

בקרת איכות דו-רמתית על ציונים לפי משתנים דמוגרפיים

דביר קלפר, אבי אללוף, אליוט טורוול,
כרמל אורן ומרינה פרונטון

כנס האגודה הישראלית לפסיכומטריקה
פברואר 2015



מבנה המצגת

❖ **מבוא** - מטרת המחקר וחשיבותו, הבחינה הפסיכומטרית

❖ **שיטה**

❖ תיאור המדגם

❖ משתני הרקע וסטטיסטיקה תיאורית

❖ 3 מודלים רב-רמתיים

- ניתוח שונות עם אפקטים מקריים
- רגרסיה של הממוצעים
- מודל רגרסיה עם מקדמים מקריים

❖ **תוצאות**

❖ **תיקוף**

❖ **סיכום**

למחקר שני היבטים:

- שיפור בקרת איכות
- פיתוח ושימוש במודלים סטטיסטיים...

מטרת המחקר וחשיבותו [1]

- ❖ ידוע שיש **קשר** בין נתונים דמוגרפיים (מין, גיל, השכלה...)
להישגים בבחינות בכלל, וכמובן שגם בבחינה הפסיכומטרית
- ❖ קשר זה מאפשר לנסות **ולנבא** ציון בבחינה הפסיכומטרית של
נבחן בודד. במחקר אפשר לאמוד את גודלה של טעות הניבוי
- ❖ צפוי כי ניבוי ציוני **קבוצה** (למשל נבחני נוסח מסוים במועד
בחינה) ובפרט ערכים מרכזיים כמו ממוצע וחציון, יהיה מדויק
יותר מניבוי ציוני נבחן בודד וטעות הניבוי תהיה קטנה יותר
- ❖ לניבוי ציוני קבוצה יש חשיבות רבה לצרכי **בקרת איכות** על
ציונים



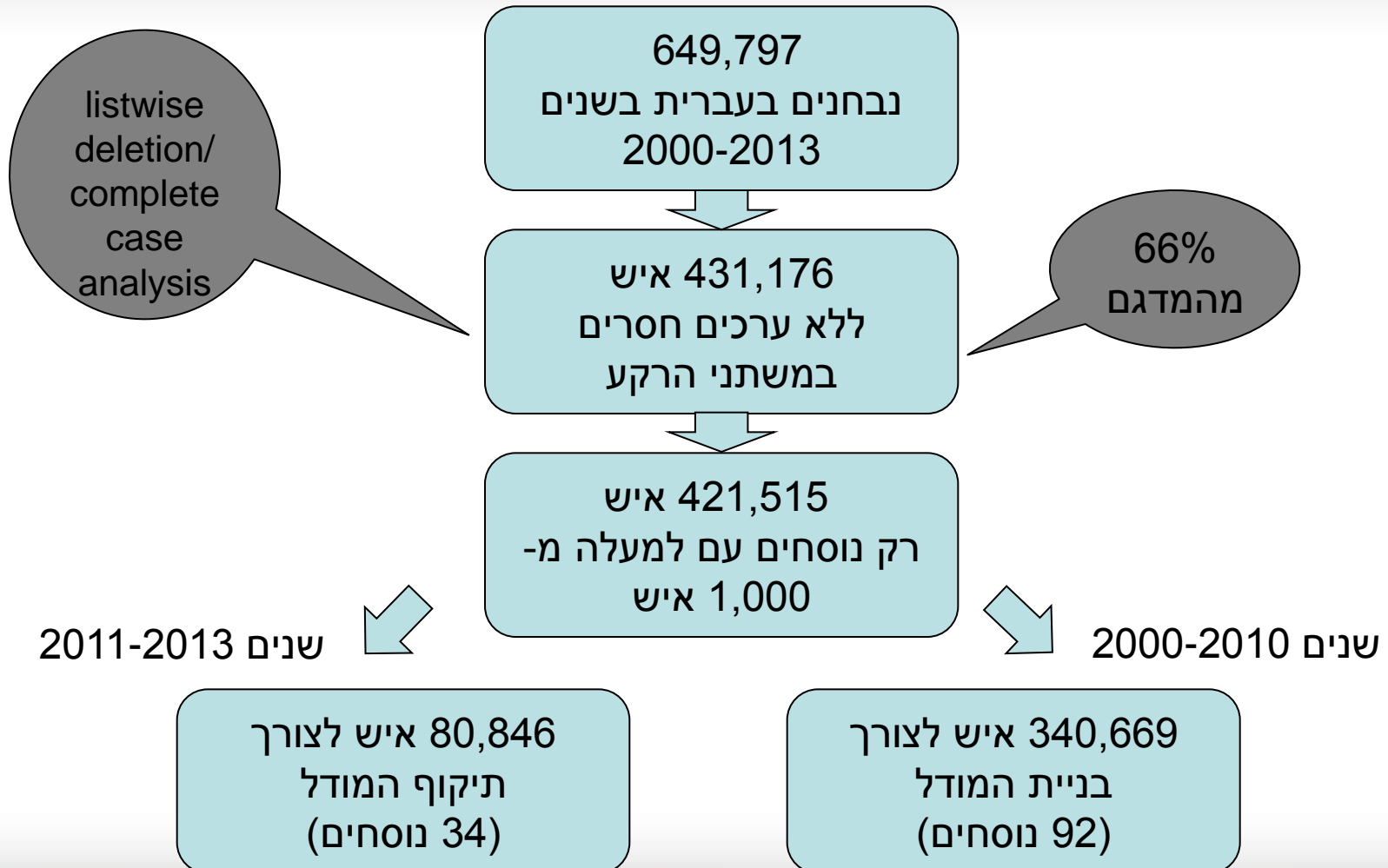
מטרת המחקר וחשיבותו [2]

