

ההודעה על מות הפסיכומטרי הייתה מוקדמת מדי:

תגובה ליוגב ואיילון (2000)

דוד נבון
גרשון בן-שחר
יוסף צלגוב

פברואר 2006



דוח מרכז 328
ISBN:965-502-119-X

ההודעה על מות הפסיכומטרי הייתה מוקדמת מדי:

תגובה ליוגב ואיילון (2000)

יוסף צלגוב

החוג למדעי ההתנהגות
אוניברסיטת בן גוריון

גרשון בן שחר

החוג לפסיכולוגיה
האוניברסיטה העברית

דוד נבון

החוג לפסיכולוגיה
אוניברסיטת חיפה

מהדורה מקורית:
אפריל 2001

הדפסה מחודשת:
פברואר 2006

מהפכות מדעיות קורות לפעמים, אך מוטב לא להיחפז לבשר עליהן לפני שבדקים אותן בשבע עיניים של מומחים מהתחום. נייר העמדה על המיון בכניסה לאוניברסיטאות שהופץ לאחרונה (יוגב ואיילון, 2000) מדגים מה קורה כשאינן נוהגים כך.

תוקפן של בחינות פסיכומטריות בחיזוי הצלחה בלימודים אקדמיים נמדד במאות מחקרים ע"י פסיכומטריקאים רבים בעולם (ראה למשל, Willingham, Lewis, Morgan & Ramist, 1990). גם בארץ נערכו בדיקות רבות כאלה החל משנות הששים (ראה למשל, אורן, 1992; אורת, 1965; בלד ובן-שחר, 1981; ספז, 1990; פרלברג, 1965; קיזרמן, 1987; קנת-כהן, אורן וברונר, 1994א, 1994ב; שילו ומאיר, 1993; שפירא ועציוני-הלוי, 1973; Beller, 1994). הממצא הטיפוסי הוא מתאם חיובי, הנע בד"כ בין 0.35 ל-0.45, והמאפשר ניבוי לא רע יחסית לחזאים אחרים בתחומי החינוך והפסיכולוגיה. גם כאשר משתמשים בחזאי נוסף, כגון ציוני הבגרות, תרומתו השולית של הציון הפסיכומטרי עצמו חיובית, שכן שילוב שני החזאים לציון סכום אחד מגדיל את התוקף. באשר לבחינה הפסיכומטרית הכלל-ארצית הנהוגה זה שנים בארץ (מכפ"ל), היא נבדקת באופן שוטף. ניתוח-על של 1861 בדיקות תוקף שנערכו בששה מוסדות לימוד ו-12 מחזורי לימוד (קנת-כהן, ברונר ואורן, 1999) מלמד, כי ברובם המכריע נמצאו מתאמים חיוביים ניכרים בין הציון הפסיכומטרי להצלחה בלימודים. המתאם הממוצע הוא 0.45 (כאשר תוקף ממוצע הבגרות הוא 0.41, ותוקף ציון הסכם של שני המרכיבים הוא 0.55).

נייר העמדה של יוגב ואיילון מערער על ממצא מבוסס זה. הבסיס לערעור הוא ניתוח אותו הם ביצעו לפני מספר שנים על נתוני שנתון לימודים אחד באוניברסיטת תל אביב (11 בדיקות סה"כ), ואשר דווח בנייר לדיון פנימי שהוצא באוניברסיטת תל אביב (איילון ויוגב, 1997). איילון ויוגב ניתחו נתונים אלה באמצעות מודל רגרסיה שהוצע ע"י Heckman (1979; ראה גם Berk, 1983) לטיפול במדגמים הסובלים מנתונים חסרים. על פי דיווחם, מקדמי הרגרסיה של ציוני המבחן הפסיכומטרי היו קטנים מאד ולעתים שליליים. בתגובה הראו קנת-כהן וברונר (1997), ששישום המודל ע"י איילון ויוגב הנו שגוי. בנייר העמדה שלהם מקבלים יוגב ואיילון (2000) חלק מביקורתם של קנת-כהן וברונר, אך ממשיכים לטעון כי אין לציון הפסיכומטרי תוקף ניבוי גבוה.

במאמר זה ננסה להסביר, בצורה פשוטה עד כמה שניתן, את הפגמים בניתוחם ובטיעוניהם של יוגב ואיילון.

כדי לאמוד את יכולת החיזוי של מדד פסיכומטרי על פיו ממינינים למסגרת מסוימת (תכנית לימודים, שרות צבאי, עבודה במפעל, וכיו"ב), אין ברירה בד"כ אלא להתבסס רק על נתוני ההצלחה של המועמדים שהתקבלו וסיימו, ולהקיש מכך על אוכלוסיית המועמדים כולה.

