

תכנון ושימוש בסימולציות מבוססות IRT

נעם כאהן וצור קרליץ



מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
المركز القطري للامتحانات والتقييم
מיסודן של האוניברסיטאות בישראל

מהי סימולציה?

- סימולציה היא הדמיה של מצב אמתי על ידי שימוש במודל (דגם של המצב האמתי).
- מטרת הסימולציה היא לייצג מאפיינים מסוימים של המצב האמתי על מנת לאפשר למשתמשים לבחון תרחישים אפשריים.



סימולציה של מבחן לפי IRT

- רוצים לקבל סט של נתוני מבחן בלי להעביר מבחן בפועל
- בסימולציה יוצרים סט של נתוני מבחן על בסיס מודל המקשר בין מאפייני הנבחנים ומאפייניי הפריטים לבין הביצוע במבחן (Rasch, 2PL, PCM...).
- לאחר יצירת סט הנתונים, ניתן לאמוד את הפרמטרים של המודל ולבחון את טיב האמידה.
- בספרות ה-IRT נעשה שימוש נרחב בסימולציות למטרות תאורטיות או פרקטיות.

מתי נשתמש בסימולציה?

ככלל, נשתמש בסימולציות כשקשה לקבל את המידע בדרכים אחרות:

- כשאיסוף הנתונים למחקר המבוקש הוא יקר/מורכב מדי – לדוגמה, רוצים לדעת איך התפלגויות שונות של יכולות הנבחנים ישפיעו על טיב אמידת המודל.
- כשנרצה לבודד אפקט ספציפי – לדוגמה, בדיקת ההשפעה של מספר הפריטים במבחן בנפרד מטיב ההבחנה שלהם.
- כשצריך להשוות דיוק של שיטות שונות לניתוח הנתונים – אומדני פרמטרים ביחס לערכם האמתי.
- מכיוון שבסימולציה הפרמטרים של הנבחנים והפריטים ידועים מראש, אפשר להעריך את מידת הטעות המשויכת לאומדנים שאנו מפיקים.

מה ניתן לבדוק בעזרת סימולציות?

סימולציה מדמה ביצוע של ניתוח נתונים ואפשר לבדוק באמצעותה מאפיינים שונים:

- תפעוליות הקוד – זמן ריצה, מבנה התוצרים, איתור באגים
- איכות הניתוח – דיוק התוצרים
- מהו גודל המדגם ומהו מספר הפריטים שנצטרך במערך המחקר כדי לקבל אומדנים מדויקים
- עמידות בתנאים שונים – חסינות התהליך לסטיות מהנחות הבסיס
- בחינת שיטות כיוול שונות במצבי עולם שונים (רעש בנתונים, הבדלי קושי בין נוסחים, שינוי ביכולות הנבחנים)

איך בונים סימולציה?

הגדרת גורמים ורמות

- גורם – מאפיין שאנחנו רוצים לבחון את ההשפעה שלו
- רמה – מקרה ספציפי של הגורם
- לדוגמה, לגורם "גודל מדגם" יכולות להיות שלוש רמות: 100, 500 ו-1000.
- השילוב של גורמים ורמות יוצר את התנאים הנבדקים בסימולציה

בחירת מודל

- בתוך כל תנאי במערך עלינו להפיק נתונים מתאימים בהתבסס על מודל IRT ועל ערכי הפרמטרים שלו.
- המודל שנבחר ישמש בחישוב ההסתברות לתשובה נכונה לכל פריט ועבור כל נבחן.

איך בונים סימולציה? (המשך)

הפקת פרמטרים

קביעת הערכים של הפרמטרים במודל נעשית על פי אחת מהגישות הבאות:

- מניחים התפלגות מסוימת לכל פרמטר ודוגמים ממנה ערכים:
 - יכולת (תטא) וקושי הפריט (b) – התפלגות נורמלית
 - הבחנה (a) – התפלגות לוג-נורמלית
 - ניחוש (c) – התפלגות אחידה קטועה
- דוגמים ערכים מתוך התפלגות אמפירית אמתית שנצפתה במחקר קודם
- דוגמים ערכים מתוך מקבץ פרמטרים שהתקבלו בניתוח קודם של נתוני אמת

