

היבחנויות חוזרות במכפ"ל

דביר קלפר, אליוט טורוול



הכינוס ה-13 של אפי
י"ג בשבט תשע"ז
9.2.2017



מבנה המצגת

- מטרת המחקר
- בסיס הנתונים
- משתני המחקר
- השלמות
- סטטיסטיקה תיאורית
- מודל צמיחת הציון בבחינות חוזרות
- מודל רגרסיה לוגיסטית לחיזוי הסיכוי לבחינה חוזרת



מטרות המחקר

- המטרה המרכזית של מחקר זה היא לענות על שתי השאלות הבאות:
 - ❖ מהו השינוי הממוצע בציון הבחינה הפסיכומטרית עבור אנשים שנבחרים מספר פעמים, ומהם הגורמים שמשפיעים עליו?
 - ❖ האם ניתן לנבא את הסיכוי שנבחן יבחן יותר מפעם אחת בבחינה הפסיכומטרית בעזרת משתני רקע? ואם כן, באיזה כיוון הם משפיעים?



בסיס הנתונים

- בסיס הנתונים המקורי מכיל את כל הנבחרים בבחינה הפסיכומטרית בשנים 2001-2010 ואשר נבחנו **רק*** בשפה העברית (סה"כ 467,873 רשומות)
 - ❖ טווח זמן זה מאפשר לנבחרים להיבחן כמה פעמים, ובנוסף מאפשר לדעת גם עבור נבחני סוף התקופה האם הוא נבחן יותר מפעם אחת (על-ידי הצצה "לעתיד")
 - ❖ מבסיס הנתונים הוחלט להוציא מהניתוח נבחרים שנבחנו יותר מחמש פעמים (ששיעורם נמוך מ- 1% מהנבחרים), וזאת משום שעבורם התקבל דפוס שונה



משתני המחקר

• כל היבחנות נחשבת כרשומה והיא מכילה את השדות הבאים:

- ❖ מספר היבחנות נוכחי + מספר היבחנות מקסימלי
- ❖ מין הנבחן (1=זכר, 2=נקבה)
- ❖ גיל נוכחי + גיל בהיבחנות ראשונה
- ❖ רקע דמוגרפי: השכלת אב (1=לא למד... 7=תואר שני ומעלה), השכלת אם, מצב סוציאקונומי (1=גבוה בהרבה מהמוצע... 6=נמוך בהרבה מהמוצע)
- ❖ ציון בגרות
- ❖ ציון בחינה נוכחי + ציון בחינה ראשונה



חסרים

- המשתנים מין וגיל הכילו אחוזים בודדים של נתונים חסרים (כ-2% לגיל ו-6% למין)
 - ❖ כאשר היה אפשר הושלמו נתונים חסרים בעזרת נתוני ההיבחניות האחרות
 - ❖ במידה שהתגלתה סתירה נבחר המידע הראשון שקיים
 - ❖ רשומה שלא כללה מין ו/או גיל הוצאה מהמדגם, כך שבסופו של דבר עבור משתנים אלו אין נתונים חסרים
- עבור השדות של מספר היבחנות (נוכחי + מקסימלי), ציון בבחינה (נוכחי + ראשון) וזמן מהבחינה הקודמת אין נתונים חסרים



השלמות בסיסי

- עבור שלושת השדות הדמוגרפיים: השכלת אב, השכלת אם, מצב סוציאקונומי והשדה ציון בגרות חסרים הרבה נתונים (ראה בהמשך סטטיסטיקה)
- כמו עבור מין וגיל בוצע ניסיון ראשוני להשלים נתונים לפי היבחניות נוספות



סטטיסטיקה על נתונים חסרים

- גם לאחר ההשלמה הבסיסית רבות מהרשומות לא היו מלאות היות ש 3 השדות הדמוגרפיים והשדה בגרות היו חסרים מאוד
- להלן הסטטיסטיקה של החסרים לפי בנאדם (או*)

אחוז רשומות חסר	המשתנים בשלשות וכולם	אחוז רשומות חסר	המשתנים בזוגות	אחוז רשומות חסר	המשתנים כבודדים
19.31%	אב+אם+סוציו	13.89%	אב+אם	13.56%	השכלת אב
42.70%	אב+אם+בגרות	19.09%	אב+סוציו	13.01%	השכלת אם
44.40%	אב+סוציו+בגרות	42.57%	אב+בגרות	16.96%	מצב סוציואקונומי
44.15%	אם+סוציו+בגרות	18.62%	אם+סוציו	41.65%	ציון בגרות
		42.30%	אם+בגרות		
44.51%	אב+אם+סוציו+בגרות	43.88%	סוציו+בגרות		