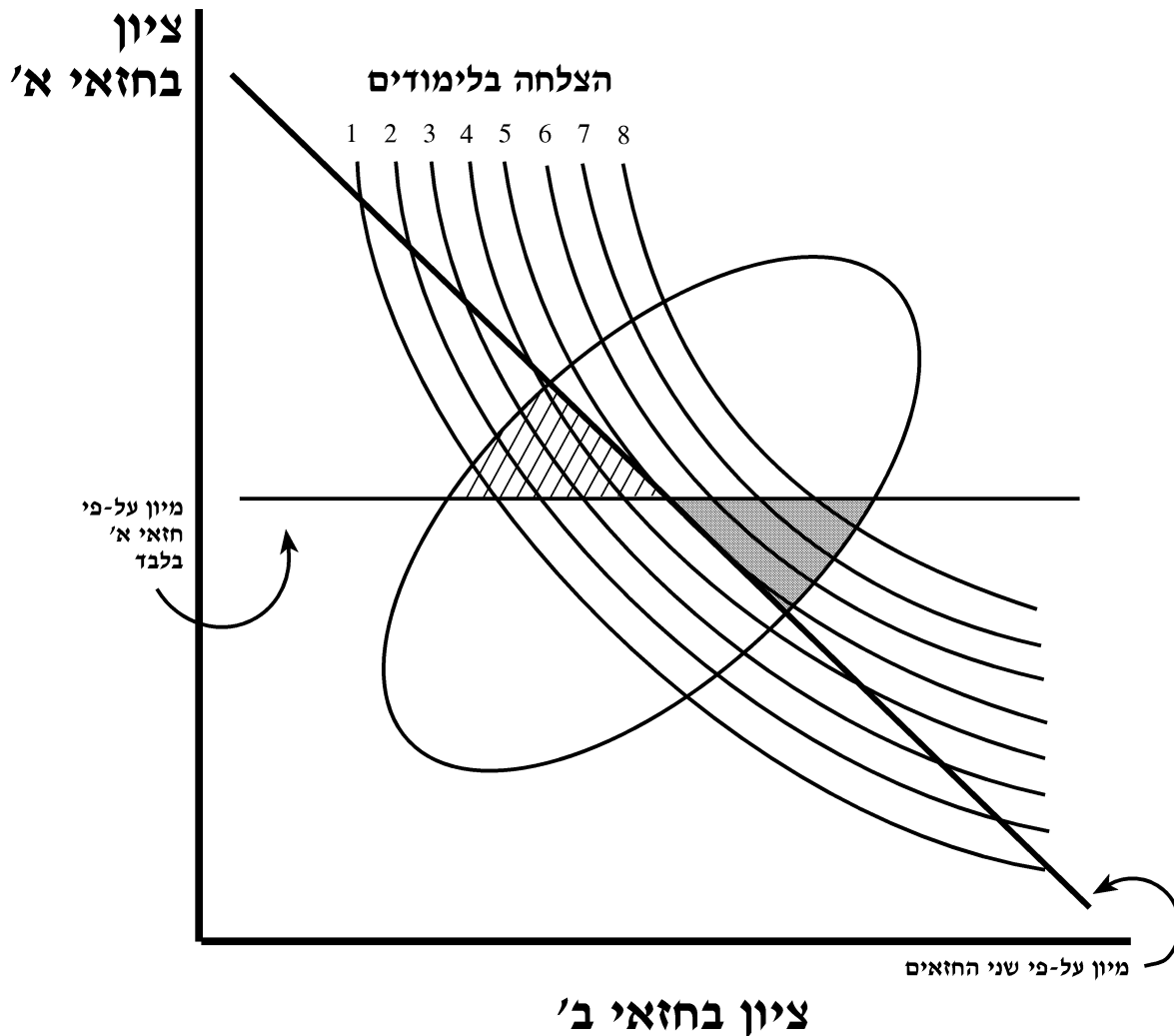
דא עקא, מקדם המתאם הנצפה בקבוצת המתקבלים¹ בין חזאי נתון (למשל, ציון בבחינת כניסה) לבין הקריטריון (למשל, הישגים בלימודים) צפוי להיות קטן ממקדם המתאם האמיתי (ראה למשל, Guilford, 1967; Lord & Novick, 1968). המקדם מתכווץ בעיקר בשל שלושה גורמים: מהימנותו המוגבלת של הקריטריון, קיצוץ תחומו של החזאי, והוספתו של חזאי נוסף (למשל, ממוצע הבגרות) למשוואת המיון. שני הגורמים הראשונים מוכרים לכל מי שלמד קורס מבוא בתורת המבחנים: הראשון קובע, כי ככל שהקריטריון פחות מהימן (כלומר, יציבותו מעבר לזמן פחותה והמתאם בין חלקיו השונים נמוך), המתאם בינו לבין כל חזאי אפשרי בהכרח נמוך יותר. הגורם השני קובע, כי מיון על פי החזאי, על פי הגדרתו, מותיר בקבוצת המתקבלים רק את בעלי הציונים הגבוהים בחזאי, והפחתה זו בשונות גוררת הפחתה בשונות המשותפת, ומכאן במתאם.