- ❖ אחת השיטות המובהקות של בקרת איכות על ציונים היא להשוות ציונים מחושבים לציונים **מצופים** על בסיס משתני רקע ידועים [הנפוץ ביותר – ציוני מועד מקביל בשנה שעברה]
- ❖ כלי המאפשר לחזות במדויק ציונים (ולקבל "ציונים מצופים") יכול לסייע בבקרת איכות, ולהדליק "נורה **אדומה**" במקרה של פער גדול בין ציונים מחושבים לציונים מצופים
- ❖ **הבחינה הפסיכומטרית** היא כלי לחיזוי סיכויי ההצלחה בלימודים במוסדות להשכלה גבוהה, והיא משמשת במיון המועמדים
- ❖ בבחינה ניתן ציון לכל חלק: כמותי, מילולי ואנגלית. את החלקים ניתן לשקלל לציון כללי. במחקר נתמקד בחיזוי הציון הכללי **TOT** שבו לציון המילולי והכמותי משקל כפול מזה של האנגלית

המדגם

- המדגם מכיל את כל הנבחרים בבחינה הפסיכומטרית אשר:
 - נבחנו בשנים 2000-2013
 - נבחנו בשפה העברית
 - לא הכילו חסרים במשתני הרקע (ראה בהמשך)
 - בנוסח שלהם השתתפו כ-1,000 איש לפחות
- המדגם פוצל לשנים
 - חלק ראשון לצורך בניית המודל
 - חלק שני לצורך תיקוף המודל

המדגם במספרים



המשתנה התלוי

- אנו ננבא את הציון הכללי בבחינה הפסיכומטרית

מקסימום	מינימום	סטיית תקן	ממוצע	N	משתנה	רמה
800	227	100.5	565.9	340,669	הציון הכולל	רמה 1: רמת הנבחן
586.4	532.4	10.3	565.2	92	ממוצע רמה 1	רמה 2: רמת הנוסח
108.3	93.3	3.1	100.2	92	סטיית תקן רמה 1	
6,655	1,132	1,423.8	3,702.9	92	גודל המדגם	

המשתנים המנבאים

• אנו משתמשים ב-6 משתנים מנבאים:

– מין

– גיל

– רמת הכנסה של המשפחה

– השכלת אב

– השכלת אם

– מספר היבחנות

• המשתנה מין הינו דיכוטומי, גיל הינו רציף (מנה)

ולשאר המשתנים נתייחס כמשתנה רוח

משתנה מנבא - מין

• מין: 0=זכר, 1=נקבה

מקסימום ציון כולל	מינימום ציון כולל	סטיית תקן	ממוצע ציון כולל	אחוזים	N	קטגוריה	רמה
800	227	96.4	587.6	45	152,163	זכר	רמה 1:
800	231	100.4	548.4	55	188,506	נקבה	רמת הנבחן
608.7	556.8	10.6	587.2		92	זכר	רמה 2:
571.2	511.3	11.9	547.1		92	נקבה	רמת הנוסח

