דוח המרכז

402

השוואה בין שתי שיטות (Bootstrap BCa והשלמות מרובות) לאמידת רווח סמך למקדם מתאם שחושב במדגם שעבר ברירה מוקדמת

אילה כהן

אטי דובא

טטיאנה אומנסקי

תמר קנת-כהן

יולי 2014

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

דוח מרכז 402

ISBN:978-965-502-183-7

© כל הזכויות שמורות
מרכז ארצי לבחינות ולהערכה

© All rights reserved
NITE

השוואה בין שתי שיטות (Bootstrap BCa והשלמות מרובות)
לאמידת רווח סמך למקדם מתאם
שחושב במדגם שעבר ברירה מוקדמת

אילה כהן, אטי דובא, טטיאנה אומנסקי
הטכניון – מכון טכנולוגי לישראל

תמר קנת-כהן
מרכז ארצי לבחינות ולהערכה (מאל"ו)

יולי 2014

המחקר נעשה במימון קרן המחקר של מאל"ו

תוכן עניינים

תקציר.. 3

הקדמה... 4

שיטה.. 6

תנאי הסימולציה.. 6

הדגימה לסימולציה.. 8

תהליך החישוב של רווח סמך בשיטת Bootstrap BCa....................................... 8

תהליך החישוב של רווח סמך בשיטת השלמות מרובות...................................... 10

תוצאות ומסקנות... 12

רשימת המקורות... 17

תקציר

מחקר זה משווה בין שתי שיטות לאמידת רווח הסמך למקדם מתאם בין חזאי לקריטריון באוכלוסייה ("אוכלוסיית המועמדים") שחושב במדגם שעבר ברירה מוקדמת ("מדגם המתקבלים"). תהליך הברירה שיושם במחקר הוא המקרה הדו-משתני שבו החזאי שבתוקף הניבוי שלו מתעניינים הוא זה ששימש בברירה.

שתי השיטות לחישוב רווח הסמך – Bootstrap BCa והשלמות מרובות – נבדלות זו מזו הן בחישוב האומד הנקודתי של מקדם המתאם באוכלוסייה והן בחישוב רווח הסמך סביבו. בשיטת ה-Bootstrap BCa האומד הנקודתי מחושב על פי הנוסחה המוכרת לתיקון לקיצוץ תחום ורווח הסמך מתבסס על האחוזונים הקיצוניים של המתאם המתוקן שהתקבלו במדגמי Bootstrap. בשיטת ההשלמות המרובות יוצרים עבור כל ערך חסר מספר ערכים משלימים. כך מתקבלים מספר אוספים של נתונים ללא ערכים חסרים. על בסיסם מחשבים את האומד הנקודתי, את שגיאת התקן של האומד ואת רווח הסמך. רווח הסמך בשיטת ה- Bootstrap BCa הושווה בעבר לשיטות אחרות ונמצא עדיף עליהן. חישוב רווח הסמך בשיטת ההשלמות המרובות ובפרט השוואתו לרווח המתקבל בשיטת ה-Bootstrap BCa הוא בבחינת חידוש של מחקר זה. הקריטריון העיקרי לבדיקת איכות רווח הסמך היה אחוז המקרים שהיו מחוץ לרווח הסמך המיועד (רמת סמך של 0.95).

איכותם של רווחי הסמך שהתקבלו בשתי השיטות נבדקה באמצעות סימולציות שנבדלו זו מזו בגודל אוכלוסיית המועמדים, ביחס הברירה ובמקדם המתאם באוכלוסייה. המדגמים לסימולציות התקבלו על ידי דגימה עם חזרות משלוש התפלגויות אמפיריות שונות, שהתקבלו עבור משתנה הברירה בטכניון – התפלגות הדומה לנורמלית, התפלגות עם אסימטריה שלילית והתפלגות הנוטה לאחידה.

ממצאי המחקר מעידים שהאומדים הנקודתיים, בשתי השיטות, מוטים כלפי מטה. ההטיה היא בדרך כלל הקטנה ביותר עבור ההתפלגות הדומה לנורמלית והגדולה ביותר עבור ההתפלגות הנוטה לאחידה, בפרט כאשר גודל מדגם המתקבלים קטן. כמו-כן נמצא שבשתי השיטות שונות האומדים גדולה, בפרט כאשר גודל מדגם המתקבלים קטן. הממצא העיקרי של המחקר, שנוגע לאיכות רווחי הסמך, מלמד שבכל ההתפלגויות, בשיטת ההשלמות המרובות רווח הסמך המתקבל משקף רמת ביטחון נמוכה מהרמה המיועדת. ממצא זה קיים גם בשיטת ה-Bootstrap BCa כאשר ההתפלגות אינה נורמלית, אולם הפער בין הרצוי למתקבל אינו כה גדול כמו בשיטת ההשלמות המרובות.

מסקנת המחקר היא ששיטת ה-Bootstrap BCa לאמידת רווח סמך עדיפה על שיטת ההשלמות המרובות.

הקדמה

במחקרי תוקף להערכת יעילות מבחני מיון מקובל להשתמש במקדם המתאם בין החזאים לקריטריון. מחקרים אלה מתבצעים לרוב במדגמים אשר עברו ברירה מוקדמת, שמקטינה את הטרוגניות המדגם, ולפיכך את מקדם המתאם שמחושב בהם לעומת המתאם בקרב המועמדים. במחקר הנוכחי אנו דנים במקרה שבו יש שני משתנים: X, החזאי שבתוקף שלו אנו מתעניינים ושלפיו נעשתה הברירה ו-Y, הקריטריון. בספרות מקובל להתייחס למקרה זה במינוח direct (explicit) range restriction. לא נתייחס ל-indirect range restriction, שבו ברירת המועמדים נעשית על בסיס משתנה שלישי Z, וקיים קשר בין Z ל-X ובין Z ל-Y. מרבית המחקרים נעשו על המקרה הראשון, ובו יתמקד גם מחקר זה.

