ארצית ללא תלות בשיעורי הלמידה	5
3	חיזוי מספר תלמידים בכיתה א' למחוז גיאוגרפי	5
4	תחזיות מספר התלמידים במגזר הערבי	7
5	תוצאות	7
6	סיכום	8
7	נספח לוחות ותרשימים	9
7.1	מבחני נורמליות לפרופורציות תלמידים במחוזות	9
7.2	לוחות ותרשימים	10

1 מבוא

במסגרת חישוב תחזיות למספר התלמידים במערכת החינוך בישראל לפי שלבי למידה פותחו מודלים סטטיסטיים לחיזוי מספר התלמידים בכיתות א'. הכנת תחזיות אלה הינה אחד השלבים בחיזוי מספר התלמידים הכולל בכל שלב חינוך, דהיינו יסודי, חט"ב, וחט"ע, המבוסס על שיעורי מעבר של התלמידים משנה לשנה. חישוב תחזיות התלמידים בכיתה א' מבוסס על נתוני מספר תלמידים בכיתה א' בין השנים 1996-2002. עקב המגבלה של מספר קטן של תצפיות הוצעו כמה שיטות שונות לחיזוי מספר התלמידים בכיתה א'. תוצאות החיזוי שנתקבלו מכל שיטה ניתנות לשילוב במערכת הכוללת לחיזוי מספר התלמידים בשלבי החינוך השונים.

2 מודל לאמידת שיעורי למידה בגילאים השונים

בניסיון לעקוב אחר הקשר שבין מספר התלמידים בכיתות א' ואוכלוסייה בגילאים מתאימים נלקחו ונותחו נתונים אודות שיעורי למידה של אוכלוסיית ילדים בגילאים שונים בתקופה 1996-2002. להלן מספר הגדרות וסימונים המתארים את הקשר הפונקציונלי בין שיעורי הלומדים למספר התלמידים בכיתות א'.

2.1 הגדרת משתנים

על סמך הנתונים של מספר התלמידים בכיתה א' בגילאים שונים ונתוני אוכלוסייה הוגדרו המשתנים הבאים:

נסמן תחילה ב- i את גיל התלמיד בכיתה א' $i \in \{5,6,7,A\}$, כאשר A היא קבוצת גילאים אחרים השונים מ-5,6,7 ו- $A = \{4,8,K,19\}$.

נסמן ב- j את השנה שאליה מתייחסים הנתונים 1996,K, 2002, $j = 1996,K$ וב- j' את השנה עברה מחושבת תחזית התלמידים בכיתה א': $j' = 2003,K, 2009$.

יהי S_i^j - מספר התלמידים בכיתה א' בגיל i , בשנה j .

יהי P_i^j - אוכלוסייה בגיל i , בשנה j .

ו- $\hat{P}_i^{j'}$ - תחזית האוכלוסייה בגיל i , בשנה j' .

יהי ω_i^j - פרופורציית הלומדים בכיתה א' מתוך אוכלוסייה בגיל i בשנה j המחושבת כדלהלן:

$$\omega_i^j = \frac{S_i^j}{P_i^j}$$

ו- $\hat{\omega}_i^{j'}$ - תחזית פרופורציית הלומדים בכיתה א' מתוך אוכלוסייה בגיל i בשנה j' .

בתרשימים א.1-4, ב.1-4 המתארים בין השנים 1996-2002 את שינוי שיעורי הלמידה בגילאים שונים על פני זמן יחד עם תחזיות על פי שתי שיטות שונות שיתוארו בהמשך. בשל מיעוט מספר התצפיות והקושי בזיהוי מגמות לאורך זמן, הוצעו מספר שיטות לאמידת $\hat{\omega}_i^{j'}$.

2.2 חיזוי שיעורי למידה באמצעות הפונקציה האמפירית

על סמך תצפיות של ω_i^j , $j = 1996, K, 2003$, נאמדה פרופורציית $\hat{\omega}_i^{j'}$ לשנים $j' > 2003$ על בסיס הפונקציה האמפירית של התצפיות.

מעונינים למצוא $\hat{\omega}_i^{j'}$, כך שהביטוי הבא יובא למינימום:

$$\min_{\hat{\omega}_i^{j'}} P\left(\left|\omega_i^j - \hat{\omega}_i^{j'}\right| > \delta\right) \quad 0 < \delta < 1 \quad (2.1)$$

מאחר ו- $\omega_i^j - \hat{\omega}_i^{j'} \in [0,1]$, כלומר הביטוי חסום מלעיל ומלרע, לא ייתכן כי $\omega_i^j - \hat{\omega}_i^{j'} \sim N(,)$. נניח כי $\omega_i^j - \hat{\omega}_i^{j'} \sim F(x)$, כאשר $F(x)$ הינה פונקציית התפלגות כלשהי. היות ו- $F(x)$ לא ידועה, נשתמש בפונקציה ההתפלגות האמפירית:

$$\hat{F}(x) = \frac{1}{n} \sum_{j=1}^n I\{\omega_i^j < x\}, \quad x \in [0,1] \quad (2.2)$$

