

בחינת הרגישות של אומדני פרמטרים של IRT לגודל המדגמים של תגובות הנבחנים

מיכל באומר, יואל רפ

המרכז הארצי לבחינות ולהערכה



רקע

- ✓ פיתוח מבחן חדש במאל"ו
- ✓ מבחן ידע באנגלית כשפה זרה
 - יחליף את כל מבחני האנגלית האחרים (המועברים ברובם במועדים בודדים)
- ✓ המיומנויות שיבדקו
 - הבנת הנקרא
 - הבנת הנשמע **חדש!**
 - כתיבה **חדש!**
- ✓ מבחן ממוחשב ואדפטיבי
- ✓ יועבר לאורך כל השנה במרכזי היבחנות ברחבי הארץ

מאלו

מבחן אדפטיבי

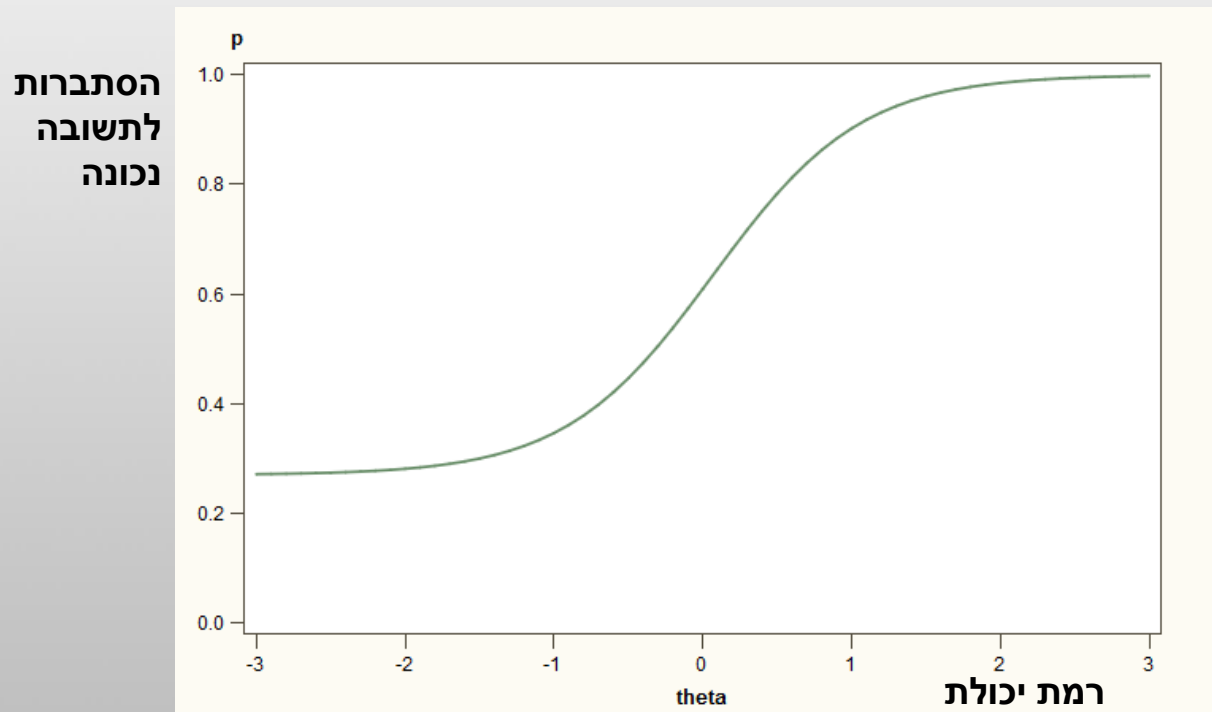
- ✓ רמת היכולת של הנבחן נאמדת תוך כדי המבחן
- ✓ הנבחן נדרש לענות על שאלות ברמת קושי התואמת את רמת היכולת שלו
- ✓ רמה זו נקבעת במהלך המבחן בכל שאלה/קבוצת שאלות על פי ביצועיו בשאלות הקודמות – והשאלה הבאה נדגמת בהתאם
- ✓ יש צורך לדעת מראש את רמת הקושי של הפריטים (ומאפיינים אחרים) ולהתאימם לרמת הנבחן
 - מודל IRT (Item Response Theory)
 - אומדן פרמטרים של IRT

מודל IRT תלת-פרמטרי

- ✓ כל פריט מאופיין על ידי שלושה פרמטרים
 - הבחנה (a), קושי (b), "ניחוש" (c)
- ✓ לכל נבחן/ת רמת יכולת בתכונה הנמדדת
 - נהוג לסמן את רמת היכולת באות θ
 - רמת היכולת יכולה להיות בין $-\infty$ ל- ∞ (בד"כ בין -3 ל-3)
- ✓ **ההסתברות למתן תשובה נכונה לפריט תלויה ברמת היכולת של הנבחן ובפרמטרים של הפריט**
- ✓ העקומה שמתארת את ההסתברות הזו נקראת **ICC – Item Characteristic Curve (אופיין הפריט)**

מודל IRT תלת-פרמטרי (המשך)

דוגמה אופיין של פריט שהפרמטרים שלו הם: $a=1.18$ (הבחנה), $b=0.07$ (קושי), $c=0.27$ ("ניחוש")