הנוסחה המקובלת לתיקון בגלל הקיצוץ, פותחה על ידי תורנדייק (Thorndike, 1949), ומוכרת כ-Thorndike's Case 2. ההנחות עבורה הן שהקשר בין החזאי לקריטריון הוא ליניארי (ליניאריות), ושהשונות המותנית של הקריטריון היא קבועה (הומוסקדסטיות).

כמו כן, מניחים שידוע אומד שונות החזאי במדגם, שגודלו N, אשר לא עבר ברירה. נסמן אומד זה ב- S_X^2. נסמן ב- s_X^2 את אומד שונות החזאי במדגם, שגודלו n, שעבר ברירה, ב-r את אומד המתאם במדגם שעבר ברירה, וב- r_c את אומד המתאם לאחר התיקון. נוסחת התיקון היא :

$$r_c = \frac{S_X r}{\sqrt{S_X^2 r^2 + s_X^2 - s_X^2 r^2}} \quad (1)$$

נציין שגם כאשר הנחות הליניאריות וההומוסקדסטיות מופרות, השימוש בנוסחה (1) לאמידת מקדם המתאם ρ עדיף על השימוש באומד ללא התיקון (Gulliksen, 1950; Levin, 1972; Lord & Novick, 1968).

עד כה התייחסנו לאומד הנקודתי, r_c. ניתן למצוא בספרות עיסוק רב באיכות האמידה הנקודתית במצבים שונים. אלא שללא הערכת שגיאת האומד הנקודתי, אין זה סביר לקבל החלטות מעשיות. לצרכים מעשיים יש צורך ברווח סמך עבור המתאם. בכך עוסק המחקר הנוכחי.

במחקר קודם (כהן, דובא, אומנסקי וקנת-כהן, 2013) הוצגו חמישה רווחי סמך, כאשר בארבעת הרווחים הראשונים חישוב האומד הנקודתי r_c היה לפי הנוסחה המקובלת לתיקון עקב קיצוץ התחום (נוסחה (1)). שני הרווחים הראשונים התבססו על הקירוב הנורמלי, כאשר בראשון האומד לשגיאת התקן של האומד התבסס על פיתוחים תיאורטיים אסימפטוטיים, ובשני, האומד לשגיאת התקן של האומד היה סטיית התקן האמפירית של

r_c^*, שהם האומדים המתוקנים המתקבלים ממדגמי Bootstrap. גם הרווחים השלישי והרביעי נבנו על בסיס מדגמי Bootstrap, אך בניגוד לרווח השני, הם לא נבנו כרווחים סטנדרטיים, אלא גבולותיהם היו האחוזונים הקיצוניים של r_c^*. ההבדל בין הרווח השלישי והרביעי הוא שברביעי נוסף תיקון להטיה של r_c^* למקרה שההתפלגות של r_c^* סביב הערך של r_c אינה סימטרית. צ'אן וצ'אן (Chan & Chan, 2004) השוו בין ארבעת רווחי הסמך הללו. רווח הסמך בשיטה הרביעית הידוע בשם Bootstrap BCa (Bias-corrected and accelerated interval) נמצא כטוב ביותר מבין הארבעה במונחי הסתברות הכיסוי.

ברווח הסמך החמישי שהוצג במחקר של כהן ואחרים חישוב האומד הנקודתי נעשה באמצעות השלמת נתונים, ובהתאם נבנה גם הרווח סביבו. האפשרות לאמוד את מקדם המתאם למדגם עם קיצוץ תחום על ידי השלמת חלק הנתונים החסר (imputation) הוצעה בעבר על ידי מנדוזה (Mendoza, 1993). על בסיס רעיון זה, אמדו וייברג וסנדסטרום (Wiberg & Sundström, 2009) את מקדם המתאם (ללא חישוב רווחי סמך) עבור אוסף הנתונים שלהן. הן אמדו לפי השיטה המקובלת (נוסחה (1)), וכמו כן בשיטה של השלמת הנתונים החסרים על ידי שימוש באלגוריתם הידוע בשם EM. זו פרוצדורה איטרטיבית המשמשת בהרבה בעיות שבהן יש צורך באמידה סטטיסטית ולא ניתן לקבל פתרון סגור (Dempster, Laird, & Rubin, 1977; Little & Rubin, 2002). בכל צעד בפרוצדורה זו יש שני שלבים: הראשון E (Expectation) והשני M (Maximization). הצעדים נעשים עד להתכנסות. במאמרן ציינו החוקרות שיש צורך במחקר נוסף על יישום השיטה של השלמת נתונים, מאחר שהיא לא נחקרה ולכן לא ברור אם היא עדיפה על השיטה המקובלת.

במחקר של כהן ואחרים (2013) הוכנס שיפור בשיטת ההשלמות שבה השתמשו וייברג וסנדסטרום. וייברג וסנדסטרום ביצעו השלמה יחידה. כלומר, במקום כל נתון חסר הוצב ערך ש"הושלם". שיטה זו אינה משקפת כראוי את אי הוודאות שבאמידת הערך החסר, ולכן ידוע שעלולים להתקבל ממנה אומדים מוטים. במקום זה, נעשה במחקר של כהן ואחרים, שימוש בשיטה של השלמות **מרובות** (Multiple Imputation). בשיטה זו, עבור כל ערך חסר יוצרים מספר ערכים משלימים, ולכן מתקבלים מספר אוספים של נתונים שכולם **ללא** ערכים חסרים. כל אחד מהם משמש לאמידת הפרמטר המבוקש. מיצוע כל האומדים נותן אומד אחד. כמו כן, על בסיס האומדים המרובים מחושבת גם שגיאת התקן של האומד ובהמשך רווח הסמך. ההשלמות המרובות נעשות בעזרת סימולציה של מדגם מקרי. נציין שמאחר שהתפלגות מקדם מתאם אינה נורמלית, מקובל בבניית רווחי סמך ובהסקה על מקדם המתאם להשתמש בהתמרה של מקדם המתאם (הטרנספורמציה של פישר). לאחר האמידה הנקודתית ובניית רווח הסמך על ההתמרה, מתמירים בחזרה ומציגים את הממצאים על מקדם המתאם. כך נעשה גם במחקר של כהן ואחרים (2013).