סטטיסטיקה על חסרים – מסקנות



- ניתן לראות בשקף הקודם שקיימת חפיפה די גדולה בין החסרים לגבי 3 המשתנים הדמוגרפיים
- המשתנה ציון בגרות חסר מאוד ובדרך כלל כשהוא חסר אז גם אין מידע על 3 המשתנים הדמוגרפיים
- בהתאם לכך, הוחלט לוותר על המשתנה ציון בגרות במחקר זה



המשתנים להשלמות מרובות

- בוצעו השלמות מרובות (תיאור השיטה מיד) ל-3 המשתנים הדמוגרפיים לפי המשתנים הבאים:
 - ❖ מספר היבחנות נוכחי + מספר היבחנות מקסימלי
 - ❖ מין הנבחן
 - ❖ גיל נוכחי
 - ❖ ציון בחינה נוכחי
- ההשלמות בוצעו ברמת הנבחן כאשר נבחרה הרשומה הראשונה עבור כל נבחן



השלמות בשיטת השלמות מרובות

- השלמות מרובות (multiple imputation) הינה שיטה מתוחכמת להשלמת נתונים חסרים
- המודל המקובל הוא המודל הנורמלי הרב מימדי שמניח כי:
 - ❖ לכל המשתנים יש התפלגות נורמלית
 - ❖ כל משתנה יכול להיות מיוצג כפונקציה לינארית של שאר המשתנים עם טעות הומוסקדסטית נורמלית
- למרות התנאים החזקים שהמודל מניח, בפועל המודל עושה עבודה טובה גם אם התנאים מופרים



סטטיסטיקה תיאורית – כמות נבחנים

- סה"כ רשומות (בחילות) בבסיס הנתונים: 436,784
- מתוכם אנשים שונים: 336,193

ספירה של אנשים

אחוזים	מספר אנשים	מספר היבחות מקסימלית
61.46	206,618	1
29.77	100,096	2
7.17	24,119	3
1.59	5,360	4

ספירה של רשומות

אחוזים	מספר רשומות	מספר היבחות
69.36	302,954	1
24.65	107,687	2
5.14	22,436	3
0.85	3,707	4

סטטיסטיקה תיאורית – זמן בין בחינות



בחינות	מתבסס על מספר רשומות	ממוצע הזמן בשנים	סטיית התקן של הזמן בשנים	חציון הזמן בשנים
1→2	107,687	2.45	2.24	1.67
2→3	22,436	1.51	1.88	0.87
3→4	3,707	1.20	1.69	0.70

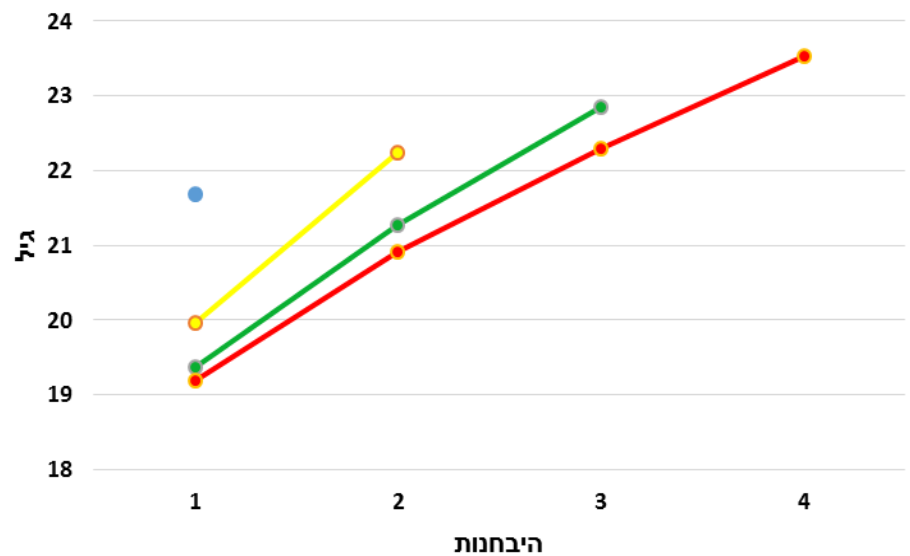
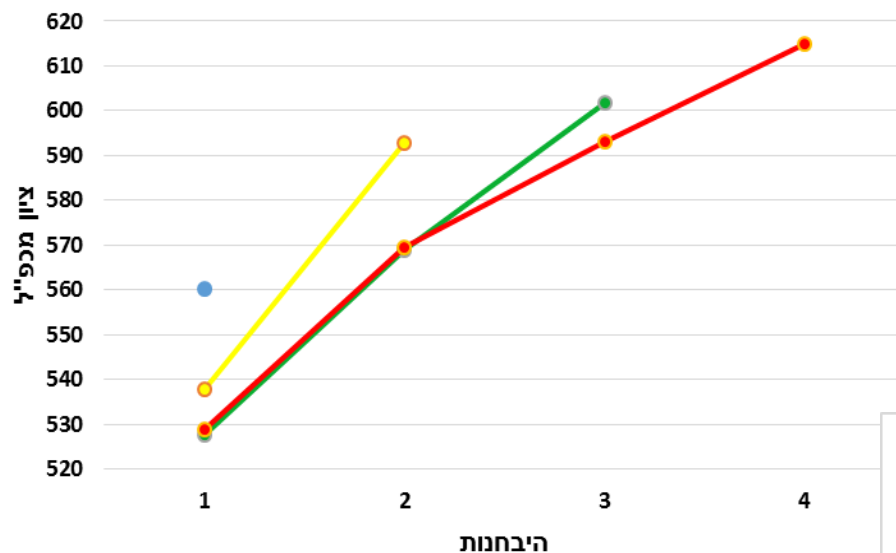
- היות שיש לנו מידע עד סוף 2015 נוכל לדעת די טוב גם עבור נבחני סוף התקופה אם בנאדם נבחן לפחות פעמיים, מה שמצדיק בדיעבד את בחירת השנים