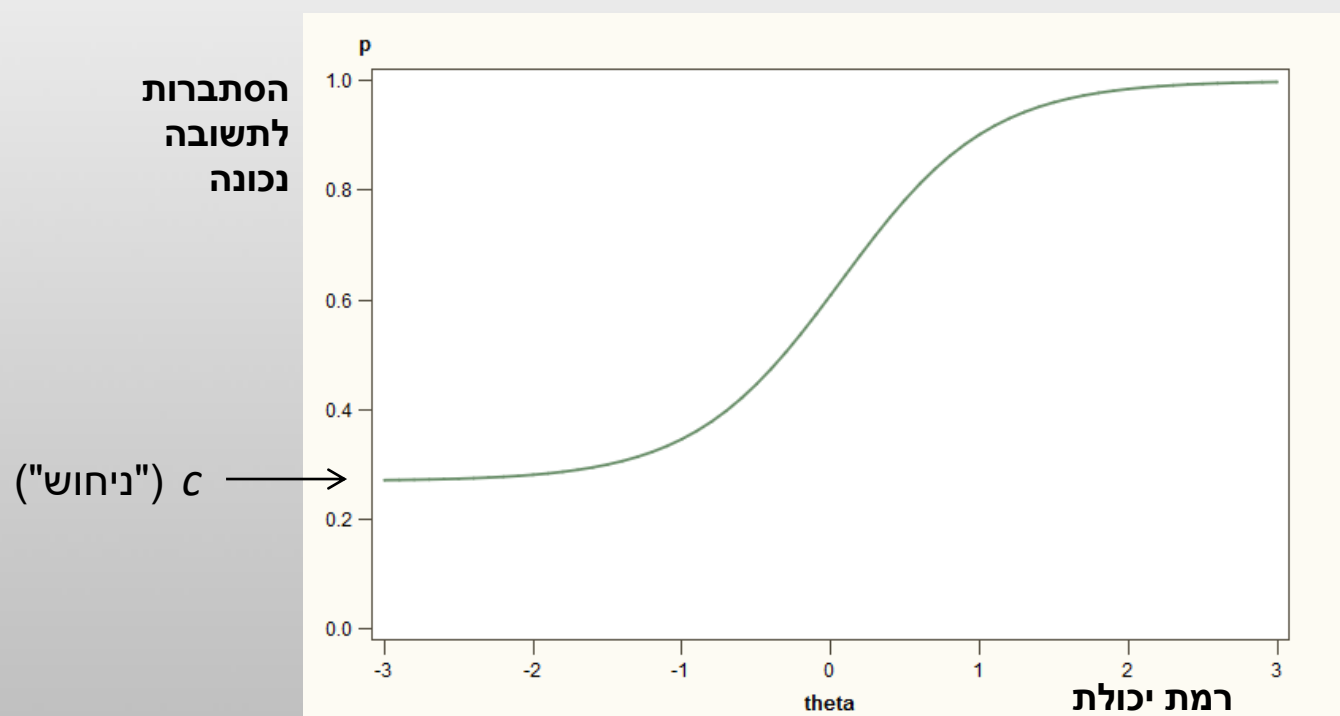


$$P(\theta) = c + (1-c) \frac{e^{1.7a(\theta-b)}}{1 + e^{1.7a(\theta-b)}}$$

θ (theta) = רמת היכולת של הנבחן/ת
... $e=2.71$

מודל IRT תלת-פרמטרי (המשך)

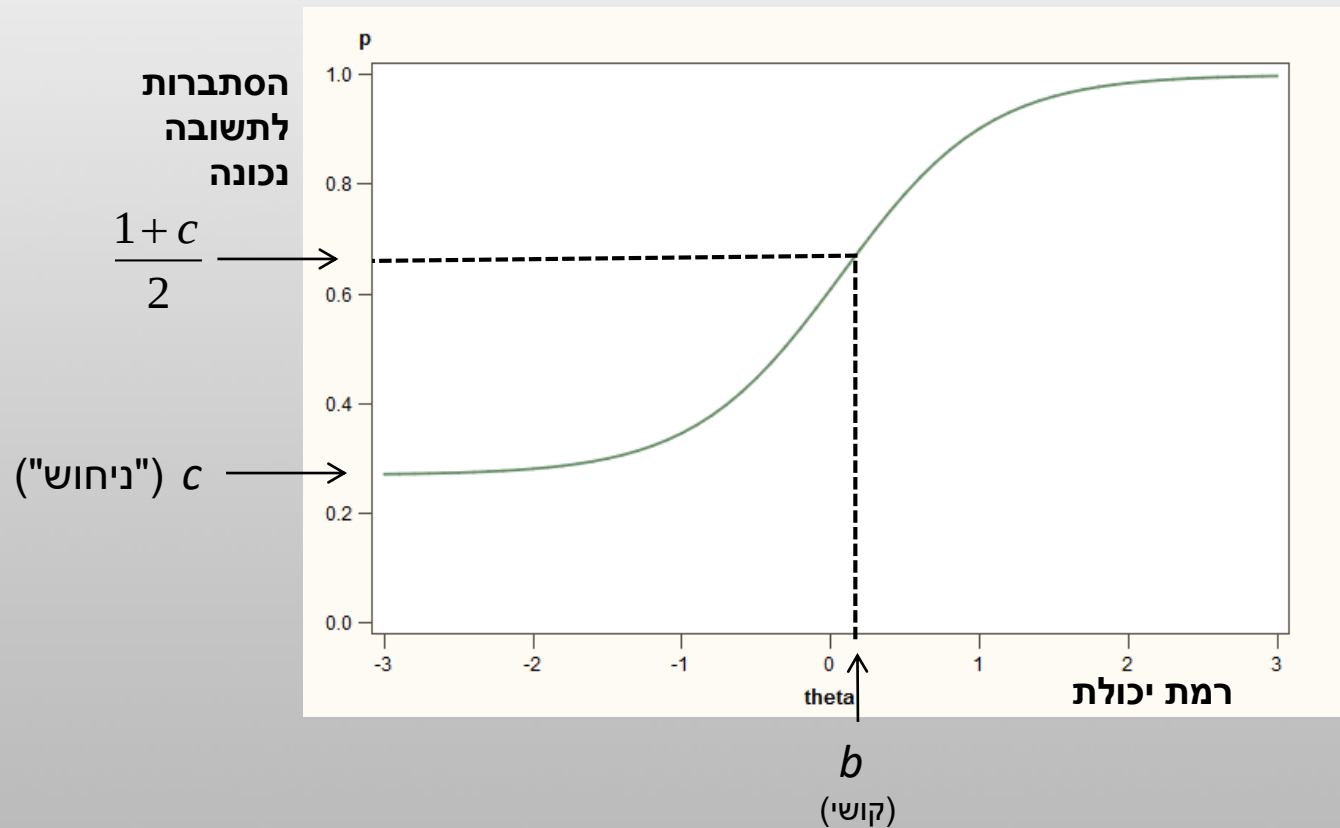
דוגמה אופיין של פריט שהפרמטרים שלו הם: $a=1.18$ (הבחנה), $b=0.07$ (קושי), $c=0.27$ ("ניחוש")