במחקרם של כהן ואחרים (2013) יושמו חמש השיטות על נתוני קבלה של הטכניון, ובנוסף בוצעו סימולציות בהיקף מוגבל. הממצאים הראו על הבדלים ברוחב רווחי הסמך שהתקבלו לפי השיטות השונות. הקריטריון החשוב לבדיקת איכות רווח הסמך – הסתברות הכיסוי – לא נבדק באותו מחקר. רווח סמך המיועד להיות ברמת סמך $(1-\alpha)$ יחשב נכון, אם בסיכוי של $(1-\alpha)$ הוא מכסה את הפרמטר. כאמור, צ'אן וצ'אן (Chan & Chan, 2004) מצאו שעל-פי קריטריון זה, שיטת ה-Bootstrap BCa היא הטובה מבין ארבע השיטות הראשונות. השאלה שנותרה פתוחה הייתה האם השיטה החמישית, אשר מעט יחסית ידוע על תכונותיה, עדיפה על שיטת ה-Bootstrap BCa. מטרת המחקר הנוכחי הייתה לענות על שאלה זו.

מחקר זה מתמקד, אם כן, בהשוואה בין שיטת ה-Bootstrap BCa לשיטת ההשלמות המרובות בהסתברות הכיסוי של רווח הסמך. על ידי סימולציות בהיקף רחב, בדקנו לכל אחת משתי השיטות ותחת תנאים (גורמים) שונים את תפקודן על פי קריטריון זה.

הסימולציות שעשינו דומות לסימולציות שבוצעו במחקר של צ'אן וצ'אן (Chan & Chan, 2004). בדומה להם בחנו את הגורמים: גודל מדגם, יחס ברירה, מקדם המתאם ρ, וההתפלגות של X. כמו כן, בדומה למגבלות אשר צוינו במחקרם, גם מחקר זה מכסה רק חלק מטווחי הגורמים השונים. כל סימולציה דורשת הרבה זמן ומספר האפשרויות הוא כמובן אינסופי. בנוסף, לא בדקנו את כל הצירופים של התנאים השונים: הצירופים שבדקנו של גודל מדגם שלא עבר ברירה ויחס הברירה נבחרו בהתאם לנתוני הטכניון. לכן, לדוגמה, צירופים כמו גודל מדגם של 80 ויחס ברירה של 0.1 לא נכללו כי לא היה בהם עניין מעשי.

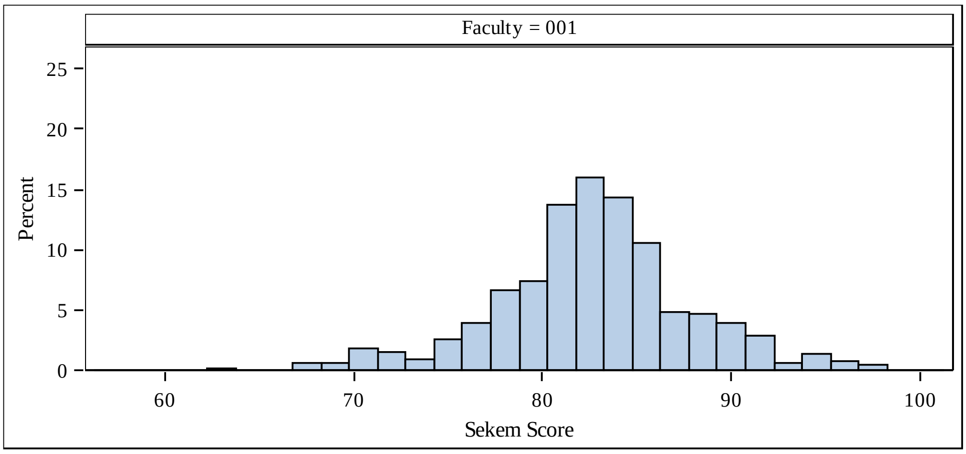
שיטה

תנאי הסימולציה

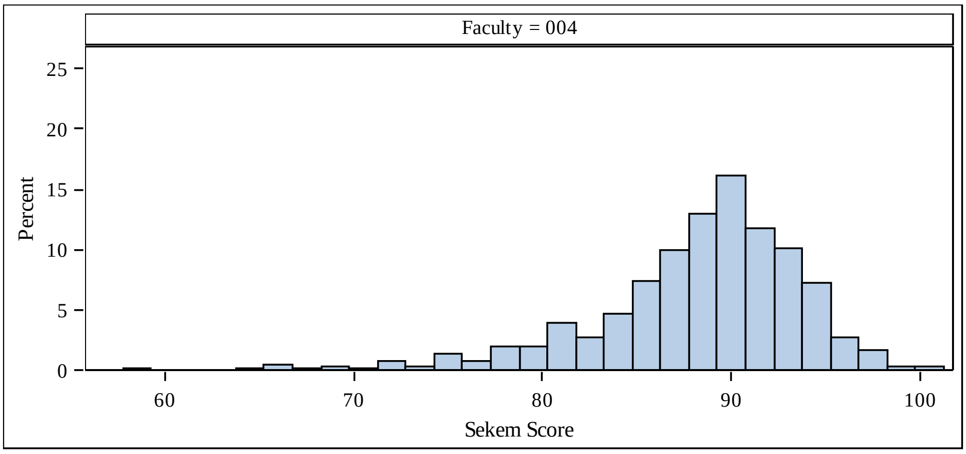
בסימולציות בדקנו 18 מצבים (ראו פירוט להלן) שנבדלו זה מזה בגודל המדגם, ביחס הברירה, ובגובה מקדם המתאם. לכל צירוף של תנאים בדקנו בעזרת הסימולציה את רווח הסמך שהתקבל בכל אחת משתי השיטות. המדגמים שיצרנו לסימולציות התקבלו על ידי דגימה עם חזרות משלוש התפלגויות אמפיריות שונות, שהתקבלו לנתוני הסכם בשלוש פקולטות בטכניון. שלוש ההתפלגויות הן: התפלגות הדומה להתפלגות נורמלית, התפלגות אסימטרית שלילית והתפלגות הנוטה להתפלגות אחידה.