משתנה מנבא – מצב כלכלי

מקסימום ציון כולל	מינימום ציון כולל	סטיית תקן	ממוצע ציון כולל	אחוזים	N	קטגוריה	רמה
800	244	97.4	608.4	2.5	8,680	1=גבוה בהרבה	רמה 1: רמת הנבחן
800	227	92.6	606.1	19.8	67,601	2=גבוה	
800	231	95.5	572.9	42.3	144,191	3=גבוה במעט	
800	232	97.5	538.5	22.0	74,922	4=נמוך במעט	
793	279	100.6	525.1	9.9	33,881	5=נמוך	
800	251	105.8	508.2	3.3	11,394	6=נמוך בהרבה	
645.4	523.5	18.8	613.4		92	1=גבוה בהרבה	רמה 2: רמת הנוסח
625.1	562.3	10.8	604.9		92	2=גבוה	
592.0	545.1	10.7	572.0		92	3=גבוה במעט	
558.6	513.5	9.2	537.9		92	4=נמוך במעט	
555.8	499.3	10.1	524.5		92	5=נמוך	
540.1	471.3	15.2	508.7		92	6=נמוך בהרבה	

משתנה מנבא – השכלת אב

מקסימום ציון כולל	מינימום ציון כולל	סטיית תקן	ממוצע ציון כולל	אחוזים	N	קטגוריה	רמה
783	251	101.1	448.3	1.6	5,497	1=לא למד	רמה 1: רמת הנבחן
781	239	99.8	513.5	3.8	12,794	2=יסודית	
797	279	96.7	537.5	12.6	43,036	3=תיכונית חלקית	
800	227	96.8	543.9	25.8	88,000	4=תיכונית מלאה	
800	231	95.4	566.8	17.5	59,501	5=על תיכונית	
800	232	95.0	593.1	19.1	65,076	6=תואר ראשון	
800	232	91.7	605.8	19.6	66,765	7=תואר שני ומעלה	
525.8	404.2	25.0	453.2		92	1=לא למד	רמה 2: רמת הנוסח
543.9	481.5	14.4	513.1		92	2=יסודית	
568.9	510.6	10.9	537.8		92	3=תיכונית חלקית	
568.7	512.7	10.7	543.0		92	4=תיכונית מלאה	
588.9	532.6	10.1	566.0		92	5=על תיכונית	
609.5	563.7	10.3	591.9		92	6=תואר ראשון	
627.4	579.7	10.4	605.4		92	7=תואר שני ומעלה	

משתנה מנבא – השכלת אם

מקסימום ציון כולל	מינימום ציון כולל	סטיית תקן	ממוצע ציון כולל	אחוזים	N	קטגוריה	רמה
786	251	110.7	463.8	2.0	6,718	1=לא למד	רמה 1: רמת הניבחן
783	239	99.5	501.7	2.7	9,035	2=יסודית	
797	273	96.6	530.5	9.5	32,275	3=תיכונית חלקית	
795	232	96.5	545.4	28.2	95,954	4=תיכונית מלאה	
800	252	96.2	567.8	17.9	60,860	5=על תיכונית	
800	231	94.8	590.4	21.7	74,007	6=תואר ראשון	
800	227	91.6	605.5	18.1	61,820	7=תואר שני ומעלה	
570.2	391.4	27.5	447.7		92	1=לא למד	רמה 2: רמת הנוסח
535.9	470.7	14.7	500.3		92	2=יסודית	
552.6	497.2	11.0	530.2		92	3=תיכונית חלקית	
562.6	514.9	10.4	544.5		92	4=תיכונית מלאה	
588.2	532.4	11.2	566.2		92	5=על תיכונית	
609.4	552.7	10.3	589.4		92	6=תואר ראשון	
629.8	579.8	10.1	605.2		92	7=תואר שני ומעלה	

משתנה מנבא – מספר היבחנות

מקסימום ציון כולל	מינימום ציון כולל	סטיית תקן	ממוצע ציון כולל	אחוזים	N	קטגוריה	רמה
800	231	101.4	554.6	68.9	234,891	היבחנות 1	רמה 1: רמת הניבחן
800	227	93.0	588.0	24.2	82,548	היבחנות 2	
800	286	93.4	597.9	5.2	17,789	היבחנות 3	
800	284	105.0	612.8	1.6	5,441	היבחנות +4	
575.8	518.0	10.1	554.0		92	היבחנות 1	רמה 2: רמת הנוסח
610.2	553.4	11.3	587.5		92	היבחנות 2	
627.6	549.5	15.5	597.7		92	היבחנות 3	
688.1	527.9	32.9	606.4		92	היבחנות +4	