פרמטר c = ההסתברות לתשובה נכונה ע"י נבחן/ת בעל/ת רמת יכולת נמוכה מאוד

מודל IRT תלת-פרמטרי (המשך)

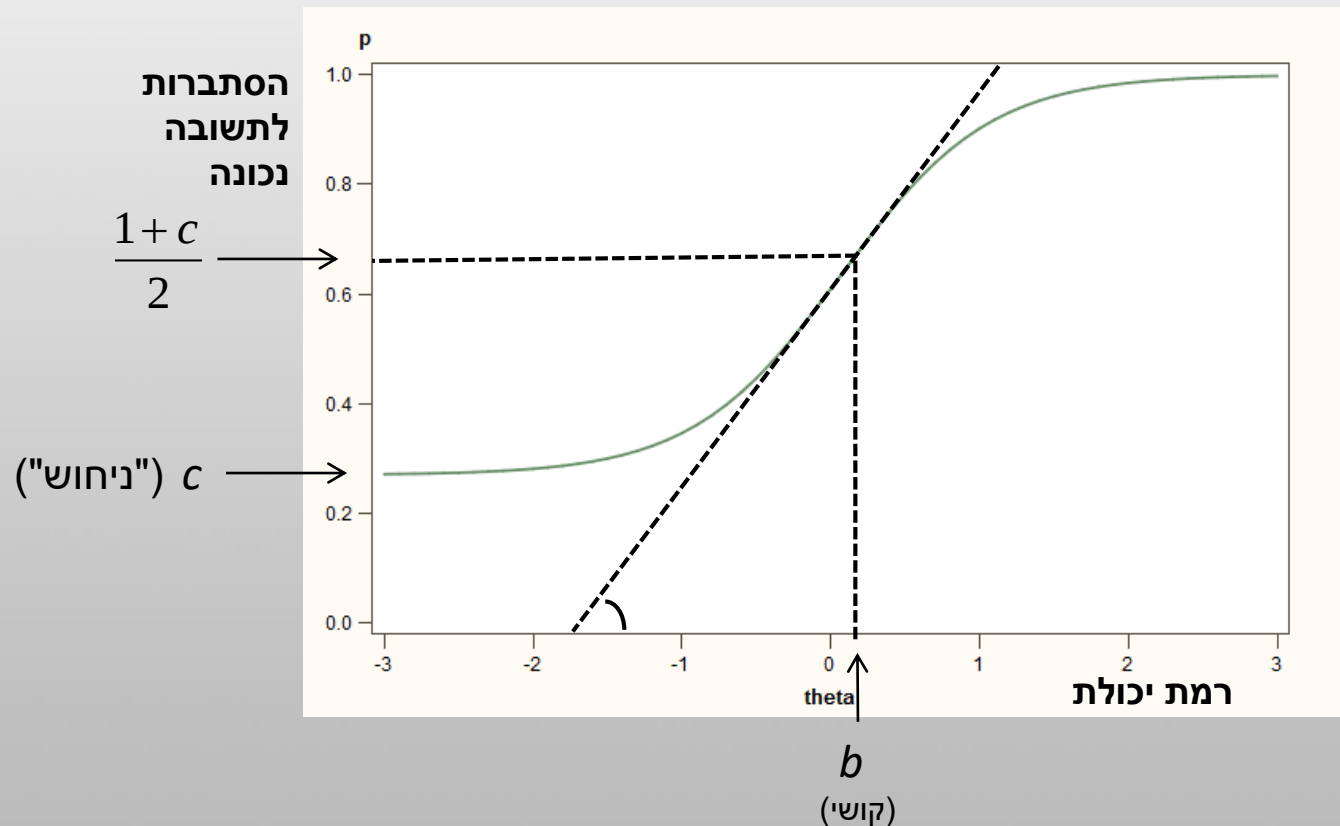
דוגמה אופיין של פריט שהפרמטרים שלו הם: $a=1.18$ (הבחנה), $b=0.07$ (קושי), $c=0.27$ ("ניחוש")