אך מאחר ומתקיים מתוך (2.1) כי

$$P\left(\left|\omega_i^j - \hat{\omega}_i^{j'}\right| > \delta\right) = P\left(\omega_i^j > \delta + \hat{\omega}_i^{j'}\right) + P\left(\omega_i^j < \hat{\omega}_i^{j'} - \delta\right) \quad (2.3)$$

האומד $\hat{\omega}_i^{j'}$ יתקבל מתוך

$$\begin{aligned} \hat{\omega}_i^{j'} &= \arg \min_{\hat{\omega}^*} \int_0^1 \frac{1}{n} \sum_{j=1}^n \left(I\{\omega_i^j < \hat{\omega}^* - \delta\} + I\{\omega_i^j > \delta + \hat{\omega}^*\} \right) d\delta = \\ & \arg \min_{\hat{\omega}^*} \int_0^1 \left(F(\hat{\omega}^* - \delta) + 1 - F(\hat{\omega}^* + \delta) \right) d\delta \end{aligned} \quad (2.4)$$

חישוב $\hat{\omega}^*$ נעשה באופן נומרי על ידי תכנות תהליך מזעור הפונקציה תוך שימוש בתוכנת R.

חזוי מספר התלמידים בכיתה א' בגיל i בשנה j ע"פ שיטת זו יתקבל מתוך:

$$\hat{S}_i^{j'} = \hat{\omega}^* \hat{P}_i^{j'} \quad (2.5)$$

ולפיכך, תחזית למספר הכולל של תלמידים בכיתה א' (כל הגילאים) תחושב באופן הבא:

$$\hat{S}^{j'} = \sum_i \hat{S}_i^{j'} \quad (2.6)$$

2.3 חיזוי שיעורי למידה בשיטה של ממוצע נע

שיטה נוספת שהוצעה לצורך אמידה וחיזוי פרופורציות הלומדים בגילאים השונים מבוססת על חישוב ממוצע נע של q תצפיות אחרונות. לפיכך תחזית פרופורציית הלומדים בגיל i בשנת j' תחושב באופן הבא:

$$\omega_i^{(j')} = q^{-1} \sum_{k=j'-q}^{j'-1} \omega_i^{(k)} \quad (2.7)$$

ולפיכך מספר התלמידים בכיתה א' בגיל i החזוי לשנת j' הינו

$$\hat{S}_i^{j'} = \hat{\omega}_i^{(j')} \hat{P}_i^{j'} \quad (2.8)$$

לצורך חיזוי שיעורי למידה בכיתה א' בגילאים השונים נלקח $q = 3$ והמספר הכולל של התלמידים חושב ע"פ (2.6).

2.4 חיזוי מספר התלמידים בכיתה א' ברמה ארצית ללא תלות בשיעורי הלמידה

בנוסף למודלים שתוארו ב- (2.4) – (2.8) הוכנו תחזיות אלטרנטיביות ללא התייחסות לגורם שכבת הגיל. המודל האלטרנטיבי הינו מודל תלות בזמן ומתוך הנחת אי התלות בין התצפיות והתפלגותן הנורמלית (לאחר טרנספורמציה לוגריתמית). לאחר בדיקות טיב התאמה של מודלים לתצפיות סך התלמידים בכיתה א' בין השנים 1996-2002 נבחרו שני המודלים הבאים:

$$\log(S^j) = \gamma \tau_j^\delta + \nu, \quad \tau_j = (j-1996)+1 \quad \nu \sim N(0, \sigma_\nu^2) \quad (2.9)$$

$$\log(S^j) = \beta_0 + \beta_1 \tau_j + \nu, \quad \tau_j = (j-1996)+1 \quad \nu \sim N(0, \sigma_\nu^2) \quad (2.10)$$

3 חיזוי מספר תלמידים בכיתה א' למחוז גיאוגרפי

על סמך הנתונים של מספר התלמידים בכיתה א' בשנים 1996-2003 חושבו לכל שנה פרופורציות של תלמידים בכיתה א' מתוך סך התלמידים בכיתה א' ברמה הארצית. נגדיר p_k^j כפרופורציית מספר התלמידים בכיתה א' במחוז k , $k = 1, K, 6$, בשנה j , $j = 1996, K, 2003$ מתוך המספר הכולל של תלמידים במגזר מסוים (עברי, ערבי).

$$\sum_j p_k^j = 1, \quad p_k^j = \frac{S_k^j}{S^j}$$

כאשר S_k^j הינו מספר התלמידים בכיתה א' במחוז k , בשנה j .

ו- S^j הינו מספר התלמידים בכיתה א' בכל הארץ בשנה j .

לפני התאמת המודלים בוצעו מבחנים לבדיקת נורמליות של פרופרציות התלמידים לאחר טרנספורמצית לוגיט (ראה לוחות 7.1 בנספח לוחות ותרשימים).