סטטיסטיקה תיאורית – ציון הבחינה + גיל + מין



אחוז נשים	גיל		בחינה		מספר רשומות	מספר היבחנות	מספר היבחנות מקסימלי
	סטיית תקן	ממוצע	סטיית תקן	ממוצע			
55.32%	3.44	21.68	106.29	560.30	206,618	1	1
57.70%	2.51	19.97	87.21	537.79	75,003	1	2
56.20%	2.73	22.24	92.96	592.71	84,905	2	
57.60%	2.30	19.38	85.51	527.62	17,421	1	3
56.22%	2.42	21.27	85.73	568.99	18,660	2	
56.58%	2.81	22.86	89.73	601.75	18,477	3	
55.65%	2.26	19.19	90.99	528.84	3,912	1	4
55.34%	2.64	20.92	92.23	569.42	4,122	2	
55.04%	2.71	22.30	92.76	593.17	3,959	3	
54.98%	3.17	23.54	93.32	614.76	3,707	4	
56.08%	3.17	21.43	100.11	564.12	436,784	סה"כ	

סטטיסטיקה תיאורית בגרפים 1



- מקסימום 1 היבחנות
- מקסימום 2 היבחנות
- מקסימום 3 היבחנות
- מקסימום 4 היבחנות



סטטיסטיקה תיאורית – השכלת אב

אחרי השלמות			לפני השלמות			מספר היבחנות	מספר היבחנות מקסימלי
ממוצע	מספר רשומות	סטיית תקן	ממוצע	מספר רשומות	סטיית תקן		
4.78	206,618	1.49	4.84	168,704	1.53	1	1
4.95	75,003	1.48	4.97	69,586	1.50	1	2
4.90	84,905	1.49	4.91	80,134	1.50	2	
5.06	17,421	1.48	5.07	16,616	1.49	1	3
5.02	18,660	1.48	5.03	18,007	1.49	2	
4.98	18,477	1.49	4.99	17,932	1.50	3	
5.17	3,912	1.48	5.19	3,730	1.49	1	4
5.17	4,122	1.49	5.18	3,982	1.50	2	
5.13	3,959	1.50	5.14	3,849	1.50	3	
5.10	3,707	1.51	5.10	3,624	1.51	4	
4.88	436,784	1.49	4.91	386,164	1.51	סה"כ	