פרמטר b = רמת היכולת שבה ההסתברות לתשובה נכונה היא $\frac{1+c}{2}$

מודל IRT תלת-פרמטרי (המשך)

דוגמה אופיין של פריט שהפרמטרים שלו הם: $a=1.18$ (הבחנה), $b=0.07$ (קושי), $c=0.27$ ("ניחוש")



פרמטר ההבחנה (a) מתייחס לשיפוע המקסימלי של הפריט, המתקבל עבור רמת יכולת זהה לפרמטר הקושי (b).
ככל שאופיין הפריט תלול יותר, הפריט מבחין יותר

האתגרים

✓ כדי לאמוד את הפרמטרים של הפריטים, משתמשים בתגובות של מאות/אלפי נבחנים על אותם פריטים.

✓ עם המעבר ממבחנים רחבי היקף המועברים להרבה נבחנים במספר קטן של מועדים, למבחנים מתמשכים, בכל פעם למספר קטן של נבחנים, יידרש פיתוח מואץ של פריטים רבים

הרצה ניסיונית ← בדיקה ← אמידת פרמטרים ← שילוב הפריטים במבחן

✓ לצורך אמידת הפרמטרים, יש לאסוף תגובות רבות, לכן יש להעביר את הפריטים החדשים לאורך מספר ימים (בכל יום למספר מצומצם של נבחנים)

מטרת המחקר

✓ לבחון את רגישות אומדני הפרמטרים של IRT לגודל המדגמים של תגובות הנבחנים

- לבדוק עד כמה מושפעים הפרמטרים של הפריטים מגודל המדגם המשמש לאמידתם
- לקבוע מהו מספר הנבחנים המינימלי הסביר לעריכת אומדן פרמטרים

שיטה

✓ נלקחו תגובות נבחנים מפרקי האנגלית במכפ"ל מ-2016 (ני"ע)

- שני נוסחים, סה"כ 132 פריטים
- שהועברו לדוברי עברית ולדוברי ערבית
- כל נבחן קיבל 44 או 66 פריטים
- סה"כ כ- 17000 נבחנים (בשני הנוסחים)

✓ אמידת פרמטרים באמצעות תוכנת IRTPro

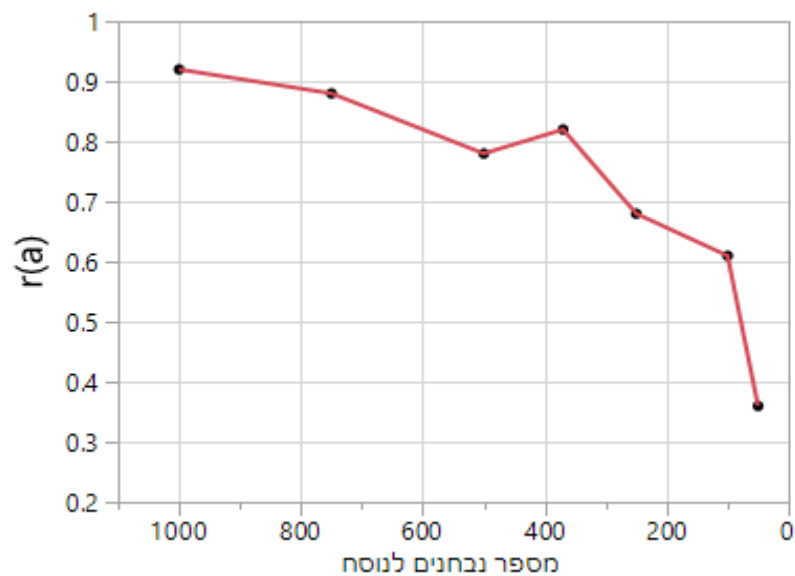
✓ הרצת הליכי אמידה שונים:

- אמידה על בסיס כל התגובות (ההליך הראשי)
- אמידה על בסיס מדגמים בגודל משתנה (מתוך בסיס התגובות)
 - ממדגם של 1000 ועד מדגם בגודל 50 לנוסח

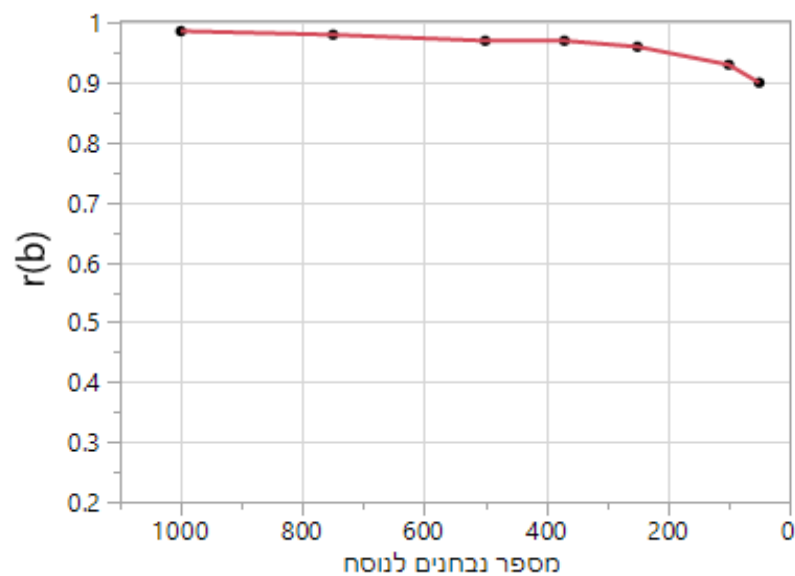
✓ ניתוח התוצאות שהתקבלו והשוואתם להליך הראשי