עבור מחוזות ירושלים ($k = 1$), ותל אביב ($k = 5$) הותאם מודל חזקה:

$$\log it(p_k^j) = a_k \tau_j^{\beta_k} + \varepsilon_j^k \quad (3.1)$$

$$k = 1, 5 \quad \tau_j = (j - 1996) + 1, \quad \varepsilon_j^k \sim N(0, \sigma_k^2)$$

עבור מחוזות צפון ($k = 2$), חיפה ($k = 3$) ומרכז ($k = 4$) הותאם מודל רגרסיה ליניארית:

$$\log it(p_k^j) = \beta_{0k} + \beta_{1k} \tau_j + \varepsilon_j^k \quad (3.2)$$

$$k = 2, 3, 4, \quad \tau_j = (j - 1996) + 1, \quad \varepsilon_j^k \sim N(0, \sigma_k^2)$$

$$\sum_k \hat{p}_k^{j'} = 1 \quad j' = 2003, K, 2009 \quad \text{נזכיר כי}$$

אי לכך, פרופורציית תלמידים בכיתה א' במחוז דרום¹ חושבה באופן שישמר את סכום הפרופורציות השווה ל-1:

$$\hat{p}_6^j = 1 - \sum_{k=1}^5 \hat{p}_k^j \quad (3.3)$$

על סמך פרופורציות תלמידי כיתה א' שנחזו באמצעות המודלים (3.1) ו- (3.2) חושבו תחזיות של מספר התלמידים \hat{S}_k^j בכל מחוז גיאוגרפי באופן הבא:

$$k = 1, K, 6, \quad j' = 2003, K, 2009, \quad \hat{S}_k^{j'} = \hat{p}_k^{j'} S^{j'} \quad (3.4)$$

כאשר $\hat{p}_k^{j'}$ - תחזית פרופורציית תלמידים בכיתה א' במחוז k לשנה j' ,

$\hat{S}^{j'}$ - תחזית מספר התלמידים הכולל בכיתה א' ברמה ארצית חזוי לשנה j' ע"פ מודל אמידת שיעורי למידה בגילאים השונים (2.6) או לחילופין ע"פ מודל חיזוי סך התלמידים בכיתה א' ברמה הארצית (2.9).

¹הסיבה שמחוז דרום נבחר להיות זה שפרופורציית התלמידים בו תחושב באמצעות נוסחה האילוף היא שטיב התאמה של המודלים שהותאמו עבור מחוז זה היה נמוך וכמו כן רמת מובהקות המודל לא הייתה מספקת (>0.05)

4 תחזיות מספר התלמידים במגזר הערבי

התחזיות נעשו על פי אותן שיטות שתוארו בפרקים קודמים, אלא שלא ניתנה התייחסות לרמה המחזית של מספר התלמידים והתחזיות הוכנו ברמה הארצית. הדבר נבע מריכוזיות מוסדות חינוך באזור הגליל ומספר מועט של מוסדות בשאר המחוזות.

5 תוצאות

תחזיות מספר תלמידים ברמה הארצית חושבו לסירוגין ע"פ המודלים הבאים:

- מודל אמידת שיעורי למידה בגילאים השונים ע"פ פונקציית התפלגות אמפירית של התצפיות ושימוש בתחזיות אוכלוסייה לצורך חיזוי מספר התלמידים הכולל (2.6-2.1)
- מודל אמידת שיעורי למידה בגילאים השונים ע"פ שיטת ממוצע נע של שלוש תצפיות אחרונות תחזיות מספר התלמידים הכולל נעשו באמצעות המודלים הבאים (2.8-2.7)
- מודל רגרסיה ליניארית לאמידה וחיזוי מספר תלמידים כלל ארצי כפונקציה של תקופת זמן (שנה) (2.9)
- מודל רגרסיה (מודל חזקה) לאמידה וחיזוי מספר תלמידים כלל ארצי כפונקציה של תקופת זמן (שנה) (2.10)

תחזיות של פרופורציות מספר התלמידים בכיתות א' מתוך המספר הכללי ארצי נעשו ע"פ מודלים (3.1)–(3.3) בהתאם למחוז הגיאוגרפי ותחזיות מספר התלמידים בכיתות א' לכל מחוז נעשו על פי מודל (3.4).

בחירת המודלים נעשתה על פי הקריטריונים הבאים: מובהקות המודל ומקסימום טיב התאמה (מקדם מתאם ריבועי)

להלן טבלה עם אפיונים סטטיסטיים של המודלים שהותאמו לפרופורציות התלמידים בכיתה א' במחוזות השונים

טבלה 5-1 : סוגי מודלים לאמידה וחיזוי פרופורציות תלמידים בכיתה א' במחוזות השונים – רמות מובהקות של המודלים

מספר התלמידים		פרופורציות תלמידים במחוזות						סוג תחזית
ארצי	ארצי	זרוב	תל אביב	מרכז	חיפה	צפון	י-ם	מחוז
חוקה	ליניארי	ראה (3.3)	חוקה	ליניארי	ליניארי	ליניארי	חוקה	סוג המודל
0.6324	0.683	-----	0.9734	0.7466	0.974	0.898	0.9907	טיב התאמה
<0.0001	<0.0001	-----	<0.0001	0.0057	<0.0001	<0.0001	<0.0001	מובהקות המודל

תחזיות פרופורציות תלמידים בכיתה א' ברמת מחוז ותחזיות מספר התלמידים ע"פ שיטות שונות שהוצעו ניתן לראות בנספח לוחות ותרשימים (ראה נספח 7).