איך בונים סימולציה? (המשך)

יצירת נתונים

מסמלים סט נתונים עבור כל תנאי במערך המחקר, לפי שלושת השלבים הבאים:

1. חישוב $P(x_{ij}=1)$, הסתברות לתשובה נכונה של נבחן j בפריט i על בסיס מודל ה-IRT והפרמטרים שהופקו.
2. שליפת U_{ij} , מספר אקראי בין 0 ל-1.
3. יצירת נתון על ידי השוואת ההסתברות והמספר האקראי וקביעת התגובה לפריט לפי הכלל הבא:
אם $U \leq P$, אזי הערך בתא יהיה 1 (תשובה נכונה), ואם $U > P$, הערך יהיה 0.

איך בונים סימולציה? (המשך)

נמשיך כך לקבוע את כל הנתונים בטבלה

פריט 3 קושי = 2	פריט 2 קושי = 0.5	פריט 1 קושי = -2	
0	0	1	נבחן 1 יכולת = -1
0	1	1	נבחן 2 יכולת = 0
1	1	1	נבחן 3 יכולת = 1

$$P_{13} = \frac{\exp(\theta_1 - b_3)}{1 + \exp(\theta_1 - b_3)}$$

$$P_{32} = \frac{\exp(\theta_3 - b_2)}{1 + \exp(\theta_3 - b_2)}$$

יצירת נתונים – רפליקציות

- את תהליך יצירת הנתונים בכל תנאי מבצעים מספר פעמים על מנת ליצור מספר סטים של נתונים אקראיים מאותו סט פרמטרים.

למה זה חשוב?

- תהליך יצירת הנתונים נעשה בצורה אקראית ולכן עלולות להתקבל גם תוצאות קיצוניות. יצירת מספר רפליקציות מקטינה את הסיכוי שנקבל החלטות על בסיס תוצאות קיצוניות.
- רצוי להשתמש במספר גדול של רפליקציות

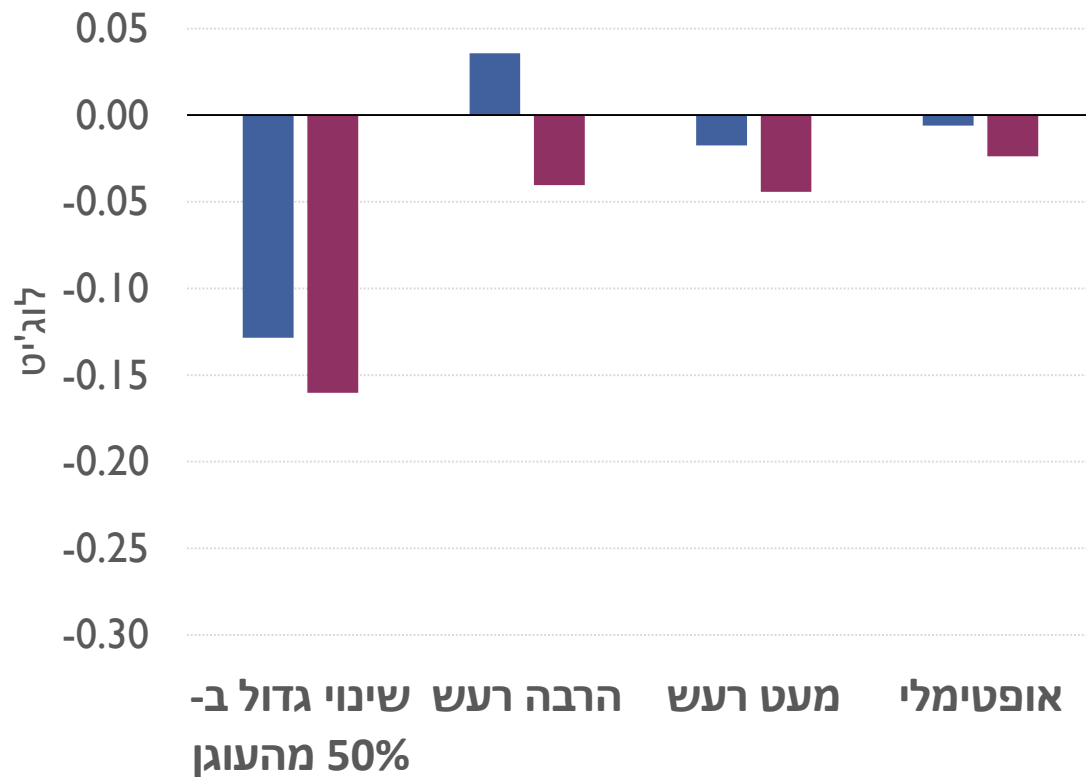
- מנתחים את הנתונים המסומלים בשיטה שאותה אנחנו רוצים לבדוק
 - בודקים את טיב התאמת המודל וקיום הנחות המודל
- בודקים את מידת הטעות באומדני הפרמטרים על ידי חישוב שלושה מדדים (Feinberg & Rubright, 2016):
 - הטיה (טעות שיטתית) - חישוב הפער הממוצע בין האומדן לערך האמתי.
 - טעות תקן (טעות מקרית) - סטיית התקן של האומדנים.
 - **RMSE** (טעות כוללת) - ממוצע ריבועי הפערים בין האומדנים לערכים האמתיים (ממוצע טעות ריבועית).