- עפ"י המתאמים בין סדרות האומדנים בכל הליך לבין ההליך הראשי
- השוואת אומדני הפרמטרים a, b, c, θ לפרמטרים בהליך הראשי

פרמטר ההבחנה (a)



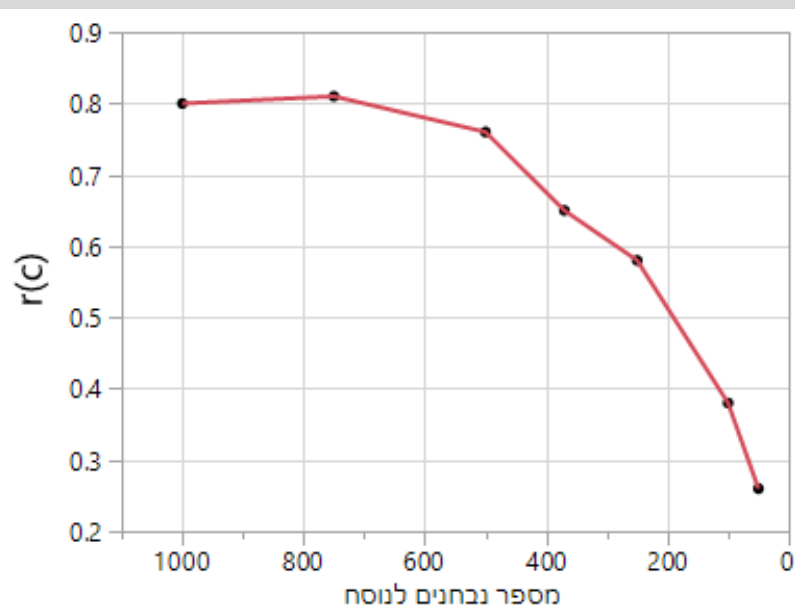
פרמטר הקושי (b)



תוצאות

1

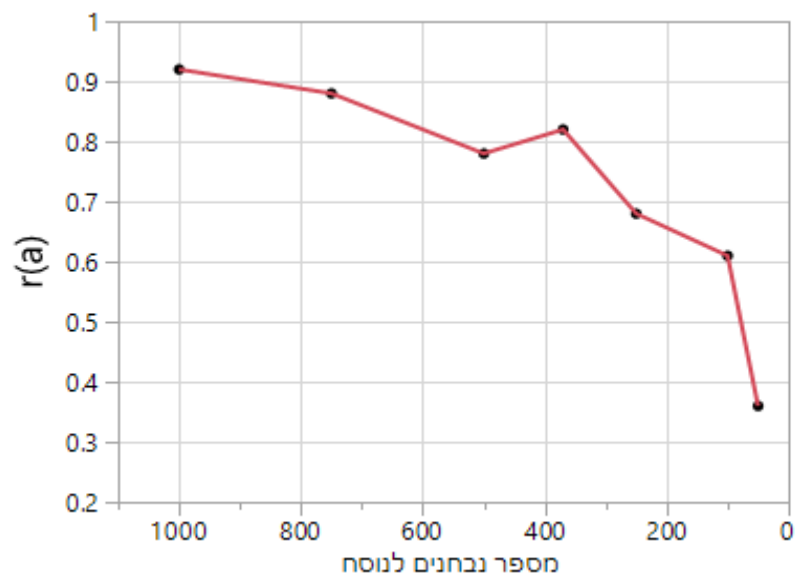
פרמטר הניחוש (c)