6 סיכום

דוח זה תיאר כאמור את שלבי העבודה שעסקה בהכנת תחזיות למספר התלמידים בכיתות א' ברמת המחוז הגיאוגרפי. עקב מספר מצומצם של תצפיות נתונות הוצעו כמה שיטות ומודלים לחיזוי מספר התלמידים בכיתה א'. השיטה הראשונה הסתמכה על הקשר בין מספר התלמידים לבין שיעורי למידה בגילאים השונים, כאשר השניה התבססה על הקשר בין בתקופה לסך התלמידים בכיתה א' ברמה הארצית. בתהליך החיזוי נעשה שימוש בפונקציית התפלגות אמפירית לצורך אמידה של שיעורי למידה בגילאים השונים וכמו כן, הוצעה שיטת ממוצע נע כדרך נוספת לאמוד את שיעורי הלמידה בגילאים השונים. ניתן לעיין בלוחות ותרשימים המתארים את ממצאי המודלים בנספח 7.

7 נספה לוחות ותרשימים

7.1 מבחני נורמליות לפרופורציות תלמידים במחוזות

7.1.א: פרופורציית תלמידים במחוז ירושלים (לאחר טרנספורמציה לוגית) Tests for Normality:

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.956287	Pr < W 0.7741
Kolmogorov-Smirnov	D 0.168425	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.033789	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.218531	Pr > A-Sq >0.2500

7.1.ב: פרופורציית תלמידים במחוז צפון (לאחר טרנספורמציה לוגית) Tests for Normality:

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.914273	Pr < W 0.3851
Kolmogorov-Smirnov	D 0.177313	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.046574	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.310297	Pr > A-Sq >0.2500

7.1.ג: פרופורציית תלמידים במחוז חיפה (לאחר טרנספורמציה לוגית) Tests for Normality:

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.96175	Pr < W 0.8266
Kolmogorov-Smirnov	D 0.141033	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.026206	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.180597	Pr > A-Sq >0.2500

7.1.ד: פרופורציית תלמידים במחוז מרכז (לאחר טרנספורמציה לוגית) Tests for Normality:

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.939864	Pr < W 0.6097
Kolmogorov-Smirnov	D 0.17256	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.03671	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.24896	Pr > A-Sq >0.2500

7.1.ה: פרופורציית תלמידים במחוז ת"א (לאחר טרנספורמציה לוגית) Tests for Normality:

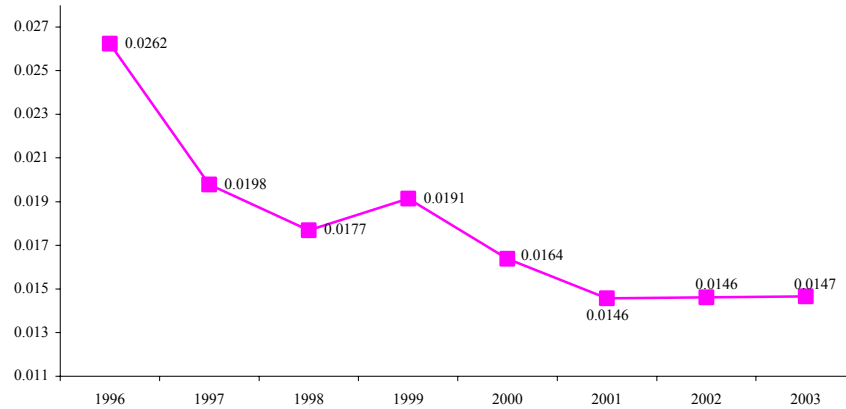
Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.926804	Pr < W 0.4875
Kolmogorov-Smirnov	D 0.181309	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.043018	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.280363	Pr > A-Sq >0.2500

7.1.ו: פרופורציית תלמידים במחוז דרום (לאחר טרנספורמציה לוגית) Tests for Normality:

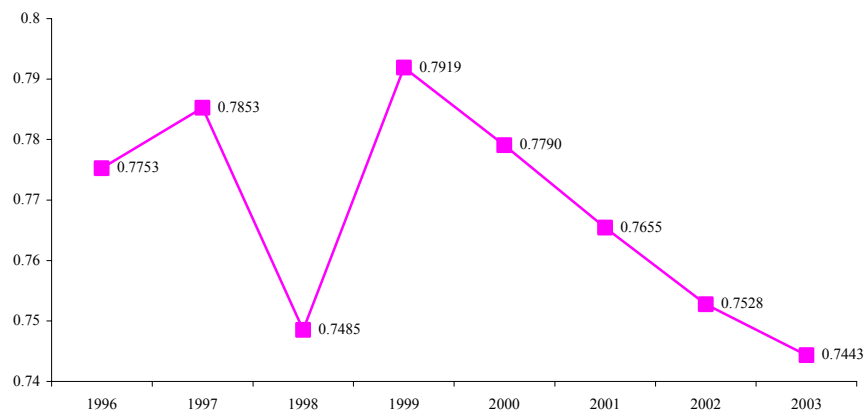
Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.944138	Pr < W 0.6522
Kolmogorov-Smirnov	D 0.191416	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.043961	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.270193	Pr > A-Sq >0.2500