סטטיסטיקה תיאורית – השכלת אם

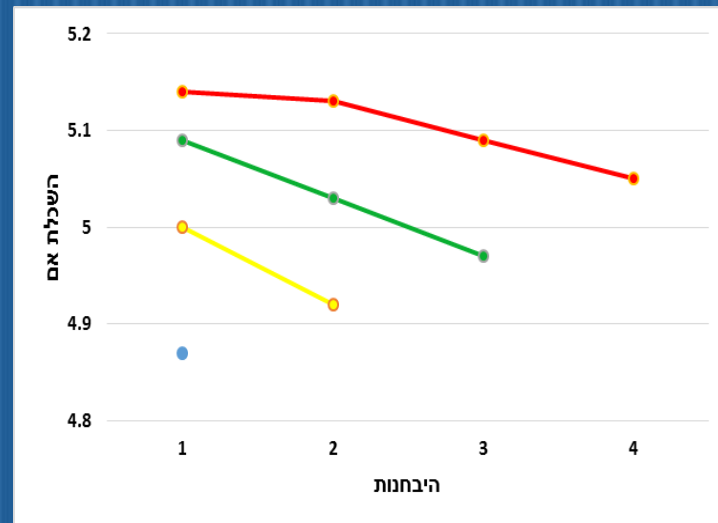
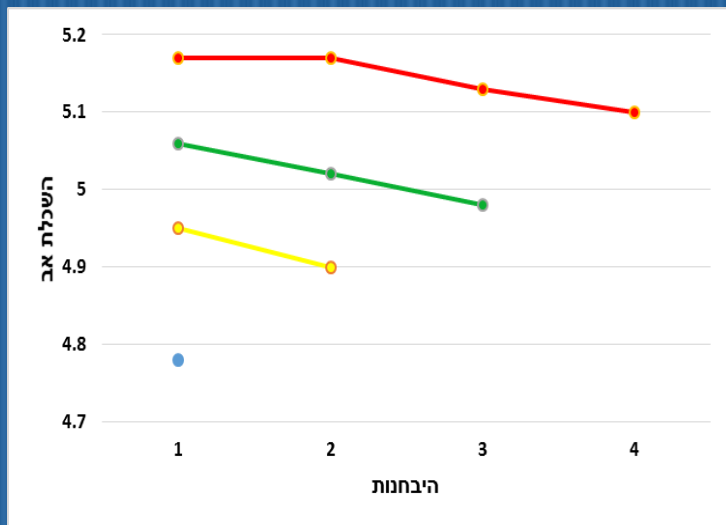
אחרי השלמות			לפני השלמות			מספר היבחנות	מספר היבחנות מקסימלי
סטיית תקן	ממוצע	מספר רשומות	סטיית תקן	ממוצע	מספר רשומות		
1.43	4.87	206,618	1.47	4.92	170,106	1	1
1.44	5.00	75,003	1.46	5.02	69,888	1	2
1.44	4.92	84,905	1.45	4.94	80,413	2	
1.46	5.09	17,421	1.47	5.10	16,675	1	3
1.46	5.03	18,660	1.47	5.04	18,067	2	
1.48	4.97	18,477	1.48	4.98	17,988	3	
1.50	5.14	3,912	1.51	5.16	3,738	1	4
1.50	5.13	4,122	1.51	5.14	3,991	2	
1.51	5.09	3,959	1.52	5.10	3,859	3	
1.52	5.05	3,707	1.52	5.06	3,628	4	
1.44	4.93	436,784	1.47	4.97	388,353	סה"כ	

סטטיסטיקה תיאורית – מצב סוציאקונומי

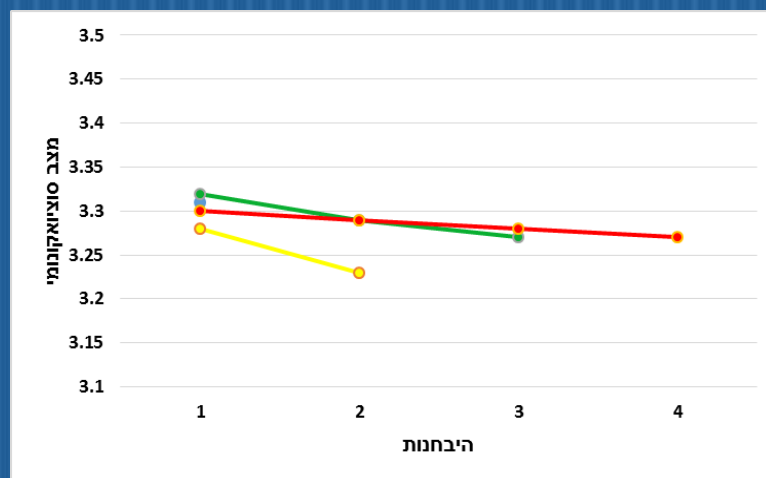


אחרי השלמות			לפני השלמות			מספר היבחנות	מספר היבחנות מקסימלי
סטיית תקן	ממוצע	מספר רשומות	סטיית תקן	ממוצע	מספר רשומות		
1.07	3.31	206,618	1.09	3.30	160,241	1	1
1.07	3.28	75,003	1.08	3.28	67,521	1	2
1.06	3.23	84,905	1.06	3.23	78,096	2	
1.09	3.32	17,421	1.10	3.32	16,304	1	3
1.08	3.29	18,660	1.08	3.29	17,697	2	
1.07	3.27	18,477	1.07	3.27	17,662	3	
1.09	3.31	3,912	1.09	3.30	3,689	1	4
1.09	3.29	4,122	1.09	3.29	3,945	2	
1.09	3.28	3,959	1.09	3.28	3,812	3	
1.07	3.27	3,707	1.08	3.27	3,593	4	
1.07	3.29	436,784	1.08	3.28	372,560	סה"כ	

סטטיסטיקה תיאורית בגרפים 2



- מקסימום 1 היבחנות
- מקסימום 2 היבחנות
- מקסימום 3 היבחנות
- מקסימום 4 היבחנות





מודל צמיחה הציון בבחינות חוזרות

- נפעיל מודל דו-רמתי עבור ניתוח נתוני אורך
- ניתן להתייחס אל נתוני אורך כנתונים היררכיים כאשר מדידות חוזרות משוכנות בתוך נבחן
- שימוש במודל רב-רמתי מאפשר להתמודד עם התלות הקיימת בציונים השונים של אותו נבחן, עם השונות שקיימת בין נבחנים שונים ועם העובדה שמערך בסיס הנתונים אינו מאוזן (קיימים דפוסים שונים של היבחניות בנתונים שברשותנו)