הגורם השלישי מוכר פחות, ועל כן נרחיב עליו כאן במידת מה. שילובו של חזאי נוסף (נקרא לו חזאי ב', ולחזאי המקורי – חזאי א') בחיזוי כאשר משוואת הקבלה היא צירוף לינארי, מביא לכך שבקרב המתקבלים, לבעלי ציונים נמוכים יחסית בחזאי א' יש סיכוי לא קטן להצליח בלימודים, משום שהם בהכרח בעלי ציונים גבוהים יחסית בחזאי ב'. עובדה זו יוצרת הטיה כלפי מטה של המתאם הנצפה בין חזאי א' לקריטריון, בפרט כאשר יחס הסלקציה נמוך (למשל, במסלולי לימוד יוקרתיים בהם מכסת המתקבלים מצומצמת יחסית לביקוש).

ציור מס' 1 מדגים זאת. מוצג בו היטל של התפלגות דו-משתנית של שני חזאים בעלי תוקף חיובי שביניהם מתאם ביניים חיובי (מיוצגת בציור ע"י האליפסה). הקריטריון מוצג באמצעות עקומות שוות-הצלחה (מודגמות בציור ע"י הקווים הקמורים). מיון על סמך חזאי א' בלבד (המיוצג כאן על גבי הציר האנכי) מביא כמובן לקיצוץ תחום המביא להקטנה מסוימת במתאם הנצפה בין חזאי א' להצלחה. איך ישפיע מיון על סמך ציון סכם של שני החזאים על מקדם מתאם זה? כאשר תוקפו של חזאי א' אינו עולה על זה של חזאי ב', מיון כזה גורע מקבוצת המתקבלים על סמך חזאי א' בלבד מועמדים (ראה השטח המקווקו בציור בקווים מלוכסנים) שציוניהם נמוכים (יחסית לקבוצה זו) הן בחזאי והן בקריטריון, ומאידך מוסיף מועמדים (ראה השטח המקווקו בציור בתבנית רשת) שציוניהם בחזאי א' נמוכים עוד יותר אך ציון הקריטריון שלהם גבוה יותר (שכן ציוניהם בחזאי ב' גבוהים במיוחד). מכאן שכאשר יחס הסלקציה נמוך

1. במונח זה נתייחס כאן ולהלן לכל המועמדים שהתקבלו, מימשו קבלתם ולא נשרו עד לסוף הדרך, כך שציון הקריטריון שלהם ידוע.

או בינוני, הוספת חזאי ב' למיון מחליפה מועמדים התורמים למתאם הנצפה בין חזאי א' לקריטריון במועמדים הגורעים ממנו. המתאם הנצפה מתכווץ אפוא.



ציור מס' 1. דוגמה להשפעת תוספת חזאי למשוואת המיון על התוקף הנצפה של החזאי הקודם. האליפסה מייצגת היטל של התפלגות דו-משתנית של שני חזאים, א' ו-ב'. הקו האופקי מייצג סף מיון על פי חזאי א' בלבד. הקו האלכסוני מייצג סף מיון על פי צירוף לינארי של שני החזאים. הקווים הקמורים כלפי הראשית מייצגים עקומות שוות-הצלחה, כאשר ההצלחה גבוהה יותר ככל שגדל המרחק מהראשית. (ראה הסבר בגוף הטכסט).

כדי להמחיש זאת, חישבנו באמצעות ניתוח מונטה-קארלו את מקדמי המתאם הצפויים להיצפות בקבוצת המתקבלים כאשר תוקפו האמיתי של חזאי א' הוא 0.45, תוקפו האמיתי של חזאי ב' הוא 0.45,

והמתאם בין החזאים הוא 0.51. החישוב התבצע לגבי שלושה יחסי סלקציה שונים, 0.25, 0.5, ו-0.75, בהנחה שההתפלגות הרב-משתנית היא טרי-נורמלית. הנתונים מוצגים בלוח מס' 1.

לוח מס' 1. מקדמי המתאם הצפויים בקבוצת המתקבלים כאשר תוקפו האמיתי של חזאי א' הוא 0.45, תוקפו האמיתי של חזאי ב' הוא 0.45, והמתאם בין החזאים הוא 0.51, עבור שלושה יחסי סלקציה שונים, 0.25, 0.5, ו-0.75.

הערה: התוצאות הן ממוצעי שתי הרצות מונטה-קארלו.

מקדם מתאם צפוי בקבוצת המתקבלים
בין חזאי א' לקריטריון

יחס הסלקציה	מיון על-פי	מקדם מתאם צפוי בקבוצת המתקבלים
0.25 (25% מתקבלים)	חזאי א'	0.23
	חזאי ב'	0.34
	ציון סכם	0.19
0.50 (50% מתקבלים)	חזאי א'	0.29
	חזאי ב'	0.36
	ציון סכם	0.25
0.75 (75% מתקבלים)	חזאי א'	0.34
	חזאי ב'	0.39
	ציון סכם	0.32

ובכן, מקדם המתאם הנצפה בין חזאי לקריטריון מהווה אומדן-חסר לתוקפו של החזאי ועל כן יש להשתמש בנוסחאות תיקון כדי לשחזר את התוקף האמיתי (0.45 בדוגמה זו). הנוסחאות הנלמדות בכל קורס לתורת המבחנים הן נוסחה לתיקון אפקט המהימנות ונוסחה לתיקון קיצוץ התחום הנובע מדחיית המועמדים הנמוכים בחזאי. כמו כן, קיימת נוסחת תיקון לקיצוץ תחום המתחשבת בריבוי חזאים (Gulliksen, 1950), שלמעשה הוצעה כבר בראשית המאה שעברה ע"י אבי רעיון המתאם הלינארי, קרל פירסון.