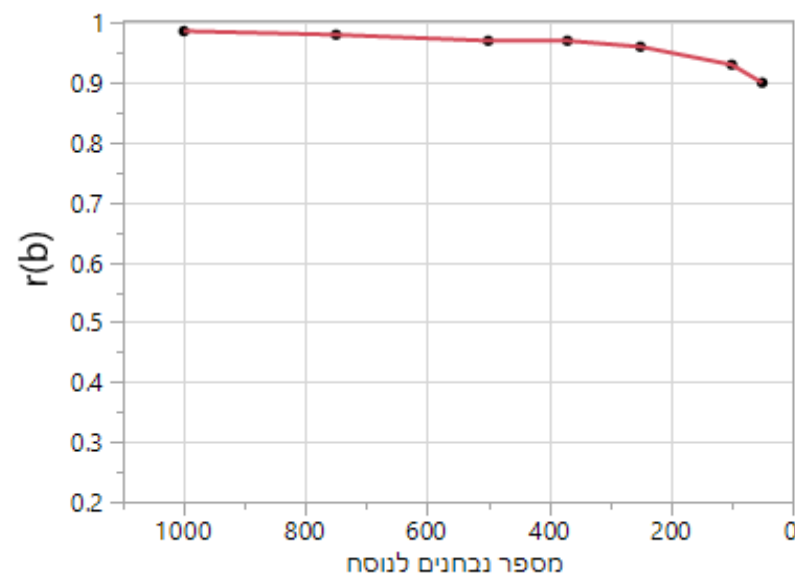
תוצאות

1

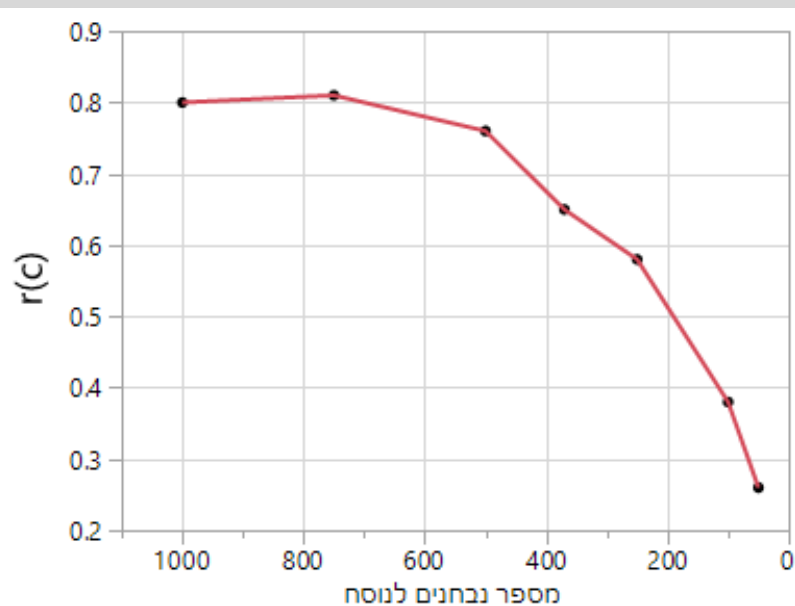
פרמטר ההבחנה (a)



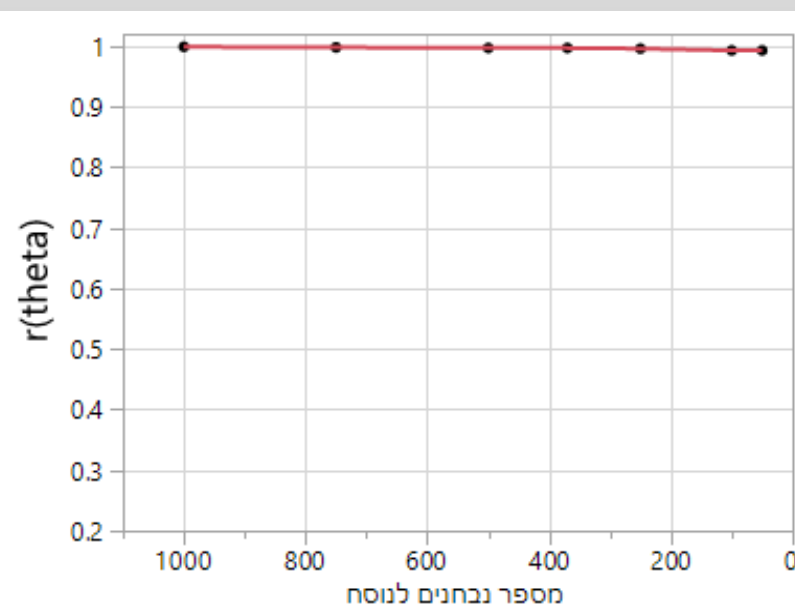
פרמטר הקושי (b)



פרמטר הניחוש (c)



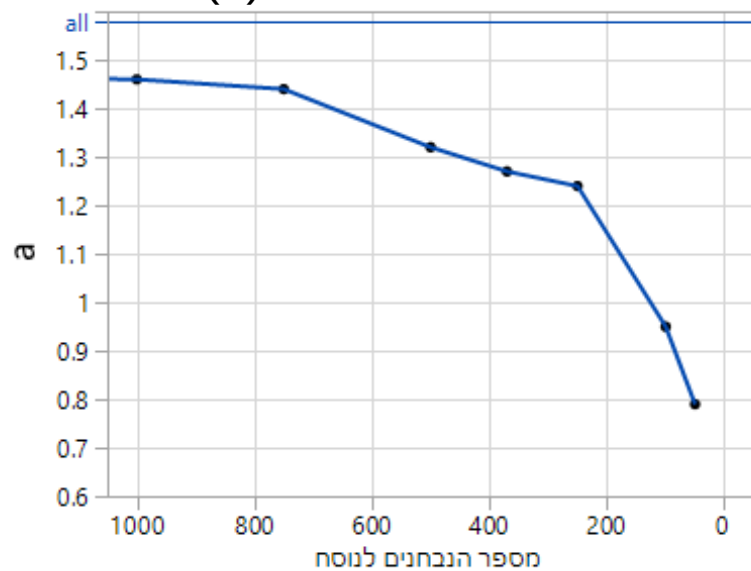
רמת היכולת (θ)