הדגמה: סימולציות לבדיקת שיטות כיול במצבים שונים

- מערך המחקר כלל ארבעה מצבים:
 - **אופטימלי** – הנתונים מייצגים במדויק את הפרמטרים מהם הופקו
 - **מעט רעש** – 5% מהנתונים נוצרו באופן אקראי (לא לפי מודל ה-IRT)
 - **הרבה רעש** – 25% מהנתונים נוצרו באופן אקראי
 - **שינוי גדול** – ב-50% מפריטי העוגן יש הבדל של 0.5-1.0 לוג'יט בין הנוסחים
- על כל מצב הופעלו שתי שיטות כיול:
 - **קיבוע פרמטרים** – מקבעים את הפרמטרים של פריטי העוגן בכל הנוסחים לערכים שנאמדו בנוסח המוביל
 - **Mean-Sigma** – מחשבים מקדמים לטרנספורמציה לינארית של הפרמטרים של פריטי העוגן בין נוסחים

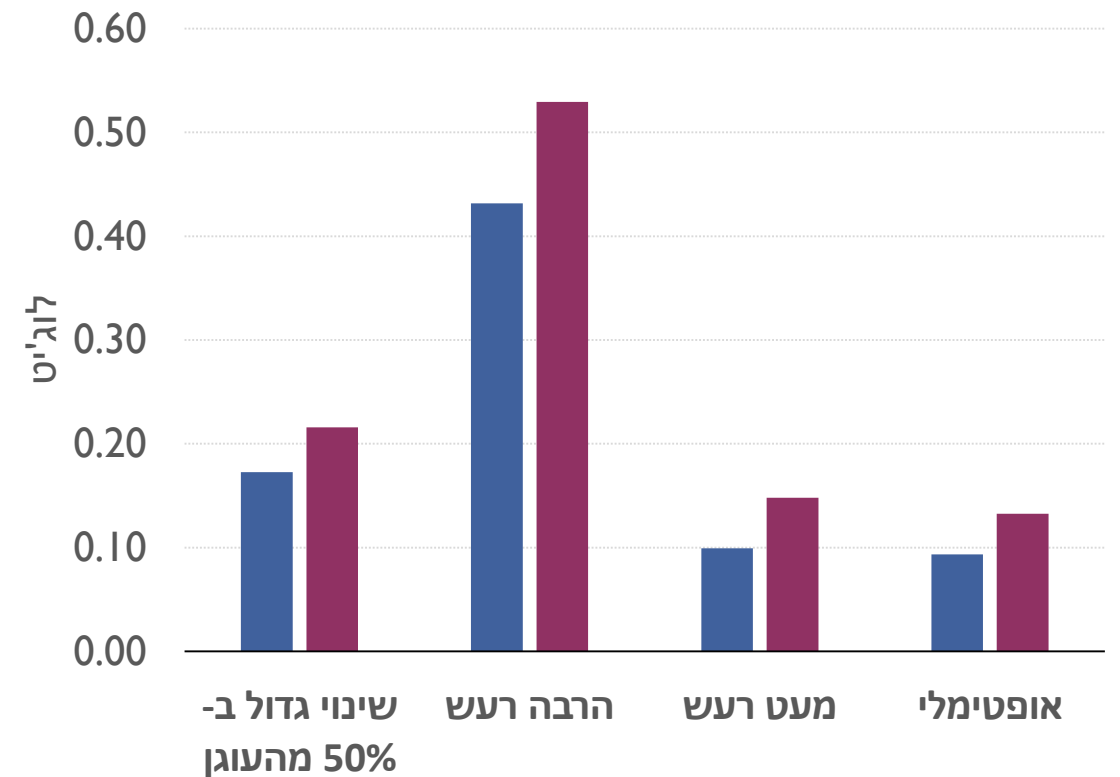
הדגמה: תוצאות עבור פרמטר הקושי

הטיה בפרמטר B



■ קיבוע ■ mean-sigma

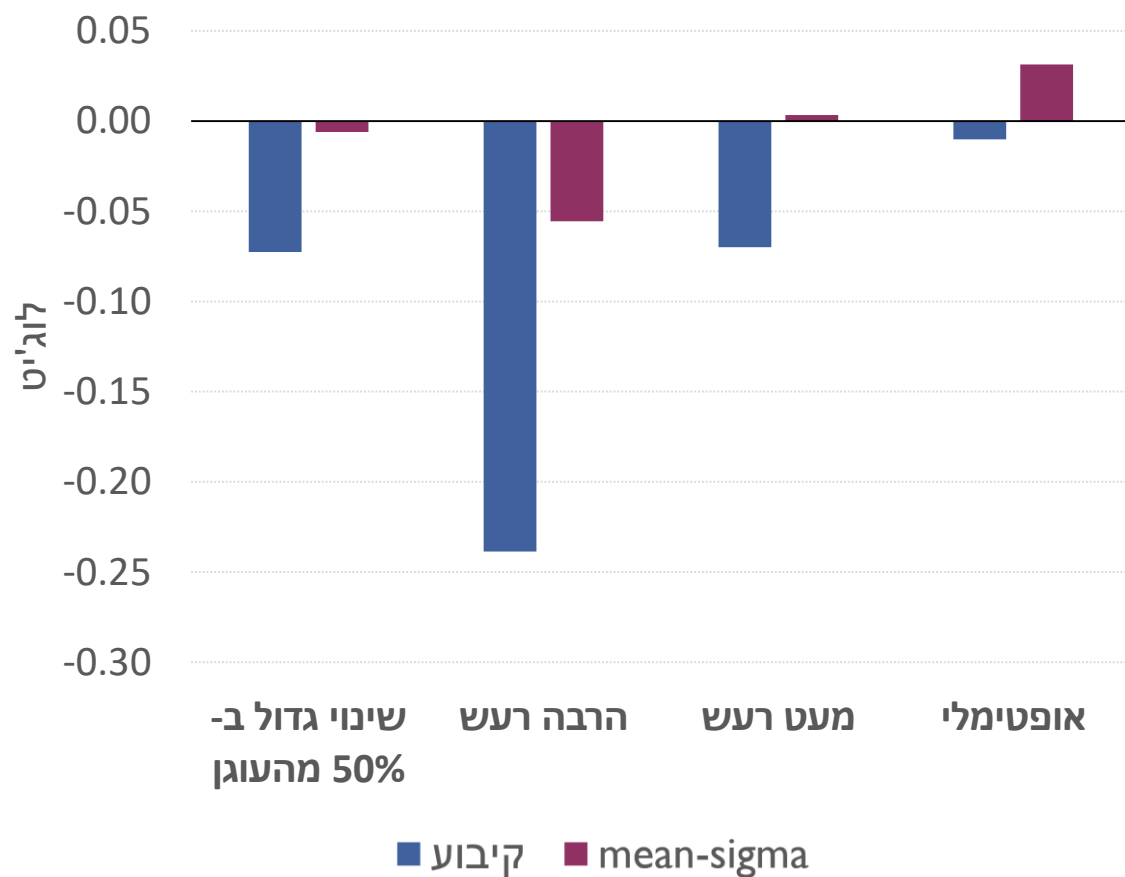
RMSE בפרמטר B



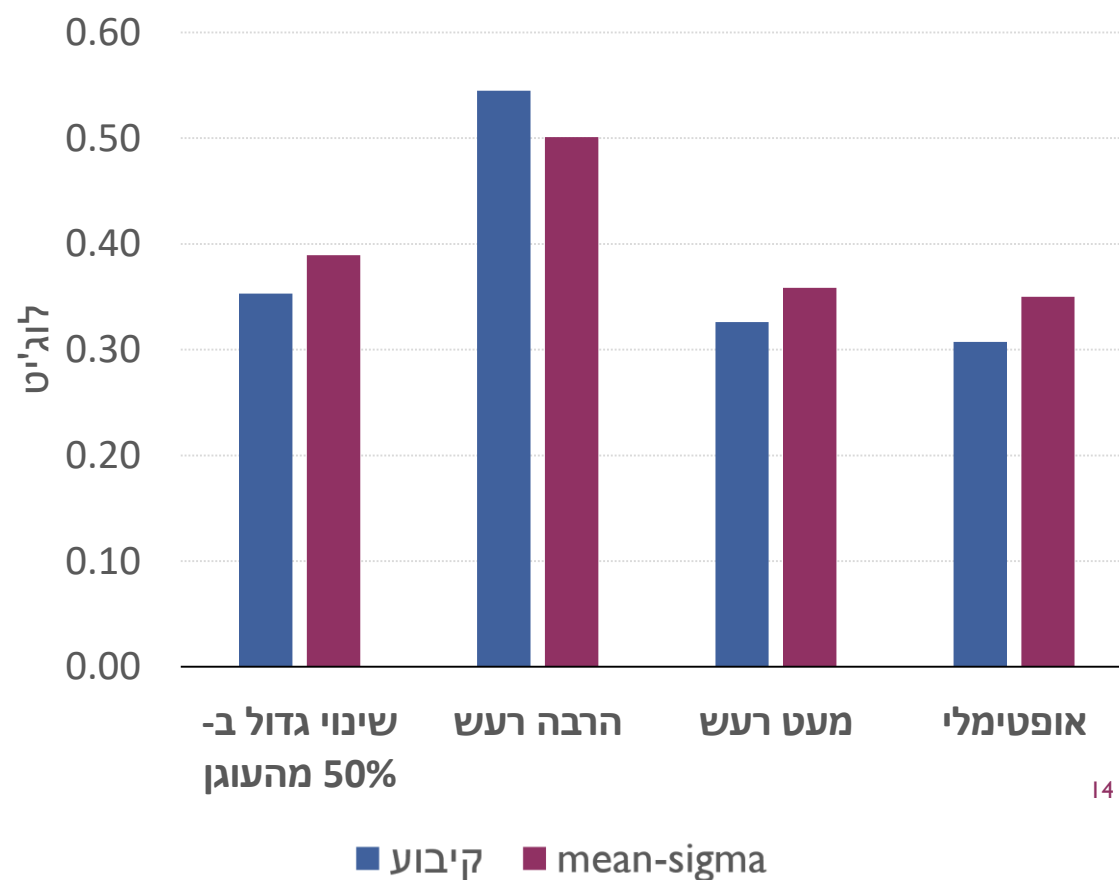
■ קיבוע ■ mean-sigma

הדגמה: תוצאות עבור אומדני היכולת

הטיה ביכולת θ



RMSE ביכולת θ



- נשתמש בסימולציות למטרות תאורטיות ופרקטיות
- היתרון הגדול: ידיעת ערכי הפרמטרים האמיתיים
- מגדירים גורמים ורמות, בוחרים מודל, מפיקים פרמטרים ומסמלצים נתונים – מספר רפליקציות
- מנתחים ומפיקים את המדדים המקובלים: הטויה, טעות תקן ו-RMSE
- על סמך הממצאים מקבלים החלטה לגבי הפעולה שנעשה במצב האמיתי



תודה על ההקשבה!

noam@nite.org.il