7.2 תרשימים

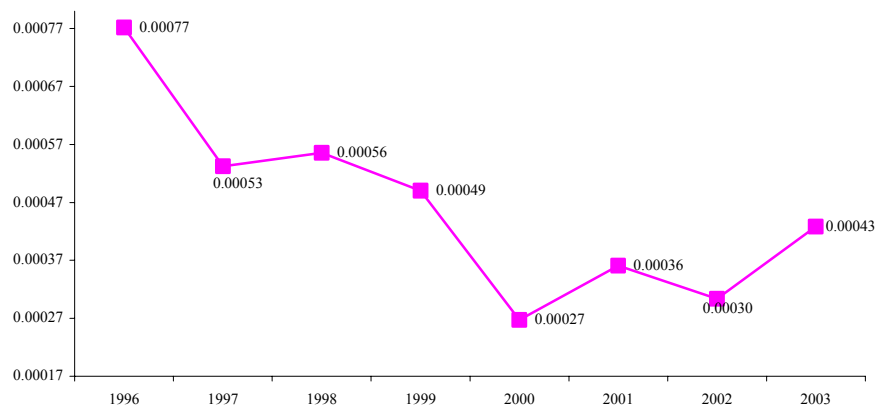
תרשים 1: שיעורי למידה בכיתה א' באוכלוסייה - גיל 5 – מגזר עברי



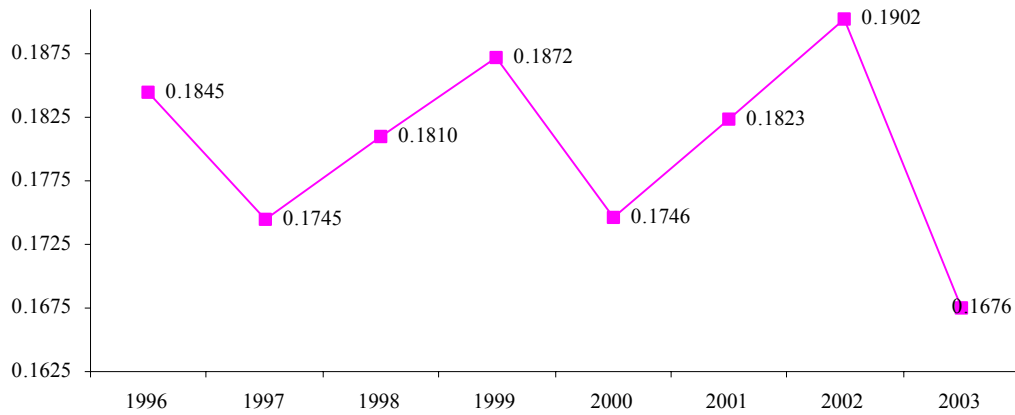
תרשים 2: שיעורי למידה בכיתה א' באוכלוסייה - גיל 6 – מגזר עברי



תרשים 3: שיעורי למידה בכיתה א' באוכלוסייה - גיל 7 – מגזר עברי



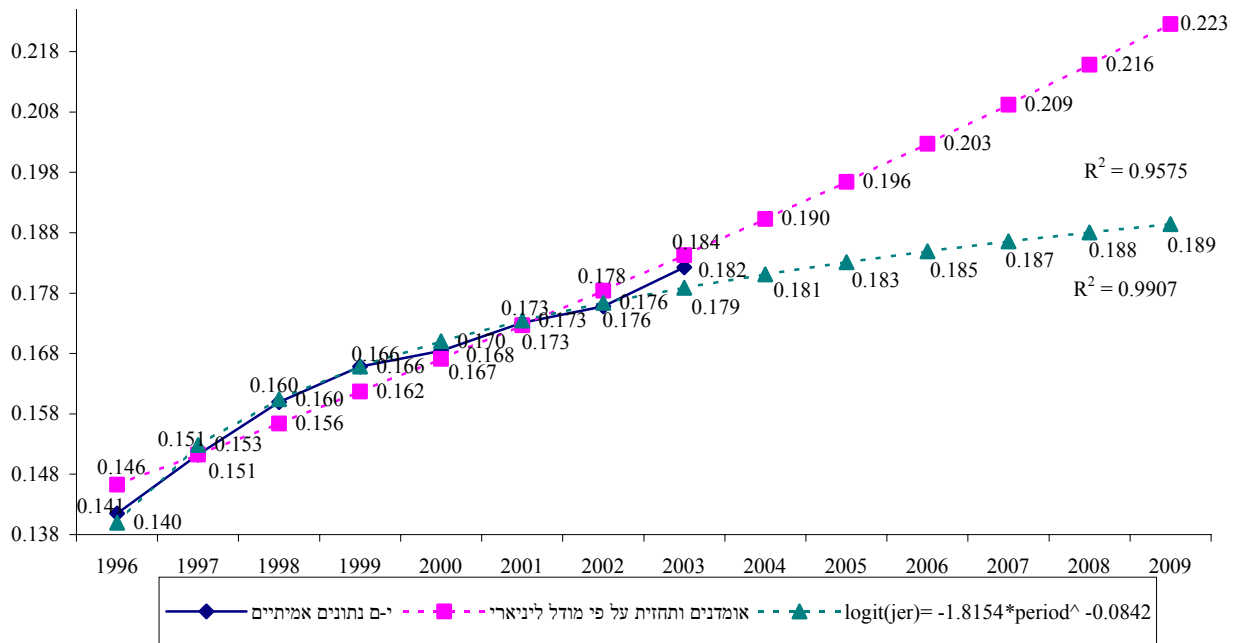
תרשים 4 : שיעורי למידה בכיתה א' באוכלוסייה - גילאים אחרים - מגזר עברי