משתנים מנבאים – רמה שנייה

- אנו מכניסים כרמה שנייה את ממוצע כל אחד מן המשתנים ברמה הראשונה

מקסימום הממוצעים	מינימום הממוצעים	סטיית תקן של הממוצעים	ממוצע הממוצעים	משתנה
0.61	0.46	0.03	0.55	מין
23.05	19.03	0.98	21.52	גיל
3.49	2.85	0.09	3.27	מצב כלכלי
5.20	4.46	0.17	4.89	השכלת אב
5.27	4.31	0.20	4.96	השכלת אם
1.60	1.15	0.11	1.40	היבחנות

מודלים רב רמתיים

- בוצע ניתוח היררכי לינארי בשתי רמות כדי לחקור את הקשר של נתוני הרקע לציון הכללי מעבר לנוסחים כאשר:
 - רמה 1: הנבחן הבודד
 - רמה 2: נוסח של בחינה
- אנו נבדוק שלושה מודלים:
 - ניתוח שונות עם אפקטים מקריים (אשר ישמש כבסיס להשוואה)
 - רגרסיה של הממוצעים
 - מודל רגרסיה עם מקדמים מקריים

מודל 1: ניתוח שונות עם אפקטים מקריים

• רמה ראשונה: $Y_{ij} = \beta_{0j} + r_{ij}$

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

- Y_{ij} הוא הציון של נבחן i בנוסח j
- β_{0j} הוא הציון הממוצע של נבחנים בנוסח j
- r_{ij} הוא השארית הקשורה בנבחן i בנוסח j ומניחים שהיא מתפלגת נורמלית עם תוחלת 0 ושונות σ^2
- γ_{00} הוא הממוצע הכולל (ממוצע מעבר לנוסחים של ממוצע נוסח)
- u_{0j} הוא אפקט מקרי הקשור בנוסח j ומניחים שהוא מתפלג נורמלי עם תוחלת 0 ושונות τ_{00}

מודל 2: רגרסיה של הממוצעים

- רמה ראשונה: $Y_{ij} = \beta_{0j} + r_{ij}$
- רמה שנייה: $\beta_{0j} = \gamma_{00} + \gamma_{01}G_j + u_{0j}$
- G_j הוא חזאי משתנה רקע מרמה 2 לציון הממוצע בנוסח j
- γ_{01} הוא מקדם השיפוע ברגרסיה של החזאי G_j על β_{0j}
- γ_{00} הוא הממוצע הכולל מותנה בחזאי G_j
- u_{0j} הוא אפקט מקרי הקשור בנוסח j מותנה בחזאי G_j

מודל 3: מודל הגרסיה עם מקדמים מקריים

$$Y_{ij} = \beta_{0j} + \beta_{1j}B_{1ij} + r_{ij} \quad \bullet \text{ רמה ראשונה:}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad \bullet \text{ רמה שנייה:}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

B_{1ij} הוא חזאי משתני רקע מרמה 1 \bullet

β_{0j} ! β_{1j} הם החותך והשיפוע בגרסיה של החזאי מרמה 1 על Y_{ij} \bullet

r_{ij} הוא השארית של החזאי מרמה 1 \bullet

γ_{10} ! γ_{00} הם הממוצע הכולל והשיפוע הממוצע מעבר לנוסחים \bullet

u_{1j} ! u_{0j} הם האפקטים המקריים של החותך והשיפוע הקשורים בנוסח j ומניחים \bullet

שהם בעלי מטריצת שונות משותפת הבאה: $\begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}$

τ_{00} הוא השונות ברמה 1 של החותך, τ_{11} הוא השונות ברמה 1 של השיפוע ו τ_{10} ! \bullet

τ_{01} הם השונות המשותפת בין החותך והשיפוע ברמה 1

תוצאות

תוצאות מודל 1: ניתוח שונות עם

אפקטים מקריים

המשמעות של זה היא שהנבחן לא צריך לדאוג לאיזה מועד הוא ניגש

$$\sqrt{10,015} = 100.07$$

$$Y_{ij} = \beta_{0j} + r_{ij}; \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\sqrt{101.37} = 10.07$$

אפקט	מקדם	רכיב שונות	טעות תקן	ערך t/Z
קבוע	565.20	γ_{00}	1.0662	530(91)
מיקרי	$\tau_{00}(u_{0j})$	101.37	15.5470	6.52
	$\sigma^2(u_{ij})$	10,015	24.2691	412.66

