הלשכה המרכזית לסטטיסטיקה
גף ניתוח סטטיסטי

אמידת היקף הנישואין בין בני-דודים
בקרב המוסלמים בישראל

פרופ׳ גד נתן ,לואיזה בורק
ד״ר גיא קינן , מרט צחמן

נובמבר 1999

אמידת היקף הנישואין בין בני- דודים בקרב המוסלמים בישראל

1. כללי

יש עניין במעקב אחרי מגמות בנישואין בין קרובי משפחה בקרב האוכלוסיה המוסלמית בישראל. נישואין כאלה מגדילים את הסיכון שיוולד בעל מום, מסיבות גנטיות. אחוז הנישואין כאלה גם יכול לשמש מדד להתחזקות או להחלשות מוסד החמולה, יחד עם מדדים אחרים.

על פי בקשתו של ד"ר גיא קינן מגף הדמוגרפיה נבדקת במסמך זה האפשרות לעקוב אחרי מגמות בתחום זה בקרב המוסלמים בישראל מנתונים הנמצאים במרשם התושבים. הנתונים המתקבלים מהמרשם הם כלל-ארציים ורב-שנתיים, העובדה שמאפשרת לעקוב אחרי מגמות בהיקף הנישואין בין בני-דודים מדרגה ראשונה: פלוני מתחתן עם הבת של אח של האב שלו.(זה סוג הקשר הקלסי , אך לא ידוע אם הוא השכיח ביותר).

האמידה נעשתה ע"ס קובץ של 154,500 זוגות שהוצא ממרשם התושבים מתאריך 31.10.95. הקובץ כולל את נתונים הבאים : מס' הזהות של בעל ואישה, שמות המשפחה ,סמל ישוב ,שנות הלידה ושנת הנישואין שלהם.

2. שיטת האמידה

2.1 סימונים

N - מספר הזוגות (ידוע);

$M_i\ (F_i)$ - החתן (הכלה) של הזוג ה- i ;

R - מספר הסבים השונים(לא ידוע);

K- מספר השמות השונים של R הסבים $(K \leq R)$

Y_{ik} - פונקצית אינדיקטור למצב שבו לשני בני הזוג i יש סב באותו שם k (ידוע)

Y_k - מס' זוגות שלשני בני זוג יש סב בשם k (ידוע)

Y - סך כל הזוגות שלשני בני הזוג יש סב באותו שם (ידוע)

X_{ik} - פונקצית אינדיקטור למצב שבו לשני בני הזוג i יש אותו סב בשם k (לא ידוע)

X_k - מס' זוגות שלשני בני זוג יש אותו סב בשם k (לא ידוע)

X - סך כל הזוגות שלשני בני הזוג יש אותו סב (לא ידוע)

M_k -מס' חתנים עם סב בשם k $(\sum_{k=1}^{K} M_k = N)$

F_k -מס' כלות עם סב בשם k $(\sum_{k=1}^{K} F_k = N)$

$M_k^* = M_k - X_k$ מספר החתנים עם סב בשם k שאין להם סב משותף עם כלותיהם

$F_k^* = F_k - X_k$ מספר הכלות עם סב בשם k שאין להם סב משותף עם חתניהם

$N^* = N - X$ סך כל הזוגות שלשני בני הזוג אין אותו סב

2.2 הנחה

א. לזוגות שאין לשני בני הזוג אותו סב, מבחינת שם הסב , הזיווג הוא מקרי , דהיינו :

$$P(Y_{ik} = 1 | X_{ik} = 0) = \frac{M_k^* F_k^*}{N^* N^*}$$

ב. F_k^* ו M_k^* קבועים.

2.3 אמידת X

האומדנים של X_k ו- X התקבלו כפתרונות של מערכת משוואות הבאה :

$$(M_k - X_k)(F_k - X_k) = (Y_k - X_k)(N - X)$$

תחת המגבלה $X_k \leq Y_k$.

האמידה נעשתה עבור 100 שמות סב השכיחים ביותר (כלומר K= 100) בכל תקופה ואזור , כאשר בחלק מהתקופות והאזורים היה כיסוי של כל שמות סב הידועים.

נציין שבכל תקופה ואזור התוצאות התקבלו מפתרון של 100 משואות המופיעות לעיל.
הערה : 100 שמות סב השכיחים ביותר מכסים חלק ניכר מסך שמות סב הידועים ובמספר תקופות ואזורים אף את כל שמות סב הידועים.