בתרשימים 1 עד 3 שלהלן מוצגות שלוש ההתפלגויות האמפיריות של נתוני הסכם שמהן דגמנו (עם חזרות).

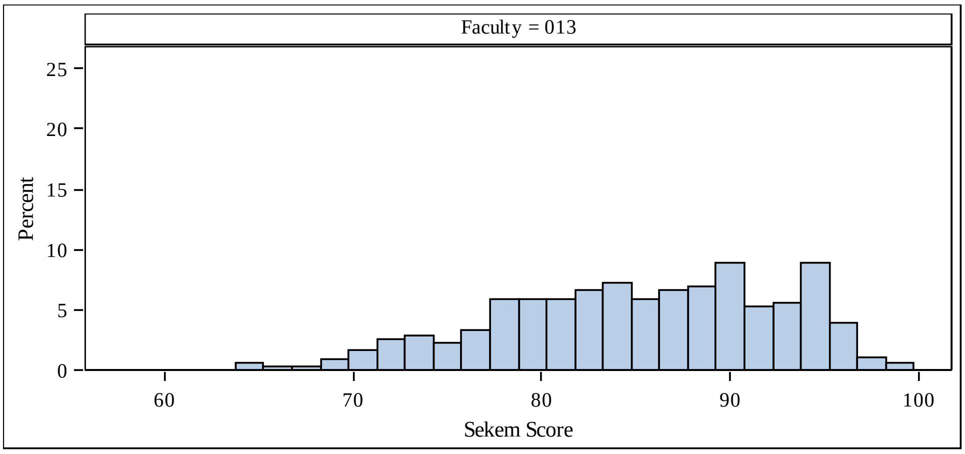
<u>תרשים 1</u>: התפלגות אמפירית הדומה להתפלגות נורמלית (N=603)



<u>תרשים 2</u>: התפלגות אמפירית עם אסימטריה שלילית (N=891)



<u>תרשים 3</u>: התפלגות אמפירית הנוטה להתפלגות אחידה (N=420)



כאמור, עבור כל אחת משלוש ההתפלגויות בדקנו צירופים שונים של גודל מדגם (N) ויחס ברירה, המגדיר את גודל המדגם אחרי הברירה (n):

- N=80, יחס הברירה 0.3, n=24
- N=80, יחס הברירה 0.5, n=40
- N=400, יחס הברירה 0.1, n=40
- N=400, יחס הברירה 0.5, n=200
- N=900, יחס הברירה 0.1, n=90
- N=900, יחס הברירה 0.3, n=270

לכל אחד מ-6 הצירופים שלעיל נבדקו שלושה ערכים של מתאם ρ: 0.4, 0.6, 0.8. סה"כ נעשו 3*6*3=54 סימולציות שונות. בכולן רמת רווח הסמך הייתה כמקובל 0.95.

הדגימה לסימולציה

המשימה הייתה לדגום מדגם בגודל N מהתפלגות דו-ממדית עם מתאם ρ והתפלגות שולית נתונה של X (אחת משלוש ההתפלגויות האמפיריות שתיארנו).
ראשית תקננו את ההתפלגות של X על ידי חלוקה בסטיית התקן והזזה, כך שהתפלגות X היא עם ממוצע 0 ושונות 1. את ערכי X של המדגם דגמנו עם חזרות מתוך ההתפלגות של הנתונים. ליצירת ערכי Y דגמנו ε מהתפלגות נורמלית עם תוחלת 0 ושונות $1-\rho^2$. המשתנה Y הוגדר כך: $Y=\rho X+\varepsilon$.

מאחר ש-X וגם ε הם משתנים עם תוחלת 0, גם Y הוא עם תוחלת 0. ומאחר ש-X ו-ε בלתי תלויים אזי $\mathbf{Var(Y)} = \rho^2 + \mathbf{1} - \rho^2 = 1$. המתאם בין X ל-Y שווה לשונות המשותפת ביניהם השווה ל-ρ.
כל אחת מ-54 הסימולציות השונות הורצה 1,000 פעמים. הסימולציות בוצעו בעזרת חבילת התוכנה SAS.
להלן תיאור של שתי השיטות לבניית רווח הסמך שאותן בדקנו.

תהליך החישוב של רווח סמך בשיטת Bootstrap BCa

על התכונות הסטטיסטיות של שיטה זו לבניית רווח סמך ניתן לקרוא למשל ב- DiCiccio & Efron, 1996. כאן נפרט רק את הפרטים הטכניים הרלוונטיים ליישום השיטה במחקר זה.
נסמן ב-(X_r, Y_r) את n נתוני המדגם המקוצץ (ה"מתקבלים") וב- X_u את (N-n) הערכים של X עבור המדגם של אלה שלא התקבלו. עבור ערכי X אלה אין ערכים מתאימים של Y.

תהליך האמידה בשיטת ה-Bootstrap מבוצע באופן הבא: תחילה על בסיס המדגם המקוצץ מחשבים את מקדם המתאם המתוקן לפי נוסחה (1). את מדגם ה-Bootstrap המקוצץ הראשון מקבלים כאשר דוגמים עם חזרות n זוגות מ-(X_r, Y_r). נסמן את המדגם המתקבל ב- $(X_r^{*(1)}, Y_r^{*(1)})$, ואת מקדם המתאם (ללא תיקון) ושונות X המחושבים על בסיס מדגם זה ב- $r^{*(1)}, s_x^{2*(1)}$ בהתאמה. בשלב השני דוגמים עם חזרות (N-n) ערכים מתוך X_u ומקבלים את מדגם ה-Bootstrap הראשון של אלה שלא התקבלו, אותו נסמן $X_u^{*(1)}$.