מודל דו-רמתי כמודל צמיחה



- מודלים רב-רמתיים (multilevel) למידול נתוני אורך (longitudinal data) נקראים גם מודלים של צמיחה (growth models)
- הרמה הראשונה היא המדידות החוזרות והרמה השנייה היא האנשים



מודל צמיחה – פורמולציה רמה 1

- נמדל את התהליך בהתאם למשוואה הבאה:

$$y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + e_{ij}$$

- כאשר:

– y_{ij} הוא ההתרחשות ה- i -ית (רמה ראשונה) עבור המדידה של אדם j (רמה שנייה)

– β_{0j}, β_{1j} הם חותך אקראי ושיפוע אקראי בהתאמה

המשמעות של האקראיות היא שהערך ההתחלתי וקצב השינוי שונים עבור אנשים שונים

– t_{ij} הוא הזמן

– e_{ij} הוא טעות ומניחים שהוא מתפלג נורמלית עם תוחלת

0 ושונות σ^2



מודל צמיחה – פורמולציה רמה 2

• רמה ראשונה: $y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + e_{ij}$

• רמה שנייה: $\beta_{0j} = \beta_{00} + \beta_{01}x_j + u_{0j}$

$$\beta_{1j} = \beta_{10} + \beta_{11}x_j + u_{1j}$$

• כאשר:

– β_{00}, β_{10} הם החותך הממוצע הכולל והשיפוע הכולל (בהתאמה) מותנים בחזאי x_j (מרמה 2)

– β_{01}, β_{11} הם מקדמי אפקטים קבועים עבור החזאי x_j

– u_{0j}, u_{1j} הם מקדמי אפקטים אקראיים לאדם j על מקדמי החותך והשיפוע מרמה 1 בהתאמה. מניחים שהם מתפלגים נורמלית עם תוחלת 0 ושונות $\sigma_{u_0}^2, \sigma_{u_1}^2$ בהתאמה



מודל צמיחה – פורמולציה אחידה

- אם נציב את המודלים מרמה 2 לתוך רמה 1 נקבל:

$$y_{ij} = (\beta_{00} + \beta_{01}x_j + \beta_{10}t_{ij} + \beta_{11}x_jt_{ij}) + (u_{0j} + u_{1j}t_{ij} + e_{ij})$$

רכיב קבוע של המודל

רכיב אקראי של המודל

טעויות
מרמה 2

השארית
מרמה 1

- באופן כללי נוכל לרשום בכתיב מטריצות:

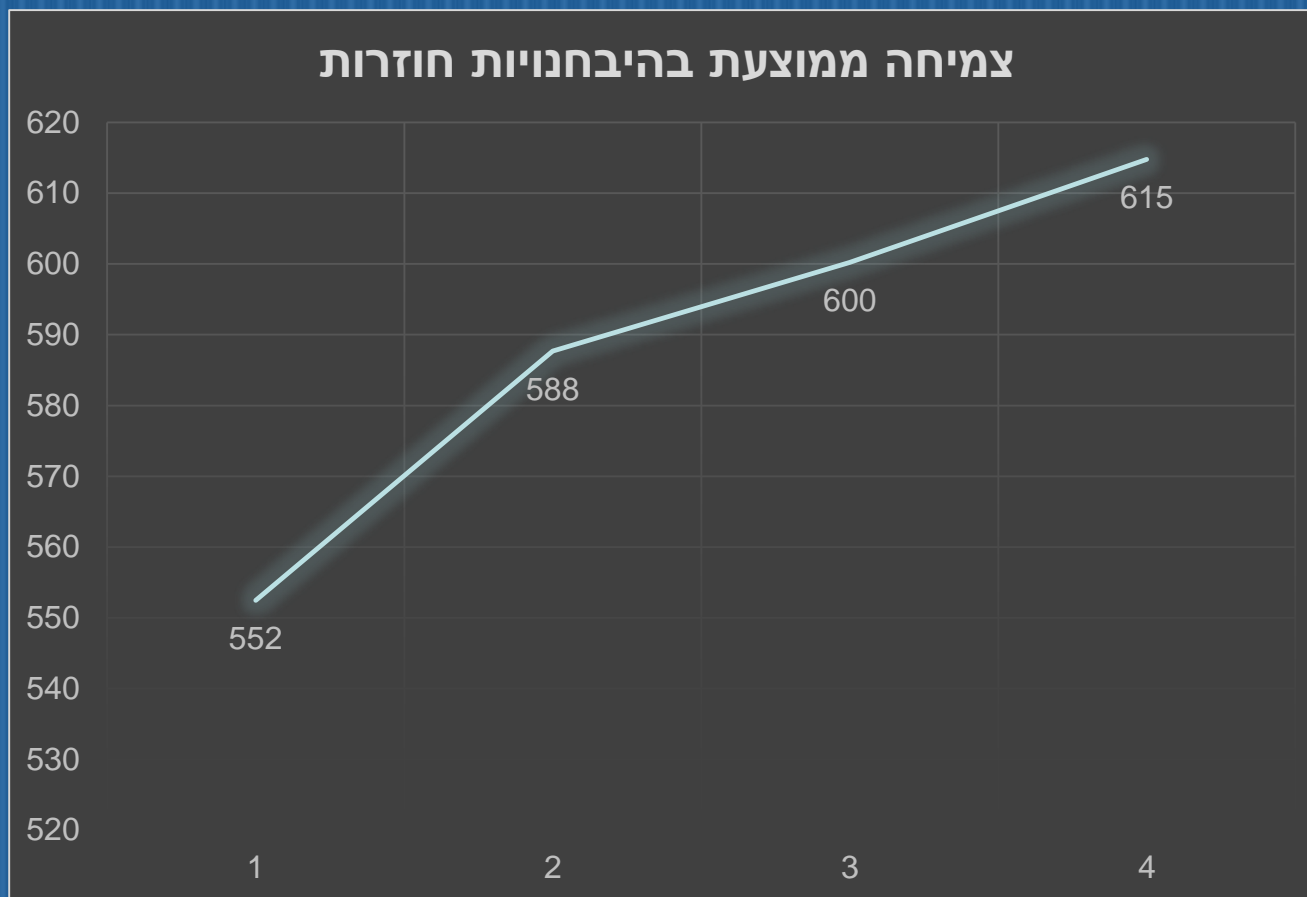
$$Y = X\beta + ZU + e$$

- כאשר: Y מסמל וקטור של תוצאות, β וקטור לא ידוע של אפקטים קבועים עם מטריצת עיצוב X ידועה, ו- U וקטור לא ידוע של אפקטים אקראיים עם מטריצת תכנון (design) Z ידועה, ולבסוף e הוא וקטור שאריות

הצמיחה הממוצעת בהיבחנויות חוזרות



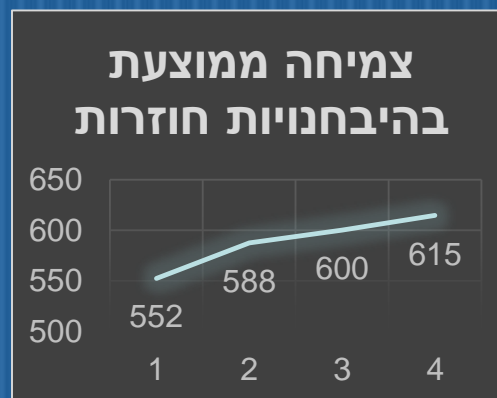
- ראשית מוצג הציון הממוצע כפונקציה של ההיבחנות





מודל צמיחה מקוטע (ספליין)

- כפי שניתן לראות בגרף, ניתן להבחין למעשה בשני חלקי צמיחה בציון
 - צמיחה גדולה בין היבחנות ראשונה לשנייה
 - צמיחה מתונה יותר (ויחסית דומה) בין היבחנות 2 ל-4
- בהתאם לכך נתאים מודל מקוטע (piecewise growth curve) המוכר גם בשם ספליין (spline)





קידוד הזמן בעקום מקוטע

- על מנת לקבל את האפקט של השבירה לאחר הבחינה השנייה נקודד את מספר ההיבחנות בעזרת שני משתנים (היבחנות 1+2) לפי הטבלה הבאה:

ציוני היבחנות			מספר היבחנות
היבחנות 2	היבחנות 1	היבחנות	
0	0	0	1
0	1	1	2
1	1	2	3
2	1	3	4