האמדנים חושבו עבור שני מצבים קיצוניים :
I. בין הזוגות האחרים והלא ידועים אין זוגות עם סב זהה.
II. לכל הזוגות האחרים והלא ידועים יש סב זהה.

2.4 אמידת שונות האומד

שונות האומד מותנית ב F_k^* ו M_k^* קבועים.

$$V(\hat{X}) = A^{-1} V(Y) (A^{-1})'$$

כאשר

$$A_{kl} = \frac{M_l F_l}{N^2} + \delta_k^l \frac{N - M_k - F_k}{N}$$

$$[V(Y)]_{kl} = \frac{M_k^* F_k^* (N^* \delta_k^l - M_l^*)(N^* \delta_k^l - F_l^*)}{(N^*)^2 (N^* - 1)}$$

ההסבר בנספח 1.

3. תוצאות וממצאים

בלוח 1 מוצגות התוצאות עבור המוסלמים בכל המדינה וכן עבור שלשה אזורים: מחוז דרום (בדוים), עיר ירושלים ושאר המדינה.

לכל אזור מופיעות 3 סדרות :
1. אחוז הזוגות עם שמות הסבים זהים.
2. אומדן לאחוז זוגות שבהם בני הזוג הם בני דודים מדרגה ראשונה כאשר בין הזוגות האחרים והלא ידועים אין זוגות עם סב זהה.
3. אומדן לאחוז זוגות שבהם בני הזוג הם בני דודים מדרגה ראשונה כאשר לכל הזוגות האחרים והלא ידועים יש סב זהה.

בדיאגרמות מופיעות שלוש הסדרות הנ״ל וכן רווחי סמך ברמה של 95% לאחוז הזוגות עם סב זהה לפי הנחה א׳ וב׳ .

בהסתכלות בלוחות ודיאגרמות המתארות את הממצאים ניתן להבחין שבסך הכל יש ירידה באחוז הנישואין בין בני-דודים , כאשר בעיר ירושלים הירידה יותר חדה . בדרום הארץ המצב יציב מאמצע שנות ה - 70 ,ובשאר המדינה קיימת ירידה מתונה באחוז הנישואין בין בני-דודים.

נציין שבתאים בהם יש אחוז גבוה של זוגות עם שמות הסבים לא ידועים ייתכן ויתקבל אומדן לפי ההנחה ב׳ שהוא גבוה יותר מאחוז הזוגות עם שמות הסבים זהים (סדרה 1) מכוון שהמדד האחרון לא כולל את הזוגות ששם הסב של לפחות אחד מבני הזוג לא ידוע.
הדבר בולט באזור ירושלים בשנים 1955-1959.

אומדני סטיות התקן שחושבו מופיעים ליד האומדן לאחוז הזוגות בתוך הסוגריים (בלוח1).

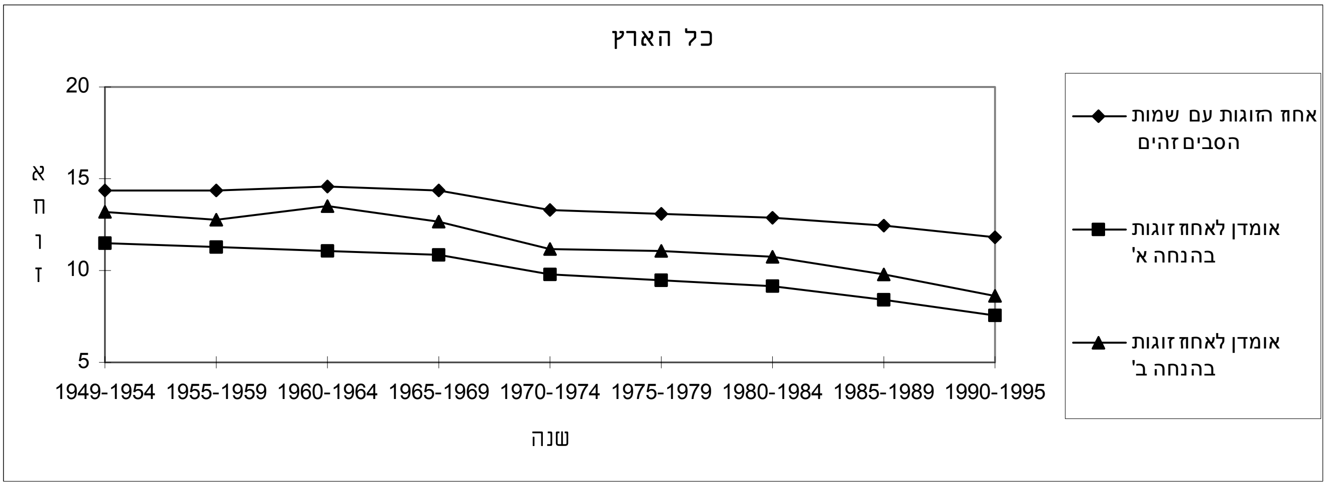
אומדני סטיות התקן התקבלו נמוכים בד״כ, לדוגמא עבור האומדן בשנים 1990-1995 שהתקבל להיות 8.6% אומדן לסטית התקן הינו 0.053% . אומדן לסטית התקן הגבוה ביותר התקבל באזור הדרום בשנים 1949-1954 והוא 1.139%.