מדגם ה-Bootstrap הכולל הראשון של X מורכב משני החלקים $X^{*(1)} = \begin{pmatrix} X_u^{*(1)} \\ X_r^{*(1)} \end{pmatrix}$. על בסיסו מחשבים את השונות $S_x^{2*(1)}$. בעזרת נוסחה (1) ניתן לחשב את אומד מקדם המתאם המתוקן $r_c^{*(1)}$, בהינתן הערכים $r^{*(1)}, s_x^{2*(1)}, S_x^{2*(1)}$.

את אותם השלבים שנעשו לחישוב $r_c^{*(1)}$ על בסיס מדגם ה-Bootstrap הראשון מבצעים B פעמים ומקבלים B אומדי Bootstrap של האומד המתוקן: $r_c^{*(1)}, r_c^{*(2)}, ...r_c^{*(B)}$.

מסדרים לפי סדר עולה את B הערכים. נסמן סטטיסטי סדר זה ב :

$$r_c^*[1] \le r_c^*[2], ... \le r_c^*[B]$$

בשלב הראשון מחשבים את הערך $z = \Phi^{-1}\left[\frac{\#(r_c^{*(b)} < r_c)}{B}\right]$,

כאשר $\Phi^{-1}(\cdot)$ היא הפונקציה ההופכית של ההתפלגות הנורמלית הסטנדרטית.

$\left[\frac{\#(r_c^{*(b)} < r_c)}{B}\right]$ היא הפרופורציה של ערכי ה-Bootstrap שערכם קטן מהערך של האומד המקורי המתוקן שהתקבל על בסיס המדגם עבור מקדם המתאם ρ. הערך z מתאר למעשה את ההטיה החציונית של r_c^*. במקרים שבהם התפלגות ערכי ה-Bootstrap סימטרית סביב הערך r_c אזי $\left[\frac{\#(r_c^{*(b)} < r_c)}{B}\right]$ יהיה קרוב לחצי, והערך של z יהיה קרוב לאפס.

בשלב השני של בניית רווח סמך זה מחשבים את הערך

$$, a = \frac{\sum_{i=1}^{N}(r_{c(\cdot)} - r_{c(i)})^3}{6\left\{\sum_{i=1}^{N}(r_{c(\cdot)} - r_{c(i)})^2\right\}^{3/2}}$$

כאשר $r_{c(i)}$ מציין את ערך ה-jackknife של r_c המתקבל על ידי חישוב הערך של r_c בהשמטת התצפית ה-i-ית, ו- $r_{c(.)} = \frac{\sum_{i=1}^{N}(r_{c(i)})}{N}$ הוא ממוצע N אומדי ה-jackknife.

הערך a ידוע בשם תאוצה, כי הוא מתאר את קצב השינוי בשגיאת התקן של r_c ביחס לפרמטר האמתי ρ.

בהינתן שני ערכי התיקון z ו-a מחשבים את הערכים

$$\alpha_1 = \Phi\left[z + \frac{z + z_{\alpha/2}}{1 - a(z + z_{\alpha/2})}\right]$$

$$\alpha_2 = \Phi\left[z + \frac{z - z_{\alpha/2}}{1 - a(z - z_{\alpha/2})}\right]$$

כאשר $z_{\alpha/2}$ הוא האחוזון ה-100(α/2) של ההתפלגות הנורמלית הסטנדרטית.

רווח הסמך ברמת סמך של (1-α) יהיה $(r_c^*[l^*], r_c^*[u^*])$, כאשר $(l^* = B\alpha_1)$, $(u^* = B\alpha_2)$.

בכל המקרים שבהם l^*, u^* אינם מספרים שלמים, מעגלים למספר הקרוב ביותר המקיים $l' < l^*$ $u' > u^*$.

תהליך החישוב של רווח סמך בשיטת השלמות מרובות

נסמן ב-Q את הפרמטר שאותו רוצים לאמוד, וב- $\hat{Q}_i, \hat{W}_i$ את האומד של הפרמטר ואת האומד של שונות האומד בהתאמה המתקבלים מההשלמה ה-i-ית. מבצעים m השלמות,

והאומד הסופי שנסמן $\hat{Q}$ מחושב כממוצע: $\hat{Q} = \bar{Q} = \frac{\sum_{i=1}^{m} \hat{Q}_i}{m}$. נסמן ב-$\bar{W}$ את הממוצע של אומדי השונויות $\hat{W}_i$. זו השונות "בתוך ההשלמות". נסמן ב-B את השונות בין ההשלמות: $B = \frac{\sum_{i=1}^{m} (\hat{Q}_i - \bar{Q})^2}{m}$. אומד השונות של האומד $\hat{Q}$ יהיה הסכום $T = \bar{W} + (1+\frac{1}{m})B$.

הסטטיסטי $\frac{\hat{Q} - Q}{\sqrt{T}}$ מתפלג בקירוב התפלגות t עם דרגות חופש

$$\nu_m = (m-1)\left[1 + \frac{\bar{W}}{(1+\frac{1}{m})B}\right]^2$$

לכן רווח הסמך המתאים עבור הפרמטר Q ברמת סמך (1-2α) הוא

$$L = \hat{Q} - \sqrt{T}t_{\nu_m}^{1-\alpha} \le Q \le \hat{Q} + \sqrt{T}t_{\nu_m}^{1-\alpha} = U$$

רווח הסמך הזה הוא רווח סטנדרטי שבו משתמשים בהתפלגות t, ולמעשה $\sqrt{T} = SE$.