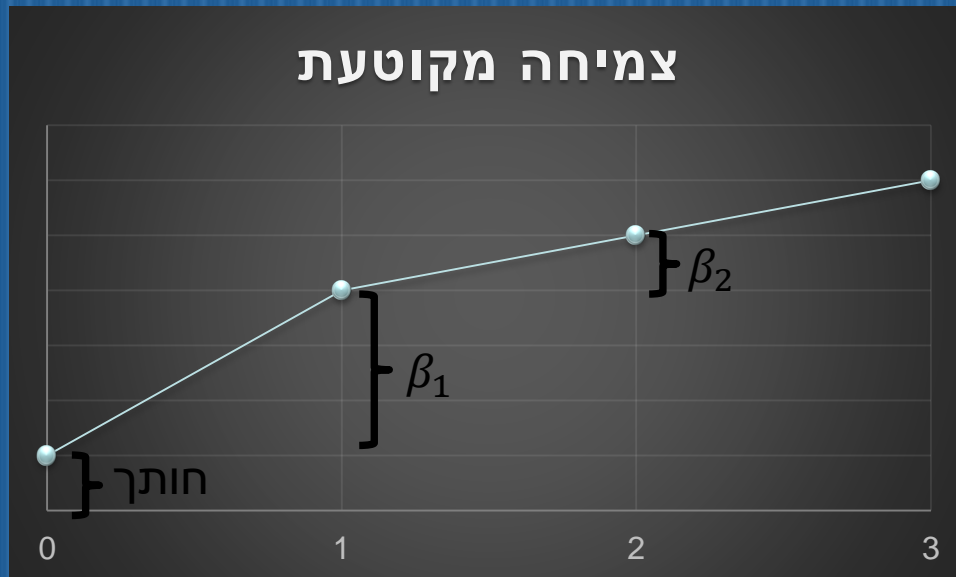
משמעות המקדמים בקידוד



- עבור המשוואה הבאה ובהתאם לקידוד שלנו נקבל:

$$\text{ציון} = \text{חותך} + \beta_1 * \text{היבחנות 1} + \beta_2 * \text{היבחנות 2}$$

- החותך (כשמו) מציין את נקודת החיתוך עם ציר ה-y



- β_1 הוא השיפוע של המקטע הראשון
- β_2 הוא השיפוע של המקטע השני

חותך ושיפועים אקראיים במודל מקוטע



- בשלב הראשון ניישם את המודל המקוטע ללא משתנים מסבירים לפי:
רמה ראשונה:

$$y_{ij} = \beta_{0j} + \beta_{1j} * 1_{ij} + \beta_{2j} * 2_{ij} + e_{ij}$$

- רמה שנייה:

$$\beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10} + u_{1j}; \beta_{2j} = \gamma_{20} + u_{2j}$$

- כאשר u_{0j}, u_{1j} ו- u_{2j} מתפלגים לפי

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} & \sigma_{u02}^2 \\ \sigma_{u01} & \sigma_{u1}^2 & \sigma_{u12}^2 \\ \sigma_{u02}^2 & \sigma_{u12}^2 & \sigma_{u2}^2 \end{pmatrix} \right]$$



תוצאות שלב ראשון

$$\text{ציון} = \gamma_{00} + \gamma_{10} * 1 \text{ היבחנות} + \gamma_{20} * 2 \text{ היבחנות} + (u_{0j} + u_{1j} * 1 \text{ היבחנות} + u_{2j} * 2 \text{ היבחנות} + e_{ij})$$

אפקט	מקדם	רכיב שונות	טעות תקן	ערך t/Z
קבוע	γ_{00}		0.1758	3,142
	γ_{10}		0.1609	281
	γ_{20}		0.2544	101
אקראי	σ_{u0}^2	9,246	29.07	318
	σ_{u1}^2	506	31.00	16.34
	σ_{u2}^2	166	20.51	8.11
	σ_{u01}^2	-187	23.46	-7.99
	σ_{u02}^2	-298	29.94	-9.95
	σ_{u12}^2	205	16.84	12.17
	$\sigma^2(e_{ij})$		861	14.57



פירוש תוצאות שלב ראשון

- $\gamma_{00}=552$ זה הממוצע בהיבחנות ראשונה
- $\gamma_{10}=45$ זה הצמיחה הממוצעת בין בחינה ראשונה לשנייה
- אומנם בגרף זה נראה כאילו זה 36, אבל צריך לזכור שאלו שנבחנו פעם אחת כנראה קיבלו ציון גבוה יותר וממוצע היבחנות ראשונה של אלו שנבחנו פעמים נמוך מ-552
- $\gamma_{20}=26$ זה הצמיחה הממוצעת בין בחינה שנייה לרביעית
- $\sigma_{u01}^2=-187$; $\sigma_{u02}^2=-298$ כלומר ככל שהציון הראשון גבוה יותר כך קצב הצמיחה בהמשך קטן יותר
- $\sigma_{u12}^2=205$ משמע שקצב הצמיחה בהתחלה מתואם חיובית עם ההמשך



שלב שני – משתני רקע במודל מקוטע

- בשלב השני נכניס את משתני הרקע
- רמה ראשונה:

$$y_{ij} = \beta_{0j} + \beta_{1j} * \text{היבחנות 1}_{ij} + \beta_{2j} * \text{היבחנות 2}_{ij} + e_{ij}$$

- רמה שנייה:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * \text{מין} + \gamma_{02} * \text{גיל ראשון} + \gamma_{03} * \text{אב} + \gamma_{04} * \text{אם} + \gamma_{05} * \text{סוציו} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * \text{מין} + \gamma_{12} * \text{גיל ראשון} + \gamma_{13} * \text{אב} + \gamma_{14} * \text{אם} + \gamma_{15} * \text{סוציו} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} * \text{מין} + \gamma_{22} * \text{גיל ראשון} + \gamma_{23} * \text{אב} + \gamma_{24} * \text{אם} + \gamma_{25} * \text{סוציו} + u_{2j}$$