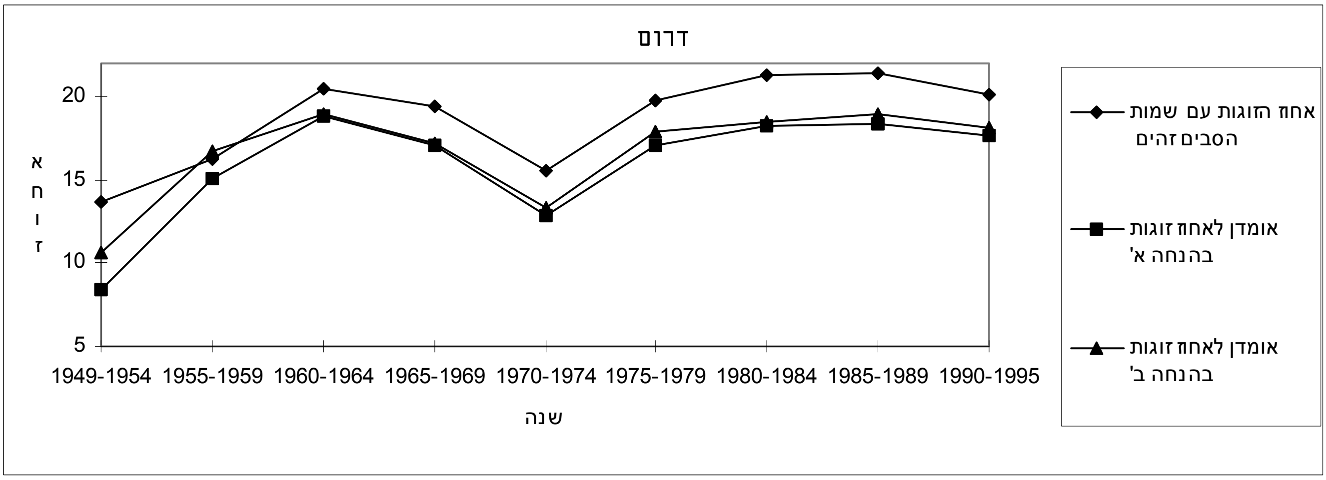
לוח 1 : אחוז הזוגות שיש לשני בני הזוג סב באותו שם ואותו סב