פרופורציות התלמידים במחוזות גיאוגרפיים - נתונים ותחזיות

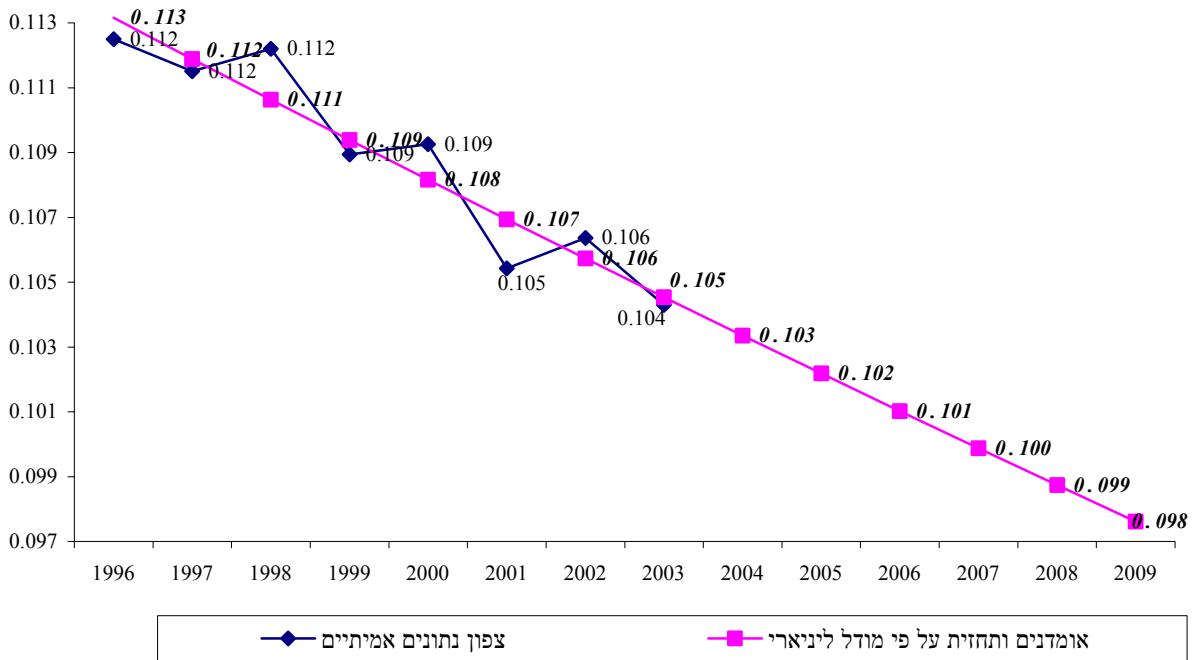
תרשים 5: פרופורציות תלמידים בכיתה א' מתוך סך התלמידים ברמה ארצית - מחוז ירושלים - מגזר עברי