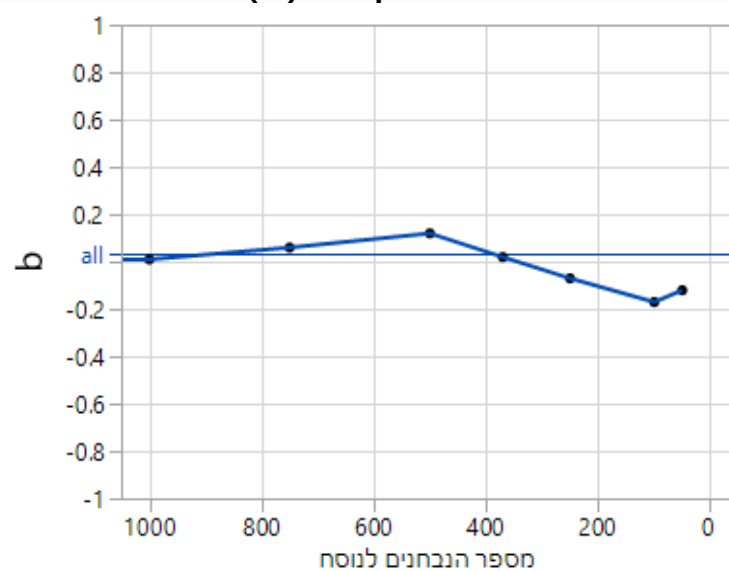
תוצאות

2

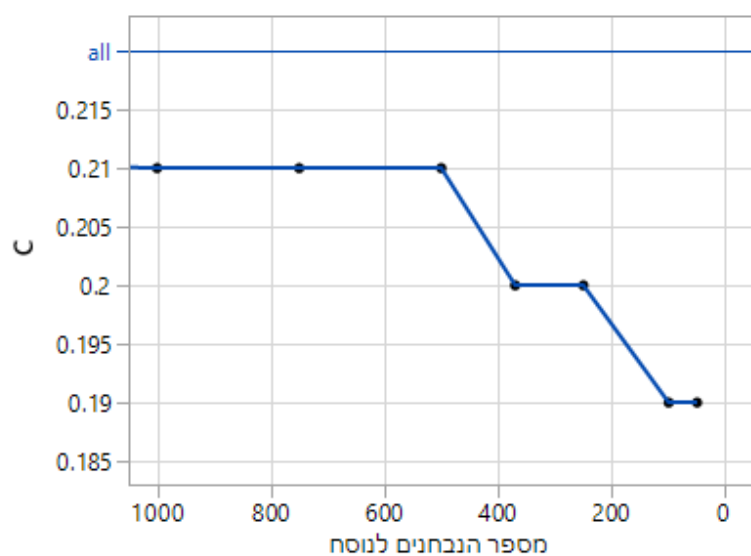
פרמטר ההבחנה (a)



פרמטר הקושי (b)



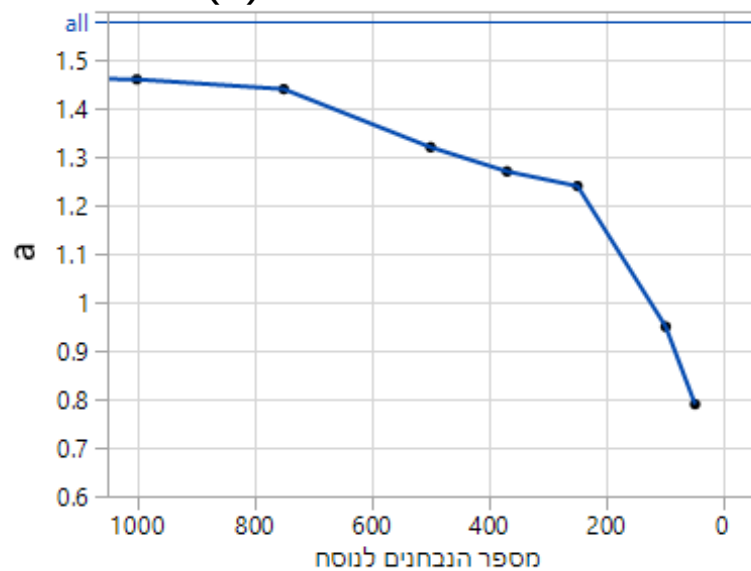
פרמטר הניחוש (c)