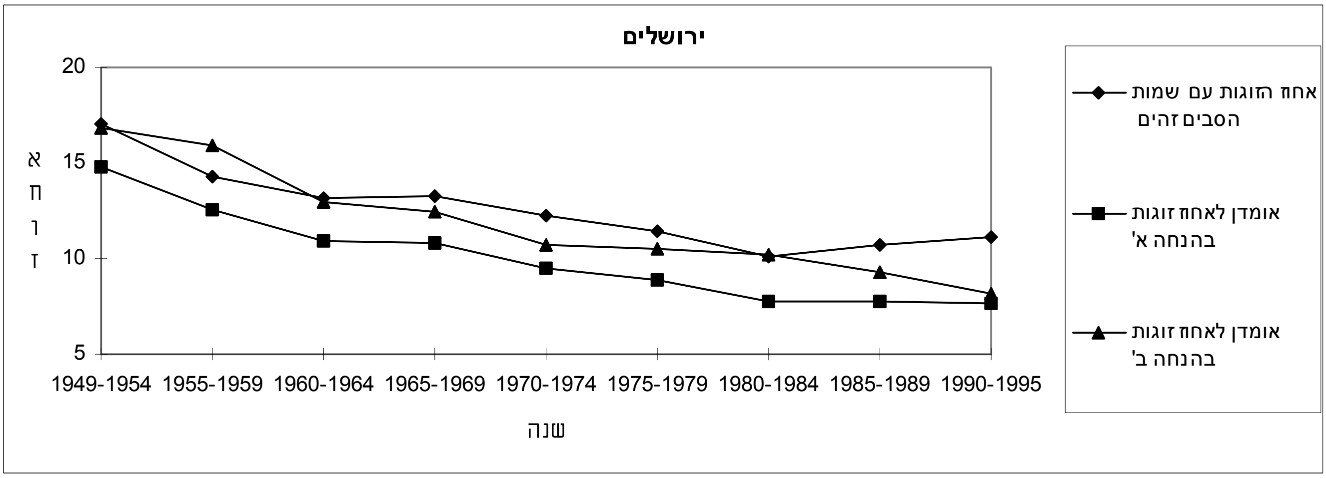
אזור			ס״ה	1949-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1995
ס״ה	אחוז הזוגות <u>עם סב באותו שם</u>		12.9	14.4	14.4	14.6	14.4	13.3	13.1	12.9	12.5	11.8
	אחוז זוגות שיש לשני	בהנחה א׳	(0.03) 8.8	(0.21) 11.5	(0.18) 11.3	(0.13) 11.1	(0.11) 10.9	(0.09) 9.8	(0.08) 9.5	(0.07) 9.1	(0.06) 8.4	(0.05) 7.6
	בני הזוג <u>אותו סב</u>	בהנחה ב׳	(0.03)10.6	(0.21)13.2	(0.18) 12.8	(0.13) 13.5	(0.11) 12.7	(0.1) 11.2	(0.1) 11.1	(0.09) 10.7	(0.09) 9.8	(0.05) 8.6
דרום	אחוז הזוגות <u>עם סב באותו שם</u>		19.9	13.7	16.3	20.5	19.4	15.5	19.8	21.3	21.4	20.1
	אחוז זוגות שיש לשני	בהנחה א׳	(0.09) 16.9	(1.14) 8.4	(1.00) 15.0	(0.65) 18.8	(0.47) 17.1	(0.4) 12.9	(0.33) 17.1	(0.28) 18.2	(0.2) 18.4	(0.16) 17.7
	בני הזוג <u>אותו סב</u>	בהנחה ב׳	(0.09) 17.7	(1.08) 10.6	(1.02) 16.7	(0.65) 19	(0.47) 17.2	(0.4) 13.3	(0.34) 17.9	(0.28) 18.5	(0.2) 18.9	(0.16) 18.1
י-ם	אחוז הזוגות <u>עם סב באותו שם</u>		11.5	17	14.3	13.2	13.3	12.2	11.4	10.1	10.7	11.1
	אחוז זוגות שיש לשני	בהנחה א׳	(0.06) 8.5	(0.37) 14.8	(0.4) 12.6	(0.35) 10.9	(0.27) 10.8	(0.2) 9.5	(0.18) 08.9	(0.15) 7.8	(0.13) 7.8	(0.12) 7.7
	בני הזוג <u>אותו סב</u>	בהנחה ב׳	(0.06) 10.3	(0.38) 16.8	(0.41) 15.9	(0.35) 13	(0.27) 12.4	(0.2) 10.7	(0.18) 10.5	(0.15) 10.2	(0.13) 9.3	(0.12) 8.2
שאר	אחוז הזוגות <u>עם סב באותו שם</u>		12.3	13.5	14.3	14.5	14.1	13.4	12.8	12.8	11.7	10.7
המדינה	אחוז זוגות שיש לשני	בהנחה א׳	(0.03) 8.1	(0.28) 10.1	(0.22) 10.8	(0.14) 10.8	(0.13) 10.5	(0.12) 9.8	(0.1) 9.1	(0.09) 8.7	(0.08) 7.5	(0.07) 6.2
	בני הזוג <u>אותו סב</u>	בהנחה ב׳	(0.03) 9.2	(0.28) 11.5	(0.22) 11.6	(0.14) 11.6	(0.13) 11.2	(0.12) 10.8	(0.1) 10.3	(0.09) 9.7	(0.08) 8.3	(0.07) 6.8