תוצאות שלב שני

אפקטים אקראיים

רכיב שונות	משתנה
7,441	σ_{u0}^2
444	σ_{u1}^2
142	σ_{u2}^2
-277	σ_{u01}^2
-275	σ_{u02}^2
148	σ_{u12}^2
857	$\sigma^2(e_{ij})$

$p < 0.0001$

אפקטים קבועים

היבחנות2	היבחנות1	חותך	משתנה
73.14	135.54	564.97	
-2.40	-3.10	-35.65	מין
-2.38	-3.80	-0.34	גיל ראשון
לא מובהק	-0.55	9.77	השכלת אב
לא מובהק	-0.97	11.39	השכלת אם
0.88	-1.03	-16.29	סוציו-אקונומי

$p < 0.0001$

היות ולא מרכזנו, ערכים בשורה זו חסרי משמעות



פירוש תוצאות שלב שני – 1

- מין: הציון הראשון של נשים נמוך ב-35.7 מזה של גברים, ובנוסף קצב השיפור שלהם מהבחינה הראשונה לבחינה השנייה נמוך ב-3.1 נקודות מזה של גברים ועבור בחינות ההמשך הקצב נמוך ב-2.4 נקודות מגברים
- גיל: ככל שהבחינה הראשונה נעשית בגיל מבוגר יותר הציון נמוך יותר (0.3 לכל שנה), וכנ"ל קצב השיפור שקטן ב-3.8 לכל שנה נוספת של גיל בין בחינה ראשונה לשנייה וב-2.4 מהבחינה השנייה לרביעית



פירוש תוצאות שלב שני – 2

- השכלת אב ואם: ככל שההורים משכילים יותר הציון הראשון גבוה יותר (9.8 ו-11.4 לכל שלב עבור אב ואם בהתאמה). לעומת זאת, קצב צמיחת הציון קטן יותר ככל שההורים משכילים יותר (0.6 ו-1.0 לכל שלב עבור אב ואם בהתאמה). השכלת הורים לא משפיעה על קצב הצמיחה בין בחינה שנייה לרביעית
- מצב סוציאקונומי: ככל שהמצב הסוציאקונומי יותר טוב הציון הראשון יותר גבוה (ב-16.3 לכל שלב). בנוסף, קצב צמיחת הציון גדול יותר (ב-1.0 לכל שלב) בין ההיבחנות הראשונה לשנייה, וקטן יותר (ב-0.9 לכל שלב) מההיבחנות השנייה לרביעית



רגרסיה לוגיסטית

- כדי לענות על השאלה השנייה לגבי האפשרות לנבא סיכוי של נבחן להיבחן יותר מפעם אחת בבחינה הפסיכומטרית בעזרת משתני רקע, נבצע רגרסיה לוגיסטית כאשר משתנה המטרה הוא נבחן פעם אחת בלבד $=0$ לעומת נבחן יותר מפעם אחת $=1$ לפי המשתנים הבאים:
 - ❖ מין, גיל בבחינה הראשונה, ציון בבחינה הראשונה, השכלת אב, השכלת אם, מצב סוציואקונומי
- המודל הופעל ברמת בנאדם כלומר על בסיס נתונים שמכיל 340,095 רשומות



תוצאות רגרסיה לוגיסטית

95% רווח בר-סמך ליחס הסיכויים			אומדן למקדם= β	משתנים
גבוה	יחס הסיכויים= e^β	נמוך		
			8.479*	חותך
0.725	0.714	0.703	-0.337*	מין
0.734	0.731	0.729	-0.313*	גיל בבחינה הראשונה
0.996	0.996	0.996	-0.004*	ציון בבחינה הראשונה
1.049	1.041	1.033	0.040*	השכלת אב
1.017	1.010	1.003	0.010*	השכלת אם
0.938	0.930	0.923	-0.072*	מצב סוציאקונומי

* $p < 0.001$, $c = 0.734$



פרשנות רגרסיה לוגיסטית

- היות ומין מקודד כזכר=1 ונקבה=2, נשים נוטות להיבחן שוב פחות מגברים
- ככל שהגיל שלך בבחינה הראשונה גדול יותר כך קטנים הסיכויים שתבחן לפחות פעמיים
- ככל שהציון בבחינה הראשונה שלך גבוה יותר כך קטנים הסיכויים שתבחן לפחות פעמיים
- משתנים דמוגרפיים: אנשים עם הורים משכילים (הן אב והן אם) ואנשים עם מצב סוציאקונומי גבוה יותר נוטים להיבחן יותר מפעם אחת

תודה



שאלות?