הנה כי כן, אומדן תוקפה של הבחינה הפסיכומטרית מחייב תיקון עבור שלושת גורמי הכיווץ במקדם המתאם הנצפה בקבוצת המתקבלים שסיימו. לא זו בלבד, שההישגים בלימודים אקדמיים נמדדים על סולם שמהימנותו לא גבוהה במיוחד, ושקיצוץ התחום הנגרם מהשימוש בציון הפסיכומטרי לצורך המיון הוא

ניכר, אלא שהשימוש גם בציוני הבגרות לצורך המיון גורם לכך שהמתאם הנצפה בין הציון הפסיכומטרי להישגים בלימודים קטן עוד יותר ממה שהיה נצפה אילו היה המיון על פי הציון הפסיכומטרי בלבד. תיקון מורכב זה מתבצע כיום במרבית בדיקות התוקף הנערכות בארץ. אם יש מי שסבור, שיש בכך משום ניפוח של ממצאים, סימן שלא הבין לאשורה את הבעיה בה עוסקים פסיכומטריקאים קרוב למאה שנה.

חזרה לאיילון ויוגב - הכשל בהמשגת הבעיה

המטרה היא אפוא להקיש על גודל המתאם האמיתי על פי המתאם הנצפה. קשה לדעת באיזו מידה היו איילון ויוגב (1997) מודעים למטרה, אך מכל מקום, יוגב ואיילון (2000) שכחו כפי הנראה מהי. הנה כך הם כותבים: "למעשה, אין כל צורך בויכוח על יישומו הנכון של מודל הקמן כדי להוכיח את חוסר הקשר בין הציון הפסיכומטרי להישגים הלימודיים באוניברסיטה. די להתבונן בהתפלגות המשותפת הבסיסית של שני גורמים אלה בקרב הסטודנטים בוגרי השנה הראשונה ללימודים... ה"עננים" המוצגים בתרשים מלמדים, כי אין כל קשר בין הציון הפסיכומטרי להישגים בלימודים של הסטודנטים...".

האמנם? והלא מכאן יצאנו. הבעיה הייתה להעריך עד כמה התכווץ מקדם המתאם, כדי לאמוד את מקדם המתאם האמיתי מן הנצפה. אם יוגב ואיילון סבורים שממצאים אלה, גם בלי להידרש לתיקון זה או אחר, "...מורים בבירור על היעדר קשר בין הציון הפסיכומטרי להצלחה בלימודים", אפשר שאין הם ערים מספיק לצורך בשחזור המקדם האמיתי. כפי שמודגם בלוח מס' 1, גם מקדם מתאם ניכר עשוי להתכווץ בקבוצת המתקבלים עד כדי כך, שדיאגרמת הפיזור תיראה כ"ענן" אמורפי למדי.

ובכל זאת, יש הגיון מסוים בהצעתם של איילון ויוגב לניתוח הנתונים: הם מעירים, כי במקרה בו יש "נשירה בלתי אקראית" (כך במקור) מקבוצת המתקבלים (נשירה הנקראת בפי Gross & McGanney "נשירה שלא ניתן להתעלם ממנה"), התיקון המקובל לקיצוץ תחום שתואר לעיל אינו מספק אומד חסר הטיה למקדם האמיתי, ולשם כך הם מציעים שימוש במודל של הקמן. אכן, אחת ההנחות בתיקון המקובל עלולה להיות מופרת עקב נשירה כתוצאה של גורמים בלתי ידועים, כגון סלקציה עצמית או פנייה למסלולי לימוד אחרים. במקרה זה התיקון המקובל צפוי להיות בדרך כלל שמרני מדי, קרי לספק אומד נמוך מדי למקדם האמיתי (Linn, Harnisch & Dunbar, 1981). מודל הקמן נועד לספק אומד מדויק יותר, ואמנם יישומו לצורך זה הוצע ע"י Gross & McGanney (1987) ו-Linn & Hastings (1984). בהמשך נבחן אם יישומם של איילון ויוגב את מודל הקמן אמנם מהווה פתרון עדיף. אך האם הם עצמם ממש הבינו את מגבלות תרומתו של המודל?