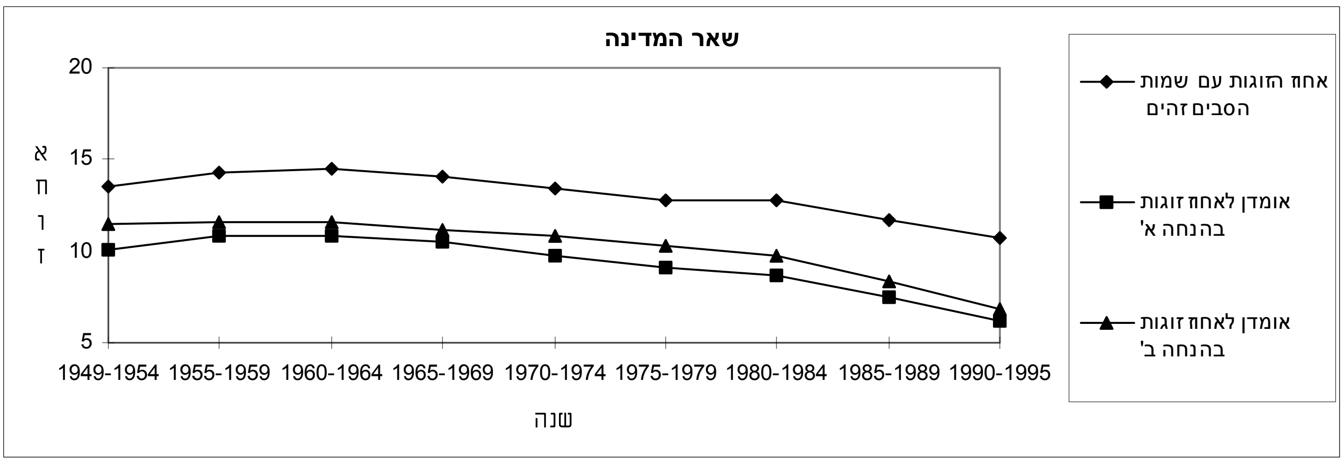
בסוגריים נתונות סטיות התקן באחוזים

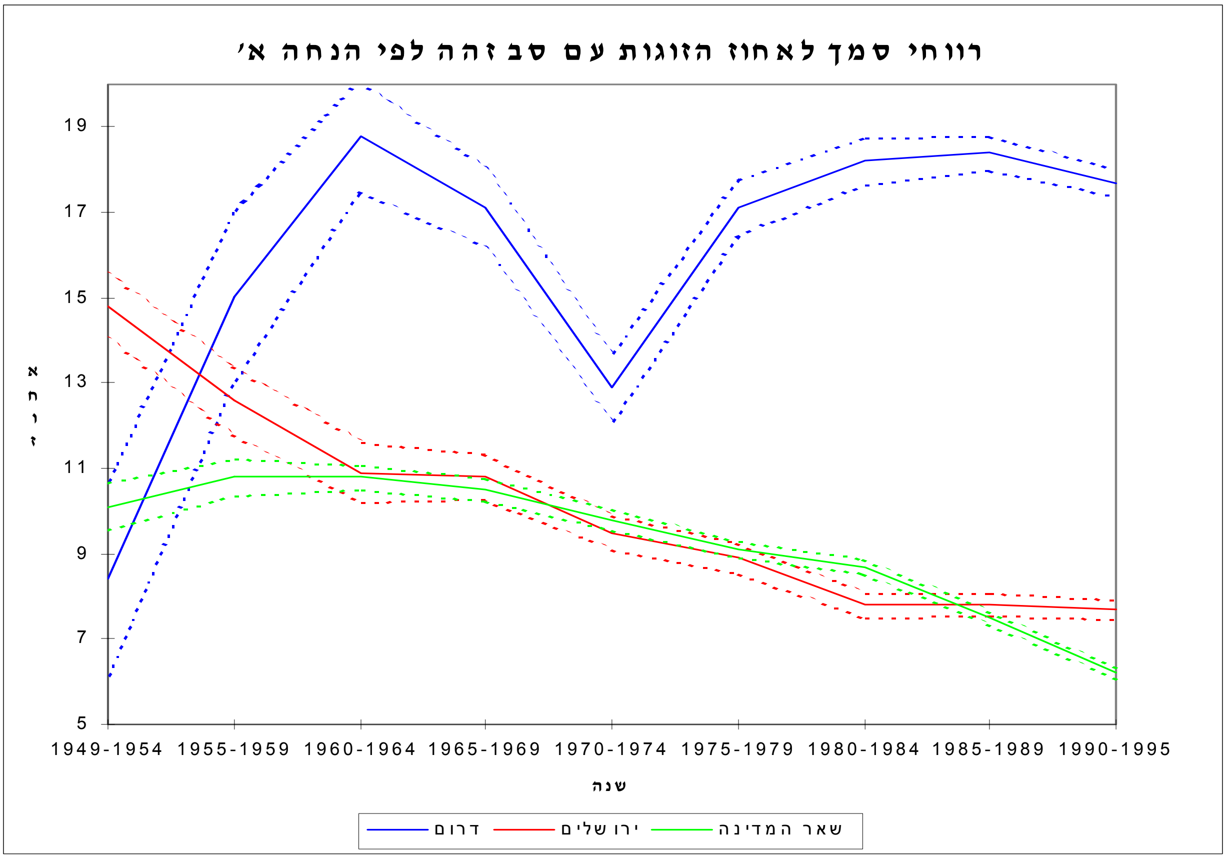
דיאגרמה 1

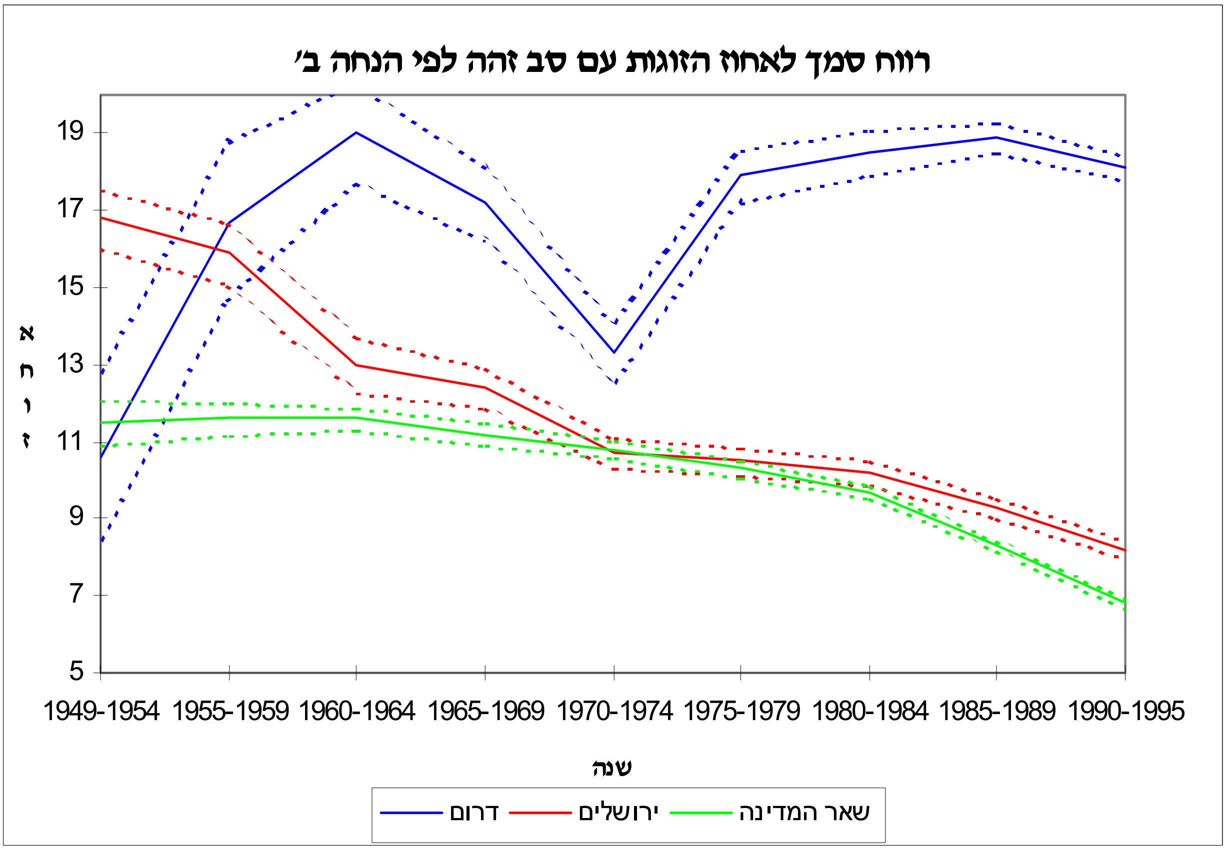

כל הארץ
20
15
10
5
אחוז
1949-1954 1955-1959 1960-1964 1965-1969 1970-1974 1975-1979 1980-1984 1985-1989 1990-1995
שנה
אחוז הזוגות עם שמות הסבים זהים
אומדן לאחוז זוגות בהנחה א'
אומדן לאחוז זוגות בהנחה ב'