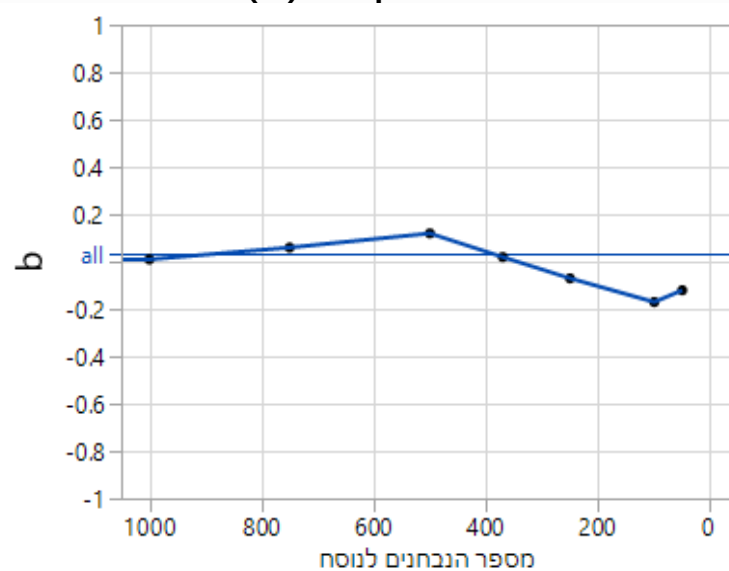
תוצאות

2

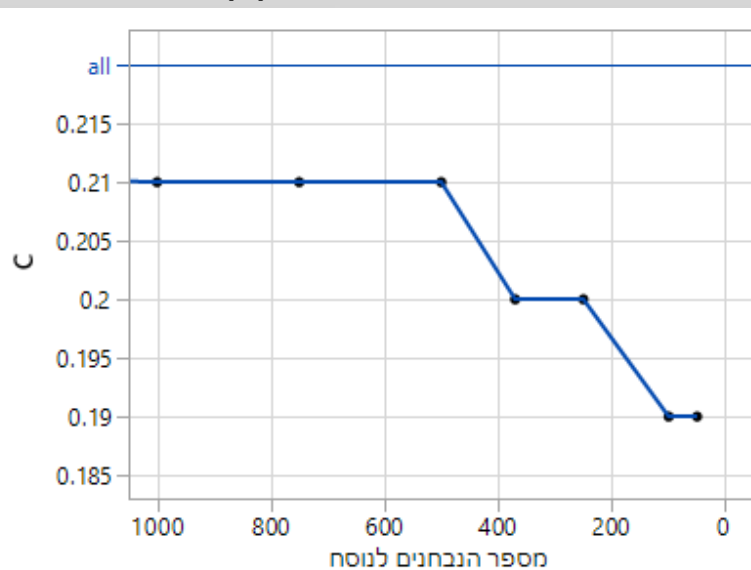
פרמטר ההבחנה (a)



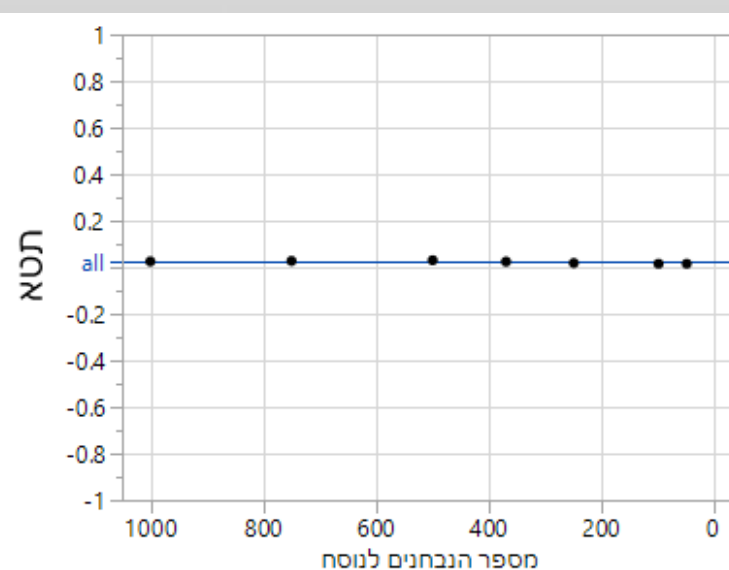
פרמטר הקושי (b)



פרמטר הניחוש (c)



רמת היכולת (θ)



תוצאות ①

- ✓ כצפוי, ככל שהמדגם קטן כך ההתאמה של האומדנים עם אלה שהתקבלו בהליך האמידה המלא קטנה
- בפרט קטנה ההתאמה עבור פרמטר ההבחנה (a)
 - מ-0.92 עד 0.36
 - קטנה ההתאמה גם עבור פרמטר הניחוש (c)
 - מ-0.8 עד 0.26
 - לעומת זאת, אומדני פרמטר הקושי (b) שמרו על תוצאות דומות למדי
 - מ-0.986 עד 0.9
 - ההתאמה של אומדני ציוני התטא (θ) קטנה במעט מאוד
 - מ-0.999 עד 0.993

תוצאות ②

✓ ככלל, הקטנת המדגמים שעליהם מתבססת האמידה מניבה אומדנים קטנים יותר

- פרמטר ההבחנה (a) ירד באופן עקבי עד כמעט פי 2
- ירידה מסוימת נרשמה גם בפרמטר הניחוש (c)
- פרמטר הקושי (b) נשאר די יציב
- תופעות אילו אינן מוכרות בספרות ויש לחקור אותן

✓ אומדני ציוני התטא (θ) נשארו יציבים ונראה כי אינם רגישים לירידה בגודל המדגם

מסקנות והמלצות

- ✓ אומדני פרמטר הקושי שמרו על תוצאות דומות
- ✓ ציוני הנבחנים אינם רגישים לאי דיוק באומדני הפרמטרים
 - כל עוד פרמטר הקושי נאמד במהימנות סבירה
- ✓ אולם, חשוב להגדיל ככל האפשר את המדגם ולפרוש אותו על הרבה אירועי היבחנות
 - כי לא ניתן להבטיח שתיאספנה תגובות ממדגם מייצג של נבחנים ברמה השנתית
 - לעיתים נדירות, עלולים להיות פריטים "בעייתיים", שקשה לאמוד את הפרמטרים שלהם, דבר שעלול להשפיע גם על הדיוק של הפרמטרים האחרים
- ✓ התוצאות מראות כי התבססות על מדגם של כ- 500-1000 נבחנים המייצגים את כלל הנבחנים יכולה להניב אומדנים דומים לאלה המתקבלים כאשר משתמשים בתגובות כלל הנבחנים

תודה!