הסברם של איילון ויוגב לקוני מכדי לדעת בבטחה. אבל יישומם מעורר חשש שהם סברו בטעות, כי התיקון המקובל אינו מתמודד עם דחייתם של מועמדים על פי החזאי הנוסף. כפי שהראו קנת-כהן וברונר

(ראה פירוט להלן), המשגה החמור של איילון ויוגב הוא בשימוש במשוואת ניבוי בה כלול חזאי אחד בלבד (למשל, הציון הפסיכומטרי), למרות שהחזאי השני המשמש לסלקציה (במקרה זה, ממוצע הבגרות) מתואם עמו. הדרך הנכונה ליישום המודל, כפי שמוסבר להלן, היא כמובן לכלול במשוואת הניבוי גם את החזאי השני. ייתכן שמשגה זה נובע מההנחה, שהחזאי הנוסף אף הוא גורם לנשירה עמה אין התיקון המקובל מסוגל להתמודד. כמוסבר לעיל, הנחה זו מוטעית. החזאי הנוסף תורם לכיוון גדול יותר במקדם המתאם הנצפה, אך נוסחת התיקון לקיצוץ תחום המתחשבת בריבוי חזאים מתקנת אף זאת.

בעיות בניתוח הנתונים

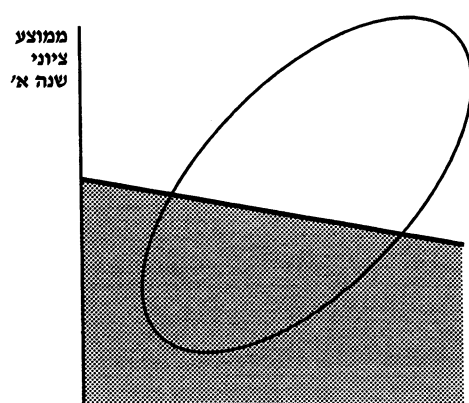
כך או אחרת, משגה חמור זה הוא שגרם לאיילון ויוגב להמעיט בתרומתו של ציון המבחן הפסיכומטרי לחיזוי. במשוואת הניבוי ששימשה לאומדן תרומה זו הם כללו את ציון המבחן הפסיכומטרי בתוספת משתנה תיאורטי (λ) המהווה אומדן להשפעה על הנשירה של כל הגורמים פרט לחזאים. מאחר שמשתנה נוסף זה קשור במידה גבוהה בצירופם של שני החזאים, הוא מנבא טוב יותר את הקריטריון מאשר החזאי היחיד המופיע במשוואה. בשל מתאם הביניים הגבוה בין החזאים, מתכווץ מקדם הרגרסיה של החזאי במידה ניכרת. כאשר יישמו קנת-כהן וברונר את המודל בדרך הנכונה, הם קיבלו ממצאים דומים לאלה השכיחים בספרות המחקרית.

אלא שגם כך כלל לא ברור אם יישומו של ניתוח זה על מסדי הנתונים בהם השתמשו החוקרים יכול לספק את המבוקש. מקדם הרגרסיה משקף את תרומת החזאי לניבוי הקריטריון בקבוצת המתקבלים. ניתן היה להקיש ממנו על מקדם הרגרסיה בקבוצת כל המועמדים, אילו אמנם התקיימו כל הנחות המודל ואילו המדגם היה גדול מספיק. במקרה אותו ניתוח איילון ויוגב, המדגמים קטנים מהנדרש. יתרה מזו, ספק אם כל הנחות המודל מתקיימות בו. ככל מודל מורכב, גם המודל של הקמן מרבה בהנחות, ולכן קיים סיכון ניכר שלפחות אחת מהנחותיו אינה תקפה (עיין גם Berk, 1983). מאחר שלא ברור עד כמה המודל חסין להפרת הנחות אלה, לא ניתן להתעלם מהאפשרות שהאומדן אותו מספק הניתוח לפרמטרים בקבוצת כל המתקבלים הנו מוטעה.

יתר על כן, משהו בממצאי הניתוחים, מעבר לערכיהם החריגים כשלעצמם של המקדמים, מעורר חשש כבד שאמנם כך הדבר: רבים ממקדמי הרגרסיה בקבוצת המתקבלים היו גבוהים מן האומדים שסיפק המודל למקדמי הרגרסיה בקבוצת כל המועמדים. א-פד'ורי, סבירותו של ממצא כזה קטנה ביותר: בעוד שכאשר לא קיימת "נשירה שלא ניתן להתעלם ממנה", מקדם הרגרסיה בקבוצת המתקבלים שווה על פי הגדרה למקדם הרגרסיה בקבוצת כל המועמדים, הרי שכאשר קיימת נשירה כזאת, סביר שמקדם הרגרסיה בקבוצת המתקבלים יהיה דווקא נמוך יותר מאשר מקבילו בקבוצת כל המועמדים (ראה למשל, Gross &

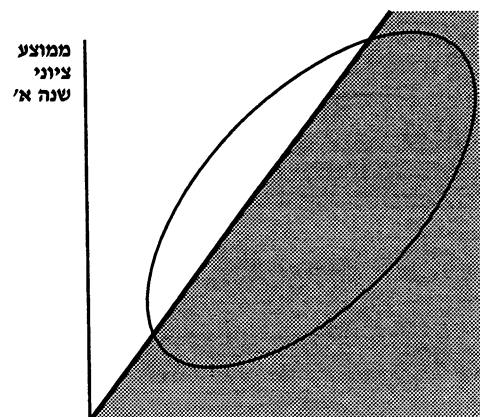
(McGanney, 1987).