נתונים אמיתיים ותחזיות



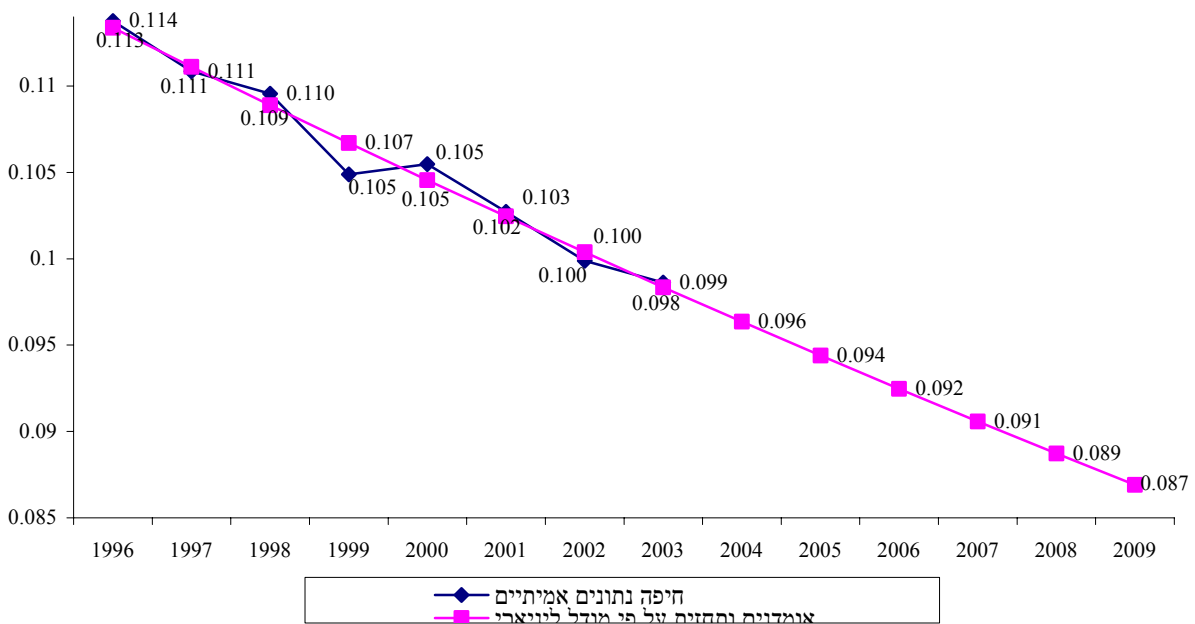
תרשים 6: פרופורציות תלמידים בכיתה א' מתוך סך התלמידים ברמה ארצית - מחוז צפון- מגזר עברי

נתונים אמיתיים ותחזיות

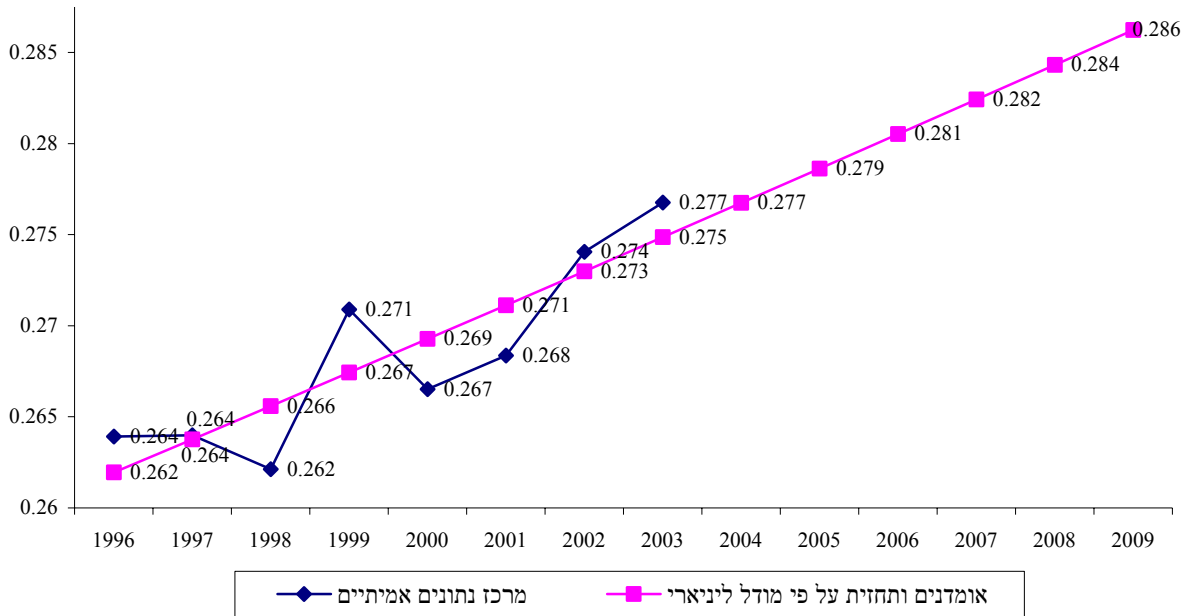


תרשים 7: פרופורציות תלמידים בכיתה א' מתוך סך התלמידים ברמה ארצית - מחוז חיפה - מגזר עברי