כאמור לעיל, מקובל להסקה סטטיסטית עבור ρ להשתמש בסטטיסטי שהוא הטרנספורמציה של פישר: $z(r) = \frac{1}{2}\ln\frac{1+r}{1-r}$. סטטיסטי זה מתפלג בקירוב נורמלית עם תוחלת $z(\rho)$ ושונות $\frac{1}{(n-3)}$. עבור אומד נקודתי $\hat{z}(r)$ ורווח סמך של (1-2α) הנבנה על בסיס אומד זה, מחשבים בעזרת הטרנספורמציה ההפוכה, $\tanh(z) = \frac{e^{2z}-1}{e^{2z}+1}$, את האומד הנקודתי המתאים $\hat{r}$ וכן את רווח הסמך למקדם המתאם: $(\frac{e^{2L}-1}{e^{2L}+1}, \frac{e^{2U}-1}{e^{2U}+1})$.

ההשלמות המרובות בוצעו על ידי שימוש בחבילת התוכנה SAS: בשלב ראשון בפרוצדורה MI, ובשלב שני, לקבלת האומד ורווח הסמך המתאים, בפרוצדורה MIANALYZE.

תוצאות ומסקנות

טבלאות 1-3 מציגות, עבור כל אחת משלוש ההתפלגויות של X, את הערך הממוצע של אומדי המתאם שהתקבלו ב-1,000 החזרות. הערכים מוצגים עבור שתי השיטות ועבור הפרמטרים השונים שנבחרו לגודל המדגם (N), גודל מדגם המתקבלים (n) וגובה המתאם. בסוגריים מוצגות סטיות התקן של 1,000 האומדים.

<u>טבלה 1</u>: ממוצעים וסטיות התקן האמפיריות (בסוגריים) של האומדים המתוקנים בשיטת Bootstrap BCa* (r_c) ושל האומדים לפי שיטת ההשלמות המרובות (r_{MI}) בהתפלגות הדומה לנורמלית

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		r_c	r_{MI}	r_c	r_{MI}	r_c	r_{MI}
80	24	.37(.27)	.36(.27)	.57(.21)	.56(.21)	.79(.11)	.78(.11)
80	40	.38(.19)	.38(.19)	.58(.14)	.58(.14)	.79(.07)	.79(.08)
400	40	.36(.29)	.35(.28)	.56(.22)	.54(.22)	.78(.11)	.77(.11)
400	200	.40(.08)	.39(.08)	.60(.06)	.59(.06)	.80(.03)	.80(.03)
900	90	.38(.19)	.37(.19)	.58(.14)	.57(.14)	.79(.06)	.79(.07)
900	270	.39(.08)	.39(.08)	.59(.06)	.59(.06)	.80(.03)	.80(.03)

* כלומר, על-פי נוסחה (1)

<u>טבלה 2</u>: ממוצעים וסטיות התקן האמפיריות (בסוגריים) של האומדים המתוקנים בשיטת Bootstrap BCa* (r_c) ושל האומדים לפי שיטת ההשלמות המרובות (r_{MI}) בהתפלגות אסימטרית שלילית

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		r_c	r_{MI}	r_c	r_{MI}	r_c	r_{MI}
80	24	.30(.44)	.28(.42)	.48(.38)	.46(.37)	.73(.25)	.70(.25)
80	40	.33(.32)	.32(.31)	.53(.26)	.52(.25)	.76(.14)	.75(.14)
400	40	.31(.42)	.29(.39)	.50(.35)	.48(.34)	.74(.21)	.72(.22)
400	200	.40(.14)	.39(.14)	.59(.10)	.59(.10)	.80(.05)	.79(.05)
900	90	.36(.29)	.34(.29)	.56(.22)	.54(.22)	.78(.11)	.77(.11)
900	270	.39(.14)	.38(.14)	.59(.10)	.58(.10)	.79(.05)	.79(.05)

* כלומר, על-פי נוסחה (1)

<u>טבלה 3</u>: ממוצעים וסטיות התקן האמפיריות (בסוגריים) של האומדים המתוקנים בשיטת Bootstrap BCa* (r_c) ושל האומדים לפי שיטת ההשלמות המרובות (r_{MI}) בהתפלגות הנוטה לאחידה

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		r_c	r_{MI}	r_c	r_{MI}	r_c	r_{MI}
80	24	.28(.49)	.26(.46)	.46(.43)	.43(.41)	.71(.29)	.68(.28)
80	40	.36(.28)	.35(.27)	.56(.21)	.54(.21)	.78(.11)	.77(.11)
400	40	.20(.60)	.18(.55)	.35(.56)	.32(.52)	.60(.45)	.56(.43)
400	200	.38(.13)	.38(.13)	.59(.09)	.58(.09)	.79(.04)	.79(.04)
900	90	.31(.47)	.29(.44)	.49(.40)	.46(.39)	.72(.26)	.70(.26)
900	270	.38(.17)	.37(.17)	.58(.12)	.57(.12)	.79(.05)	.79(.06)

* כלומר, על-פי נוסחה (1)

ניתן להיווכח שלהוציא מקרים ספורים שבהם ערכו של האומד הנקודתי שווה לערך במדגם לפני הברירה (6 מקרים עבור r_c ו-2 מקרים עבור r_{MI}), האומדים הנקודתיים מוטים כלפי מטה. ההטיה היא בדרך כלל הקטנה ביותר עבור ההתפלגות הדומה לנורמלית והגדולה ביותר עבור ההתפלגות הנוטה לאחידה, בפרט כאשר גודל המדגם המצומצם קטן. לדוגמה, עבור מתאם של 0.6, כאשר N=400 ו-n=40, אזי כאשר ההתפלגות של X קרובה לנורמלית ממוצעי האומד קרובים למדי לערך האמתי (0.56 ו-0.54 בשיטת Bootstrap BCa ובשיטת

ההשלמות המרובות בהתאמה), וכאשר ההתפלגות של X נוטה לאחידה הם נמוכים ממנו באופן משמעותי (0.35 ו-0.32 בהתאמה).