הסיבה לכך היא שכל המשתנים העומדים במתאם חיובי עם ציון הקריטריון (כגון: כושר מנטלי, מוטיבציה וכיו"ב) נוטים להשפיע על הנשירה באותו כיוון. אם למשל, סטודנטים נוטים לנשור יותר ככל שכושרם נמוך, סביר שנטייתם לנשור תגבר ככל שהמוטיבציה שלהם נמוכה. יתר על כן, מאחר שהמשתנה המתווך כאן הוא ככל הנראה הסיכוי להצליח כפי שהוא נתפס ע"י הסטודנט, אזי תהיה תחלופה בין כל המשתנים הללו. לדוגמה, סביר ששיעור הנשירה של סטודנטים שציוניהם בשנה א' נמוכים יהיה קטן יותר ככל שכושרם המנטלי גבוה יותר (ראה ציור 2א) או ככל שהמוטיבציה שלהם גבוהה יותר. לא סביר כלל ששיעור נשירה זה יהיה גדול יותר ככל שכושרם המנטלי גבוה יותר (ראה ציור 2ב). ברור שבמקרה הראשון הנשירה יוצרת קיצוץ המקטין את שיפוע קו הרגרסיה (יחסית לשיפוע בקבוצת המועמדים כולה), ואילו במקרה השני היא יוצרת קיצוץ המגדיל שיפוע זה.



כושר מנטלי

(א)



כושר מנטלי

(ב)

ציור מס' 2. דוגמאות להשפעת נשירה על שיפוע קו הרגרסיה בתוך קבוצת המתקבלים (ראה טכסט). האליפסה מייצגת היטל של התפלגות דו-משתנית של חזאי וקריטריון, והקו מייצג גבול דמיוני בין אזור נשירה גבוהה מהמדגם ואזור נשירה מועטת. שיפועי הרגרסיה עצמם אינם מוצגים כאן, אך קל לראות כי בפאנל א' מודגם מקרה בו השיפוע קטן מהשיפוע בקרב קבוצת כל המועמדים, וכי בפאנל ב' מודגם מקרה בו השיפוע גדול מהשיפוע בקרב קבוצת כל המועמדים.

זה מן הסתם הבסיס להנחה המקובלת, במפורש או במובלע, בספרות הפסיכומטרית (ראה למשל, Linn, 1983), כי במידה שמקדם הרגרסיה משתנה עקב הקבלה והנשירה, שיפועו צפוי לקטון, וכי על כן נוסחת התיקון המקובלת מספקת אומדן-חסר למקדם המתאם האמיתי (ראה Linn, Harnisch & Dunbar, 1981). זו מן הסתם הסיבה לכך, שבמאמרם הנזכר לעיל, Gross & McGanney דנים רק במקרה כזה לצורך הדגמת הטיפול של מודל הקמן בנשירה שלא ניתן להתעלם ממנה. ובכן, למרות שמודל הקמן ישים לחלוטין לשחזור מקדמי רגרסיה במסדי נתונים דומים כאשר כל הנחותיו מתקיימות (ראה תוצאות ההדמיה בקנת-כהן וברונר, 1997), הגידול בשיפועי הרגרסיה במדגמים אותם ניתחו איילון ויוגב, ואף קנת-כהן וברונר אחריהם, מעלה חשש להטיה באומדים עקב הפרה קיצונית של הנחות המודל. נראה שהמהלך המתודולוגי הסביר בדילמה כזו הוא לדבוק בניתוח המקובל, שהינו חסכוני יותר בהנחותיו וככל הנראה שמרני יותר בתוצאות. לבסוף, הדיון בממצאיהם של איילון ויוגב לוקה בשל השימוש הבלעדי במקדמי רגרסיה לשם אומדן תקפות, קל וחומר למטרת השוואה בין תקפיות. כידוע, גודלם של מקדמי הרגרסיה אינו עומד ביחס לינארי לגודלן של התקפיות, כי הוא תלוי גם בגודל סטיות התקן של החזאים. יתר על כן, ערכם של מקדמי רגרסיה כמדדים בפני עצמם ליכולת חיזוי שנוי במחלוקת בספרות (ראה Budescu, 1993).