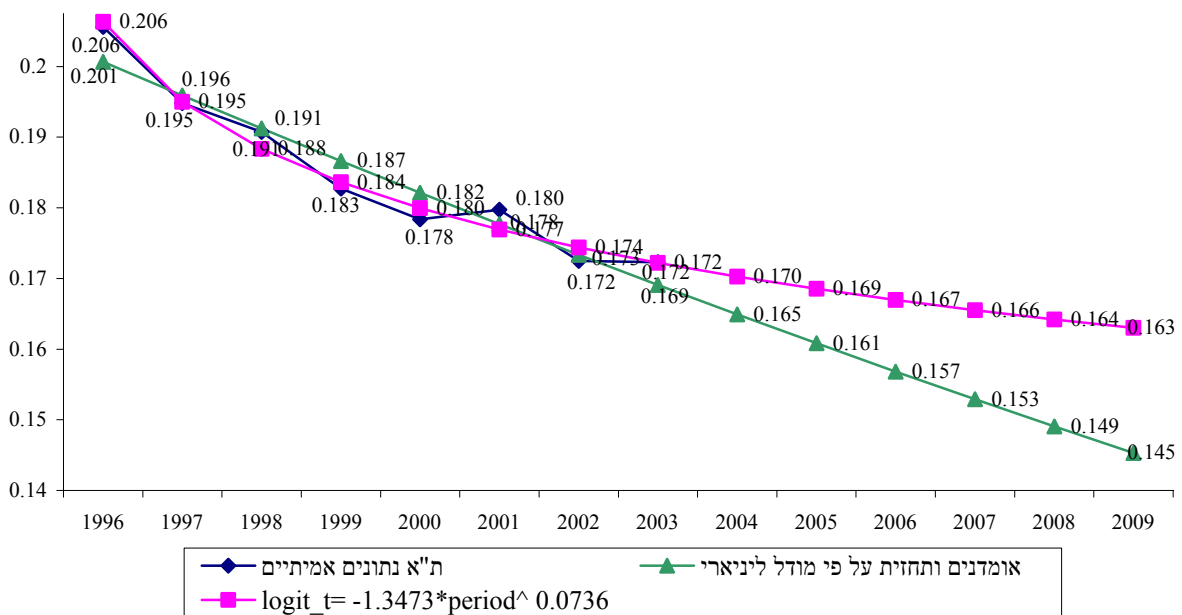
נתונים אמיתיים ותחזיות



תרשים 8: פרופורציות תלמידים בכיתה א' מתוך סך התלמידים ברמה ארצית - מחוז מרכז - מגזר עברי

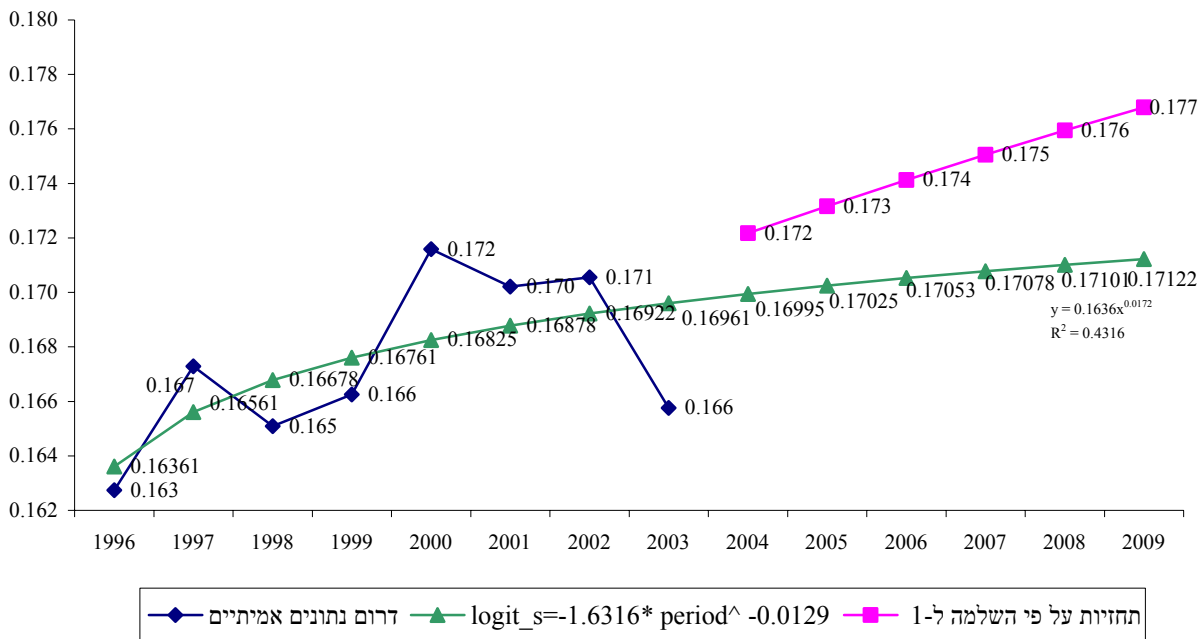


תרשים 9: פרופורציות תלמידים בכיתה א' מתוך סך התלמידים ברמה ארצית - מחוז תל אביב - מגזר עברי
נתונים אמיתיים ותחזית



תרשים 10: פרופרציות תלמידים בכיתה א' מתוך סך התלמידים ברמה ארצית - מחוז דרום - מגזר עברי

נתונים אמיתיים ותחזיות



תרשים 11: תלמידים בכיתה א' - נתונים אמיתיים ותחזיות - מגזר עברי - מחוז ירושלים