הממוצע הכולל הוא 565.20 עם טעות תקן של 1.07, כלומר רווח בר סמך ברמה של 95% של ממוצע הממוצעים הוא $565.2 \pm 1.96 * 1.07 = (563.1, 567.3)$

עבור $\tau_{00} = 101.37$! $\sigma^2 = 10,015$ מתקבל $ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \frac{101}{101 + 10,015} = 0.01$

כלומר 1% בלבד מהשונות מתקבל בין הנוסחים בעוד שרוב השונות מגיע מתוך השונות בנוסחים עצמם. שונות (101) זו נמוכה יחסית אבל עדיין מובהקת

רווח בר-סמך ברמה של 95% לממוצע של נוסח הוא $565.2 \pm 1.96 * \sqrt{101} = (545.5, 584.9)$

תוצאות מודל 2: רגרסיה של הממוצעים (גרסה ראשונית)

$$Y_{ij} = \beta_{0j} + r_{ij} ; \beta_{0j} = \gamma_{00} + \gamma_{01}G_j + u_{0j}$$

היות ולא
מרכזנו, זה
חסר משמעות

אחוז שונות מוסברת	τ_{00}	γ_{01}	γ_{00}	חזאי הנתונים
20.71	80.40***	144.08***	485.96***	מין
0	101.76	0.94	544.91***	גיל
8.01	93.27***	-31.23**	667.33***	רמת הכנסה
17.26	83.89***	25.69***	439.53***	השכלת אב
17.08	84.08***	20.47***	463.70***	השכלת אם
9.84	91.42***	28.38**	525.44***	היבחנות
		68.46**		מין
		-0.88		גיל
		-40.15***		רמת הכנסה
69.02	31.40***	27.59**	394.00***	השכלת אב
		11.56		השכלת אם
		65.30***		היבחנות

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

$\frac{101.4 - 31.4}{101.4} =$

זו לא טעות!
הסקלה הפוכה

ככל שיש יותר נשים
במועד הממוצע
גבוה יותר

גיל
והשכלת
אם לא
יוצאים
מובהקים

תוצאות מודל 2: רגרסיה של הממוצעים (גרסה סופית)

$$Y_{ij} = \beta_{0j} + r_{ij} ; \beta_{0j} = \gamma_{00} + \gamma_{01}G_j + u_{0j}$$

אחוז שונות מוסברת	τ_{00}	γ_{01}	γ_{00}	חזאי רמה 2
69.02	31.73***	71.09***	332.91***	מין
		-31.82***		רמת הכנסה
		42.74***		השכלת אב
		62.95***		היבחנות

* $p < 0.05$
 ** $p < 0.01$
 *** $p < 0.001$

$$\sigma^2 = 10,015$$

נגענו רק ברמה 2 ולכן זה לא השתנה (בהשוואה למודל 1)

$$\text{רמת הכנסה} * 31.82 - \text{מין} * 71.09 + 332.91 = \text{ממוצע נוסח}$$

$$+ 42.74 * \text{השכלת אב} + 62.95 * \text{היבחנות}$$

$$\text{כלומר יש שיפור משמעותי בניבוי! } RMSE = \sqrt{31.73} = 5.6 < 10.3$$

תוצאות מודל 3: מודל רגרסיה עם

מקדמים מקריים

$$= \sqrt{\tau_{11}}$$

$$= \sqrt{\tau_{00}}$$

$$Y_{ij} = \beta_{0j} + \beta_{1j}B_{1ij} + r_{ij}; \beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10} + u_{1j}$$

כל
התוצאות
יוצאות
מובהקות
ברמה
 $p < 0.001$

אחוז שונות מוסברת	σ^2	$\gamma_{10}(SD)$	$\gamma_{00}(SD)$	חזאי רמה 1
3.95	9,619.50	-39.95(4.28)	587.15(10.36)	מין
0.41	9,974.43	-0.89(1.66)	584.85(34.51)	גיל
7.92	9,222.85	-26.32(1.92)	651.32(12.33)	רמת הכנסה
9.44	9,070.51	20.24(1.03)	466.21(9.10)	השכלת אב
9.58	9,056.02	21.42(2.48)	458.87(14.41)	השכלת אם
2.64	9,751.23	22.58(4.03)	533.59(10.20)	היבחנות
		-37.25(4.61)		מין
		-0.88(1.20)		גיל
		-16.57(1.76)		רמת הכנסה
20.25	7,987.19	8.34(1.14)	533.39(29.35)	השכלת אב
		11.19(1.39)		השכלת אם
		21.21(3.62)		היבחנות

$$\frac{10,015 - 7,987}{10,015} =$$

$$RMSE = \sqrt{7,987} = 89.37$$

תיקוף - כללי

• מודל 3

- מצליח לחזות 20% מן השונות של הציון של נבחן בודד
- טווח הציונים עבור נבחן הוא 227 עד 800 (עם סטיית תקן של 100)
- מתקבל ערך RMSE של 89.37