שימוש בנתונים סלקטיביים

איילון ויוגב ניתחו 11 מסדי נתונים, אותם קיבצו ממחלקות שונות באופן שניתן לחלוק עליהם. קנת-כהן וברונר (1997) ניתחו 25 מסדי נתונים מאותו מחזור לימוד ואותו מוסד ממנו נלקחו נתונייהם של איילון ויוגב, במטרה לשמור על דמיון רב ככל האפשר בין מסדי הנתונים. יוגב ואיילון אף הציגו נתונים נבחרים מתוך כל אלה. לעומת זאת, קנת-כהן, ברונר ואורן (1999) כללו בניתוח-העל שלהם 1861 מסדי נתונים, מכל מסלולי הלימוד מרובי התלמידים בכל האוניברסיטאות ומ-12 מחזורי לימוד, לרבות מחקרי אורך. בתוכם נמצאו בדרך כלל מתאמים נצפים חיוביים ניכרים לציון הפסיכומטרי, אם כי היו גם מקרים אחדים בהם התקבלו מתאמים אפסיים. באוסף הנתונים הזעיר יחסית בו השתמשו איילון ויוגב נמצאו מתאמים כאלה, לא במפתיע בחוגי הלימוד הסלקטיביים יותר באוניברסיטה בה שיעור הקיצוץ גדול בשל היקף ההרשמה הגבוה במיוחד, ושבה ציון הבגרות מהימן במיוחד בשל ההרכב ההומוגני במיוחד של אוכלוסיית המועמדים, הרכב שאינו מאפיין את יתר האוניברסיטאות. מותר להניח, שמתאמים נמוכים יחסית אלה נדגמו למעשה מהעשירון התחתון של סולם התוקף.

מה יכול היה להביא את איילון ויוגב לנתח 11 מסדי נתונים שמידת יציגותם מפוקפקת א-פריורי במקום להשתמש בכלל הנתונים הזמין היום? מה יכול היה להביאם להיחפז ולהסיק על סמך ניתוח של כחצי אחוז מכלל הנתונים, כי יש דופי בממצאיהם של כל יתר החוקרים לאורך עשרות שנים?

התעלמות מממצא קריטי

לבסוף, כשיושבים בחדר סגור ומתווכחים שעות אם יש גשם בחוץ, יש ערך מסוים לעדות של מי שעלה בדעתו להציץ החוצה דרך החלון.

לא כל הבדיקות בנושא נזקקו לתיקון מקדמי מתאם מכווצים. באחת מהן (קידרמן, 1987; Ben-Shakhar & Beller, 1996; Kiderman & Beller, 1996) נבדקה קבוצה, שנוצרה ע"י קבלה אקראית של מועמדים שציון החזאי שלהם (המבוסס על שילוב של ציון המבחן הפסיכומטרי וממוצע הבגרות) היה נמוך מסף הקבלה. במקרה זה גורם הכיווץ היחיד הוא מהימנות הקריטריון, ולכן המתאם הנצפה בין החזאי להצלחה בלימודים הוא כמעט אומדן בלתי מוטה של המתאם באוכלוסייה. המתאם בין החזאי לבין קריטריון דיכטומי של הצלחה בלימודים אקדמיים (כאשר הצלחה הוגדרה כממוצע ציונים של 70 לפחות) היה 0.53, כלומר גבוה מהמתאמים הנצפים בד"כ בקרב מדגמים שקוצצו ע"י קבלה על סמך החזאי, בדיוק כצפוי על פי המודלים המקובלים בתחום. והעיקר, בקרב התלמידים שהתקבלו למרות שציון החזאי שלהם היה מתחת לסף הקבלה שיעור המצליחים היה 18.5%, בעוד שבקרב תלמידים שעברו את הסף היה שיעור המצליחים 71.2% (Ben-Shakhar & Beller, 1983). בסופו של דבר, זה המבחן המעשי העיקרי לערכו של כלי מיון, ולכן חשוב יותר מגודלו של מקדם המתאם ומתקפות ההנחות המאפשרות שחזורו ממקדם מתאם נצפה.

סיכום

יוגב ואיילון ערכו מחקר לקוי מבחינה מקצועית. איילון ויוגב אף הרחיקו לכת וכתבו נייר עמדה מגמתי, בו הם מנסים לטשטש את המשגה החמור במחקרם המוקדם, מסכמים את ממצאי קנת-כהן וברונר בדרך שאינה מקובלת בהסקה ממקבץ נתונים סטטיסטיים, ומביאים "ראייה" אחרת המסתמכת על נתונים מעטים וסלקטיביים, מה גם שהנתונים הם דיאגרמות פיזור נצפות שכידוע אינן משקפות כשלעצמן את מלוא תוקף הניבוי. חמור מכל, למרות שאת ממצאיהם וטיעוניהם הם לא העמידו במבחן המקובל באקדמיה, קרי פרסום בכתב עת הכפוף לשיפוט עמיתים, הם לא נרתעו שלוש שנים לאחר הפצת הדו"ח הראשון מניסוח המלצות לשינוי מדיניות הקבלה לאוניברסיטאות על בסיס רעוע זה. המעט שניתן לומר על כך הוא, כי הבשורה המדעית על פי יוגב ואיילון נולדה טרם זמנה.