דרום
20
15
10
5
אחוז
1949-1954 1955-1959 1960-1964 1965-1969 1970-1974 1975-1979 1980-1984 1985-1989 1990-1995
שנה
אחוז הזוגות עם שמות הסבים זהים
אומדן לאחוז זוגות בהנחה א'
אומדן לאחוז זוגות בהנחה ב'


ירושלים
20
15
10
5
אחוז
1949-1954 1955-1959 1960-1964 1965-1969 1970-1974 1975-1979 1980-1984 1985-1989 1990-1995
שנה
אחוז הזוגות עם שמות הסבים זהים
אומדן לאחוז זוגות בהנחה א'
אומדן לאחוז זוגות בהנחה ב'


שאר המדינה
20
15
10
5
אחוז
1949-1954 1955-1959 1960-1964 1965-1969 1970-1974 1975-1979 1980-1984 1985-1989 1990-1995
שנה
אחוז הזוגות עם שמות הסבים זהים
אומדן לאחוז זוגות בהנחה א'
אומדן לאחוז זוגות בהנחה ב'


רווחי סמך לאחוז הזוגות עם סב זהה לפי הנחה א'
19
17
15
13
11
9
7
5
אחוז
1949-1954 1955-1959 1960-1964 1965-1969 1970-1974 1975-1979 1980-1984 1985-1989 1990-1995
שנה
דרום ירושלים שאר המדינה


רווח סמך לאחוז הזוגות עם סב זהה לפי הנחה ב'
19
17
15
13
11
9
7
5
אחוז
1949-1954 1955-1959 1960-1964 1965-1969 1970-1974 1975-1979 1980-1984 1985-1989 1990-1995
שנה
דרום ירושלים שאר המדינה

נספח 1

מטריצת השונויות של התפלגות מולטינומית מותנית בשוליים

בלי הגבלת הכלליות נסתכל בלוח 3×3 :

$$
\begin{array}{ccc|c}
x_{11} & x_{12} & x_{13} & m_1 \\
x_{21} & x_{22} & x_{23} & m_2 \\
x_{31} & x_{32} & x_{33} & m_3 \\
\hline
n_1 & n_2 & n_3 & \mathrm{N}
\end{array}
$$

נסמן

$$\underline{x} = (x_{11}, x_{12}, x_{13})'$$
$$\underline{m} = (m_1, m_2, m_3)'$$
$$\underline{n} = (n_1, n_2, n_3)'$$

אזי ההסתברות ל- $\underline{x}$ בהינתן השוליים $\underline{m}$ ו $\underline{n}$ קבועים ובהנחת אי תלות הינה

$$P_N(\underline{x}|\underline{m},\underline{n}) = \frac{n_1!n_2!n_3!m_1!m_2!m_3!}{N!x_{11}!x_{12}!\ldots x_{33}!}$$

(המאמר "Exact Inference for Categorical data" של Cyrus R. Mehta and Nitin R. Patel)

עתה נגדיר

$$\Gamma = \{\underline{x}: \sum_j x_{ij} = m_i ; \sum_i x_{ij} = n_j\}$$

על מנת לחשב את השונות של x_{11} נחשב קודם הביטוי הבא :

$$\mathrm{E}[x_{11}(x_{11}-1)|\underline{m},\underline{n}] = \sum_{\underline{x}\in\Gamma} x_{11}(x_{11}-1)P_N(\underline{x}|\underline{m},\underline{n}) =$$

$$= \sum_{x\in\Gamma} \frac{n_1(n_1-1)(n_1-2)!n_2!n_3!m_1(m_1-1)(m_1-2)!m_2!m_3!}{N(N-1)(N-2)!(x_{11}-2)!x_{12}!\ldots x_{33}!}$$

$$= \frac{n_1(n_1-1)m_1(m_1-1)}{N(N-1)} \sum_{\underline{\tilde{x}}\in\tilde{\Gamma}} P_{\tilde{N}}(\underline{\tilde{x}}|\underline{\tilde{m}},\underline{\tilde{n}})$$