• מודל 2

- מצליח לחזות 69% מן השונות של הציון הממוצע בנוסח
- טווח הציונים בנוסח נע בין 532 ל 586 (עם סטיית תקן של 10.3)
- מתקבל ערך RMSE של 5.6

• לכן, ניתן לומר שהחיזוי של מודל 3 חלש ולעומתו החיזוי של מודל 2 חזק מאוד

• בהתאם לכך, נתקף רק את מודל 2

תיקוף מודל 2

• כדי לאמת את המודל נדרוש ש 95% מהתחזיות (32.2) יפלו בטווח של $1.96 \pm$ סטיות תקן (5.6) כלומר טעות של עד 11 לכל כיוון
 • אכן רק תצפית אחת (16) חורגת מטווח זה!
 • בנוסף ניתן לראות שממוצע הטעות הוא רק 0.82

כולל מטלת כתיבה

חזוי- אמיתי	אמיתי	חזוי	#
0.87	552.22	553.09	20
8.14	561.28	569.43	21
7.37	562.11	569.49	22
4.48	561.48	565.96	23
8.02	552.59	560.61	24
-4.26	559.66	555.39	25
-3.06	572.21	569.15	26
-8.89	555.11	546.23	27
-1.46	574.42	572.96	28
1.82	564.56	566.38	29
9.65	562.94	572.59	30
8.85	547.40	556.25	31
-2.25	554.10	551.85	32
1.21	571.41	572.62	33
7.72	564.25	571.97	34
0.82	565.98	566.80	ממוצע
6.60	8.24	8.32	ס.ת.
-19.46	547.40	546.23	מינימום
9.86	579.51	579.44	מקסימום

חזוי- אמיתי	אמיתי	חזוי	#
0.15	569.93	570.08	1
0.14	573.71	573.85	2
-2.35	578.68	576.34	3
7.34	560.11	567.45	4
-4.09	566.28	562.19	5
0.93	571.33	572.27	6
3.87	571.84	575.70	7
5.44	573.70	579.13	8
-7.67	565.65	557.98	9
-3.95	567.94	564.00	10
-4.21	565.19	560.97	11
-0.07	579.51	579.44	12
-6.75	574.42	567.67	13
3.13	569.44	572.57	14
8.65	570.59	579.24	15
-19.46	575.01	555.55	16
-8.63	578.52	569.89	17
9.86	561.49	571.35	18
7.25	554.26	561.51	19

סיכום

- מחקר זה השתמש באנליזה רב-רמתית כדי לחקור את הקשר בין משתני הרקע של נבחנים לבין הציון הכללי בבחינה הפסיכומטרית
- נבדק הקשר הן ברמת הנבחן והן ברמת הנוסח
 - ברמת הנבחן נמצא קשר חזק בין משתני הרקע לתוצאת המבחן והקשר היה שונה בנוסחים שונים, אבל היכולת לחזות ציון לנבחן בודד נמצאה כחלשה
 - ברמת הנוסח נמצא קשר חזק בין משתני הרקע המקובצים לממוצע הציון בנוסח ובנוסף יכולת ניבוי הממוצע היא טובה מאוד
- המודל עם יכולת הניבוי הטובה נבדק על בסיס נתונים אחר ונמצא תקף

הצעות למחקרי המשך

- משתנים מנבאים נוספים

- מועד היבחנות (או לעשות ניתוח נפרד לכל חודש)

- בגרות [יש לשים לב שעבור משתנה זה בלבד חסרים כחצי מהנתונים וביחד עם שאר המשתנים יש מידע מלא רק על כ-40% מהנבחים]

- ניבוי כל תחום לחוד וחישוב ציון כללי על סמך שלושת הניבויים

- קיימת אפשרות לנסות ולהשלים נתונים בשיטות שונות למשל: זקיפה מרובה (multiple imputation)