מראי מקום

- אורן, כ. (1992). מערכת המיון להשכלה הגבוהה בארץ בשנים 1984-1989: סיכום מקוצר של ממצאי תוקף ניבוי (דוח מס' 160). ירושלים: מרכז ארצי לבחינות ולהערכה.
- אורתור, ג. (1965). בחינות הכניסה למחלקה לפסיכולוגיה באוניברסיטה העברית. מגמות, יג, 331-334.
- איילון, ח., ויוגב, א. (1997). גורמי הטיה בחיזוי ההצלחה בלימודים באוניברסיטה (נייר דיון מס' 97-3). תל-אביב: המרכז לפיתוח ע"ש פנחס ספיר ליד אוניברסיטת תל-אביב.
- בלר, מ., ובן-שחר, ג. (1981). הערכת תהליך בררת הסטודנטים באוניברסיטה העברית בירושלים. מגמות, כז, 22-36.
- יוגב, א., ואיילון, ח. (2000). קריטריוני המיון בכניסה לאוניברסיטה – לאן? (נייר עמדה). ירושלים: מכון ון ליר.
- ספז, ג'. (1990). תוקף ניבוי של הציונים הפסיכומטריים וממוצע ציוני הבגרות באוניברסיטת חיפה בשנים תשמ"ה, תשמ"ו, תשמ"ז, תשמ"ח. אוניברסיטת חיפה: היחידה לתכנון אקדמי.
- פרלברג, א. (1965). חיזוי הישגים הלימודיים של סטודנטים בטכניון. מגמות, יג, 42-56.
- קידרמן, א. (1987). יישום מודל קבלת החלטות להשוואה בין כדאיות השימוש בין שתי מערכות מיון בסיטואציה של בחירה למספר מקומות בלתי מוגבל. עבודת גמר לתואר מוסמך בפסיכולוגיה, האוניברסיטה העברית בירושלים.
- קנת-כהן, ת., וברונר, ש. (1997). תוקף מרכיבי מיון המועמדים להשכלה גבוהה: תיקון עבור הטיה הנובעת מברירה באמצעות המודל של הקמן (דוח מס' 236). ירושלים: מרכז ארצי לבחינות ולהערכה.
- קנת-כהן, ת., אורן, כ., וברונר, ש. (1994א). תוקף הבחינה הפסיכומטרית ותעודת הבגרות ומשקלם היחסי של חזאים אלה בניכוי הישגי שנה א' באוניברסיטאות בישראל (דוח טכני מס' 29). ירושלים: מרכז ארצי לבחינות ולהערכה.
- קנת-כהן, ת., אורן, כ., וברונר, ש. (1994ב). תוקף הבחינה הפסיכומטרית ותעודת הבגרות ומשקלם היחסי של חזאים אלה בניכוי ציוני בוגר באוניברסיטאות בישראל (דוח טכני מס' 36). ירושלים: מרכז ארצי לבחינות ולהערכה.
- קנת-כהן, ת., ברונר, ש., ואורן, כ. (1999). ניתוח-על של תוקף ניבוי של מרכיבי מערכת המיון לאוניברסיטאות בישראל כלפי מידת ההצלחה בלימודים. מגמות, מ, 54-71.

שילד, ש., ומאיר, א. (1991). תוקף הניבוי של מבחני הכושר ב"מכון הדסה לייעוץ בבחירת מקצוע" כלפי הצלחה במסלולי לימוד והכשרה: ניתוח-על. מגמות, לה, 202-215.

שפירא, ד., ועציוני-הלוי, ח. (1973). הצלחה בלימודים אקדמיים: באיזו מידה ניתנת היא לניבוי? מגמות, יט, 215-230.

Beller, M. (1994). Psychometric and social issues in admissions to Israeli universities. Educational Measurement: Issues and Practice, 13, 12-20.

Ben Shakhar, G., & Beller, M. (1983). An application of decision-theoretic model to a quota-free selection problem. Journal of Applied Psychology, 68, 137-146.

Ben-Shakhar, G., Kiderman, I., & Beller, M. (1996). Comparing the utility of two procedures for admitting students to Liberal Arts: An application of decision-theoretic models. Educational and Psychological Measurement, 56, 90-107.

Berk, R.A. (1983). An introduction to sample selection bias in sociological data. American Sociological Review, 48, 386-389.

Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. Psychological Bulletin, 114, 542-551.

Gross, A.L., & McGanney, M.L. (1987). The restriction of range problem and nonignorable selection processes. Journal of Applied Psychology, 72, 604-610.

Guilford, J.P. (1987). Fundamental statistics in psychology and education. New York: McGraw-Hill.

Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.

Heckman, J.J. (1979). Sample selection bias as a specification error. Econometrica, 47, 153-161.

Linn, R.L. (1983). Predictive bias as an artifact of selection procedures. In H. Wainer & S. Messick (Eds.), Principles of modern psychological measurement: A festschrift for Federic M. Lord. Hillsdale, NJ: Erlbaum.

Linn, R.L. & Hastings, C.N. (1984). Group differentiated prediction. Applied Psychological

Measurement, 8, 165-172.

Linn, R.L., Harnisch, D.L., & Dunbar, S.B. (1981). Correction for range restriction: An empirical investigation of conditions resulting in conservative correction. Journal of Applied Psychology 66, 655-663.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, Addison-Wesley.

Willingham, W.W., Lewis, C., Morgan, R., & Ramist, L. (1990). Predicting college grades: An analysis of institutional trends over two decades. Princeton, NJ: Educational testing Service.