בהשוואת שתי השיטות על בסיס ממוצעי האומדים וסטיות התקן שלהם לא ניתן לראות הבדל משמעותי בין שתי השיטות. כאמור, בשתיהן יש הטיה קטנה כלפי מטה, פרט למקרים שצוינו לעיל. כמו כן, בשתיהן שונות האומדים גדולה, בפרט כאשר המדגם המצומצם קטן. לכן, לעתים עלולים לקבל אפילו אומדים שליליים. השונות הגדולה מעוררת שאלה לגבי מהימנות אומדים כמו אלה שהתקבלו בעבר עבור נתוני הטכניון. במחקר קודם (כהן ואחרים, 2013) הוצגו אומדים שהתקבלו עבור נתוני הקבלה של הטכניון לשנת 2009 לפי פקולטות. עבור הפקולטה למתימטיקה למשל, עם N=83 ו-n=16, היה ערך אומד המתאם המתוקן גבוה יחסית – 0.74. אולם, על בסיס הסימולציות במחקר זה ניתן לראות שאפילו אם בפועל הערך האמתי הוא רק 0.4, הרי עם גודלי מדגם כאלה יש סבירות לא קטנה לקבל ערך גבוה כמו 0.74.

הממצא החשוב העיקרי בהתאם למטרת מחקר זה נוגע לרווחי הסמך המתקבלים בשתי השיטות. טבלאות 4-6 מציגות, עבור כל אחת משלוש ההתפלגויות, את אחוז המקרים מתוך 1,000 החזרות שנפלו מחוץ לרווח הסמך. הרווח אמור לכסות ב-95% מההרצות את הפרמטר האמתי.

<u>טבלה 4</u>: אחוז המקרים שהיו מחוץ לרווח הסמך של 0.95 בשיטת Bootstrap BCa (*BCa*) ובשיטת ההשלמות המרובות (*MI*) בהתפלגות הדומה לנורמלית

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		BCa	*MI*	*BCa*	*MI*	*BCa*	*MI*
80	24	4.9	7.0	6.1	6.9	5.6	6.3
	40	4.9	5.0	4.6	5.2	5.2	5.5
400	40	5.5	8.0	5.4	6.7	4.9	5.8
	200	4.7	5.1	5.2	5.2	5.3	4.6
900	90	5.2	6.5	5.2	5.6	5.6	5.0
	270	5.4	6.0	5.6	5.7	5.0	5.0

<u>טבלה 5</u>: אחוז המקרים שהיו מחוץ לרווח הסמך של 0.95 בשיטת Bootstrap BCa (*BCa*) ובשיטת ההשלמות המרובות (*MI*) בהתפלגות אסימטרית שלילית

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		BCa	*MI*	*BCa*	*MI*	*BCa*	*MI*
80	24	7.2	9.6	6.7	8.6	5.6	6.6
	40	7.2	8.5	6.1	7.6	6.6	6.8
400	40	5.7	8.1	6.4	7.0	6.2	6.4
	200	5.8	5.9	5.8	5.9	5.3	6.1
900	90	5.7	7.5	5.9	7.1	5.3	5.6
	270	6.1	5.9	5.5	5.5	5.8	5.8

<u>טבלה 6</u>: אחוז המקרים שהיו מחוץ לרווח הסמך של 0.95 בשיטת Bootstrap BCa (*BCa*) ובשיטת ההשלמות המרובות (*MI*) בהתפלגות הנוטה לאחידה

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		BCa	*MI*	*BCa*	*MI*	*BCa*	*MI*
80	24	5.9	9.7	5.4	8.5	5.4	6.5
	40	4.5	6.3	4.6	5.4	5.0	4.7
400	40	7.1	13.0	6.4	12.0	6.4	11.0
	200	4.2	5.5	4.6	4.5	4.5	4.1
900	90	6.2	11.0	6.2	9.3	5.6	8.1
	270	4.4	4.9	5.0	4.7	4.4	4.1

במחקר הקודם, שבו נבנו רווחי הסמך לאומדי המתאם עבור נתוני הטכניון, התקבלו רווחים צרים יותר בשיטת ההשלמות המרובות. אולם לא ניתן היה לדעת אם הם טובים יותר או פחות מאילו שהתקבלו בשיטות האחרות שכן לא היה ידוע הערך האמתי של הפרמטר עבורו נבנה הרווח. ממצאינו עתה מראים שהרווח הצר מהווה מגרעת ובכך מצביעים על חיסרון השיטה של ההשלמות המרובות. ברוב המקרים המוצגים בטבלאות 4-6, בשיטת ההשלמות המרובות אחוז רווחי הסמך שאינו מכסה את הפרמטר הנכון אינו 5% אלא גדול מזה. כלומר, רווח הסמך המתקבל בשיטת ההשלמות המרובות צר מדי ומשקף רמת ביטחון נמוכה מהרמה שהוא אמור לתת. גם בשיטת ה-Bootstrap BCa, כאשר ההתפלגות אינה נורמלית, הרווח המתקבל משקף רמת ביטחון נמוכה מהמבוקשת, אולם הפער בין הרצוי למתקבל אינו כה גדול כמו בשיטת ההשלמות המרובות. לדוגמה, עבור מתאם של 0.4, בתנאי של N=400 ו-n=40, כאשר ההתפלגות היא אסימטרית שלילית, אזי במקום להיות

0.95, רמת הסמך בשיטת ה-Bootstrap BCa היא 0.943 ובשיטת ההשלמות המרובות 0.919.

טבלאות 9-7 מציגות, עבור כל אחת משלוש ההתפלגויות, את הממוצעים וסטיות התקן של רוחבי רווחי הסמך ב-1,000 החזרות.