כאשר

$$\underline{\tilde{x}} = (x_{11}-2, x_{12}, \ldots, x_{33}) \qquad \tilde{N} = N-2$$
$$\underline{\tilde{n}} = (n_1-2, n_2, n_3) \qquad \underline{\tilde{m}} = (m_1-2, m_2, m_3)$$
$$\tilde{\Gamma} = \{\underline{\tilde{x}}: \sum_j \tilde{x}_{ij} = \tilde{m}_i ; \sum_i \tilde{x}_{ij} = \tilde{n}_j\}$$

אזי

$$\mathrm{E}[x_{11}(x_{11}-1)|\underline{m},\underline{n}] = \sum_{\underline{x}\in\Gamma} x_{11}(x_{11}-1)P_N(\underline{x}|\underline{m},\underline{n}) = \frac{n_1(n_1-1)m_1(m_1-1)}{N(N-1)}$$

כי

$$\sum_{\underline{\tilde{x}}\in\tilde{\Gamma}} P_{\tilde{N}}(\underline{\tilde{x}}|\underline{\tilde{m}},\underline{\tilde{n}}) = 1$$

לכן

$$\mathrm{E}[x_{11}^2|\underline{m},\underline{n}] = \frac{n_1(n_1-1)m_1(m_1-1)}{N(N-1)} + \frac{n_1m_1}{N} = \frac{n_1m_1}{N}\left[\frac{(n_1-1)(m_1-1)+(N-1)}{N-1}\right]$$

מכיוון ש

$$\mathrm{E}[x_{11}^2|\underline{m},\underline{n}] = \mathrm{E}[x_{11}(x_{11}-1)|\underline{m},\underline{n}] + \mathrm{E}[x_{11}|\underline{m},\underline{n}]$$

ולכן

$$\mathrm{V}(x_{11}|\underline{m},\underline{n})=\frac{n_1m_1}{N(N-1)}[(n_1-1)(m_1-1)+(N-1)]-\frac{n_1^2m_1^2}{N^2}=\frac{n_1m_1}{N^2(N-1)}[(N-n_1)(N-m_1)]$$

באופן דומה

$$\mathrm{E}[x_{11}x_{22}|\underline{m},\underline{n}]=\sum_{\underline{x}\in\Gamma}x_{11}x_{22}P_N(\underline{x}|\underline{m},\underline{n})=\frac{n_1m_1n_2m_2}{N(N-1)}\sum_{\underline{x}\in\Gamma}\frac{(n_1-1)!(n_2-1)!(m_1-1)!(m_2-1)!n_3!m_3!}{(N-2)!(x_{11}-1)!(x_{22}-1)!x_{12}!\ldots x_{33}!}$$

$$=\frac{n_1m_1n_2m_2}{N(N-1)}\sum_{\underline{x}\in\Gamma^*}P_{N-2}(\underline{x}^*|\underline{m}^*,\underline{n}^*)$$

כאשר

$$\underline{x}^*=(x_{11}-1,x_{12},x_{13},x_{22}-1,\ldots,x_{33})$$

$$\underline{n}^*=(n_1-1,n_2,n_3)\qquad \underline{m}^*=(m_1-1,m_2,m_3)$$

$$\Gamma^*=\{\underline{x}^*:\sum_j x^*_{ij}=m^*_i;\sum_i x^*_{ij}=n^*_j\}$$

לכן

$$\mathrm{E}[x_{11}x_{22}|\underline{m},\underline{n}]=\frac{n_1m_1n_2m_2}{N(N-1)}$$

ואז

$$Cov[x_{11},x_{22}|\underline{m},\underline{n}]=\frac{n_1m_1n_2m_2}{N(N-1)}-\frac{n_1m_1n_2m_2}{N^2}=\frac{n_1m_1n_2m_2}{N^2(N-1)}$$

ובאופן כללי עבור לוח $n\times n$ נקבל

$$Cov[x_{kk},x_{ll}|\underline{m},\underline{n}]=\frac{m_kn_k(N\delta_k^l-m_l)(N\delta_k^l-n_l)}{N^2(N-1)}$$

עבור $1\le k,l\le n$.