<u>טבלה 7</u>: ממוצעים וסטיות תקן (בסוגריים) של רוחב רווח הסמך של 0.95 בשיטת Bootstrap BCa (*BCa*) ובשיטת ההשלמות המרובות (*MI*) בהתפלגות הדומה לנורמלית

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		BCa	*MI*	*BCa*	*MI*	*BCa*	*MI*
80	24	1.03(.27)	0.94(.23)	0.86(.31)	0.76(.26)	0.51(.28)	0.43(.21)
80	40	0.76(.18)	0.73(.15)	0.59(.19)	0.56(.16)	0.32(.13)	0.30(.11)
400	40	1.03(.26)	0.96(.24)	0.85(.30)	0.78(.28)	0.05(.26)	0.43(.23)
400	200	0.34(.05)	0.34(.04)	0.25(.04)	0.24(.03)	0.12(.02)	0.12(.02)
900	90	0.73(.16)	0.71(.15)	0.56(.17)	0.53(.16)	0.28(.11)	0.26(.10)
900	270	0.32(.04)	0.31(.03)	0.23(.04)	0.22(.03)	0.11(.02)	0.11(.02)

<u>טבלה 8</u>: ממוצעים וסטיות תקן (בסוגריים) של רוחב רווח הסמך של 0.95 בשיטת Bootstrap BCa (*BCa*) ובשיטת ההשלמות המרובות (*MI*) בהתפלגות אסימטרית שלילית

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		BCa	*MI*	*BCa*	*MI*	*BCa*	*MI*
80	24	1.38(.34)	1.29(.36)	1.25(.41)	1.15(.42)	0.93(.48)	0.81(.45)
80	40	1.07(.27)	1.03(.26)	0.90(.32)	0.85(.31)	0.55(.30)	0.50(.27)
400	40	1.31(.31)	1.25(.33)	1.18(.38)	1.09(.40)	0.84(.44)	0.73(.42)
400	200	0.53(.09)	0.52(.09)	0.39(.09)	0.37(.09)	0.19(.05)	0.18(.05)
900	90	1.01(.23)	0.98(.24)	0.83(.28)	0.79(.28)	0.47(.25)	0.43(.23)
900	270	0.55(.10)	0.54(.10)	0.40(.10)	0.39(.10)	0.18(.05)	0.18(.05)

<u>טבלה 9</u>: ממוצעים וסטיות תקן (בסוגריים) של רוחב רווח הסמך של 0.95 בשיטת Bootstrap BCa (*BCa*) ובשיטת ההשלמות המרובות (*MI*) בהתפלגות הנוטה לאחידה

N	*n*	*ρ=0.4*		*ρ=0.6*		*ρ=0.8*	
		BCa	*MI*	*BCa*	*MI*	*BCa*	*MI*
80	24	1.46(.34)	1.37(.38)	1.36(.41)	1.25(.45)	1.07(.51)	0.93(.51)
80	40	1.01(.24)	0.96(.23)	0.83(.28)	0.77(.27)	0.48(.24)	0.43(.21)
400	40	1.59(.37)	1.49(.45)	1.54(.42)	1.43(.49)	1.36(.53)	1.21(.59)
400	200	0.50(.08)	0.49(.08)	0.36(.08)	0.35(.08)	0.17(.04)	0.16(.04)
900	90	1.38(.33)	1.32(.38)	1.26(.41)	1.19(.45)	0.96(.49)	0.85(.49)
900	270	0.65(.12)	0.64(.12)	0.48(.13)	0.46(.12)	0.22(.07)	0.21(.07)

ניתן לראות שבכל המקרים (למעט 4 מקרים שבהם הממוצעים שווים) רוחב רווח הסמך בשיטת ההשלמות המרובות קטן מרוחבו בשיטת ה-Bootstrap BCa, וכאמור, מצב כזה מתקיים עם רמת ביטחון נמוכה מהרמה המיועדת. כך, למשל, הפער הגדול ביותר בין רמת הביטחון בפועל לרמה המבוקשת התקבל בתנאי שבו המתאם הוא 0.4, N=400 ו-n=40, כאשר ההתפלגות נוטה לאחידה והרווח נבנה בשיטת ההשלמות המרובות. במקרה זה במקום 5%, התקבל שאחוז המקרים שבהם הרווח לא כיסה את הפרמטר הנכון היה 13% (לעומת 7.1% בשיטת ה-Bootstrap BCa). במקביל, הרוחב הממוצע של רווח הסמך בשיטת ההשלמות המרובות היה במקרה זה הרבה יותר צר מאשר רוחבו בשיטת ה-Bootstrap BCa (1.49 לעומת 1.59).

מסקנתנו על בסיס הסימולציות האלה היא שהשיטה של ההשלמות המרובות פחות טובה מהשיטה של Bootstrap BCa.

רשימת המקורות

כהן, א., דובא, א., אומנסקי, ט. וקנת-כהן, ת. (2013). **יישום גישות שונות לחישוב רווח סמך ולאמידת מתאם לנתונים עם קיצוץ תחום.** דו"ח מחקר שהוגש לקרן מאל"ו.

Chan, W., & Chan, D. W. L. (2004). Bootstrap standard error and confidence intervals for the correlation corrected for range restriction: A simulation study. *Psychological Methods*, *9*, 369–385.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*(39), 1-38.

DiCiccio,T.J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*,*11(3),* 189-228.

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

Levin, J. (1972). An occurrence of an increase in correlation by restriction of range. *Psychometrika, 37*, 93-97.

Little, J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.) Hoboken, NJ: Wiley.

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

Mendoza, J. L. (1993). Fisher transformations for correlations corrected for selection and missing data. *Psychometrica, 58*, 601-615.

Thorndike, R. L. (1949). *Personnel selection.* New York, NY: Wiley.

Wiberg, M., & Sundström, A. (2009). A comparison of two approaches to correction of restriction of range in correlation analysis. *Practical Assessment, Research & Evaluation*, *14(5)*, 1-9.