מירכוז משתנים בחישוב טבלאות ANOVA

איך למה חשוב למרכז משתנים בחישוב טבלאות ANOVA

# מהן טבלאות ANOVA

טבלאות ANOVA מפיקות מבחני מובהקות עבור כל משתנה (או משתנה אינטרקציה) במודל הסטטיסטי שבנינו[[1]](#footnote-1). לרוב טבלאות ANOVA מופקות עבור מודלים בהם יש לפחות משתנה קטגוריאלי אחד (כאשר יש בעיקר משתנים קטגוריאליים, ורק משתנה רציף אחד או שניים, הטבלה מכונה טבלת ANCOVA), אבל ניתן לחשב אותם גם עבור מודל בו כל הגורמים הם משתנים רציפים[[2]](#footnote-2) (או כל מקבילה דומה לכל המודלים הללו ב־HLM/LMM / רגרסיה לוגיסטית).

# חובה למרכז משתנים

**עובדה:** מרכוז של משתנים מנבאים רציפים גורם לשינויים בתוצאות והמסקנות מטבלאות ANOVA/ANCOVA[[3]](#footnote-3). אם כן, איזו תוצאה היא הנכונה? האם צריך למרכז או לא משתנים?

**טענה עיקרית:** כאשר המודל שלנו כולל אינטרקציות, על מנת שניתן יהיה לפרש את האפקטים / אינטרקציות מטבלת הANOVA, **חובה למרכז את כל המשתנים** טרם הכנסתם למודל[[4]](#footnote-4).

# למה למרכז משתנים?

מירכוז של משתנים במודל שבו יש גורם אינטרקציה משנה את חישוב השונות המוסברת במשתנים מסדר נמוך יותר (למשל במודל בו יש אינטרקציה זוגית, מירכוז ישפיע על האפקטים העיקריים; במודל בו יש אינטרקציה משולשת, מירכוז ישפיע על האינטרקציות הזוגיות ועל האפקטים העיקריים, וכן הלאה; [Dalal & Zickar, 2011](https://doi.org/10.1177%2F1094428111430540)). למעשה, רק לאחר מירכוז של המשתנים, מבחני המובהקות של האפקטים מסדר נמוך יותר באמת מייצגים את מה שאנחנו בכל מקרה חושבים שהם מייצגים – את מובהקות השונות המוסברת על ידי המשתנה (/ אינטרקציה) מעבר להבדלים באפקט של המשתנה הממותנים על ידי גורם אחר כלשהו (מה שנקרא "אפקט עיקרי"; [AFNI, website](https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/STATISTICS/center.html#centering-with-one-group-of-subjects)).

# איך למרכז משתנים?

כאשר אנחנו מחשבים טבלת ANOVA, השונות של כל משתנה מחושבת סביב ה0. מרכוז הוא תהליך חישובי שבו הופכים את ערך ה0 במשתנה הב״ת לבעל משמעות כלשהי (ממש כמו בכל מרכוז שנעשה עבור רגרסיה רגילה).

## איך למרכז משתנה רציף

די פשוט, מחסירים מכל תצפית את הממוצע של כלל התצפיות:

לאחר טרנספורמציה זו, הממוצע של המשתנה יהיה 0, והמשתנה יחשב ל"ממורכז".

יש לציין כי לא חייבים שהערך שמפחיתים יהיה ממוצע המדגם המקרי שלכם, אבל חשוב שהאפס במשתנה יהיה ערך שסביר שייצג את האוכלוסיה עליה מסיקים ([AFNI, website](https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/STATISTICS/center.html#centering-with-one-group-of-subjects)). למשל, אם המשתנה הרציף הוא IQ, לפני המרכוז, 0 ייצג ערך שלא קיים בכלל בסקאלה. ניתן אם כן למרכז לפי ממוצע המדגם, אך אפשר גם למרכז לפי 100 – שהוא הממוצע המוגדר של מבחני IQ. דוגמה נוספת: אם המשתנה הרציף שלנו הוא גיל, לפני המרכוז, 0 הוא ערך אפשרי – ייצג תינוק בן יומו – אך הוא כנראה לא מייצג האוכלוסיה עליה אנחנו רוצים להסיק.

## איך למרכז משתנה קטגוריאלי

לכאורה משימה בלתי אפשרית – למשתנים קטגוריאליים אין ערכים מספריים, אז אי אפשר להפחית מהם ערך כלשהו. זה נכון, אבל הרעיון של מרכוז הוא שכאשר אנחנו ממדלים את הערך 0 עבור אותו משתנה, המודל יספק לנו תוצאות "מעבר" לכל הרמות של אותו משתנה. לשם כך, משתני דמה שמקודדים ביחס לקבוצת בסיס כלשהי לא יעשו את העבודה, כי כאשר כל משתני הדמה מקבלים את הערך 0, המודל "השולי" (marginal) מתייחס לקבוצת הבסיס. הפתרון הוא לקודד משתני דמה במה שנקרא [effects coding](https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqwhat-is-effect-coding/) שכן נותנים את התוצאה הרצויה – כאשר נזין לקידוד הזה 0, המודל יספק לנו תוצאות "מעבר" לכל הרמות של אותו משתנה ([Aiken, et al., 1991, pp. 127](https://books.google.co.il/books?hl=iw&lr=&id=LcWLUyXcmnkC); [Singmann & Kellen, 2017, pp. 24](http://singmann.org/download/publications/singmann_kellen-introduction-mixed-models.pdf)).

יתרון נוסף של מירכוז זה של משתנים קטגוראליים הוא שחישוב טבלאות הANOVA מתמודד בצורה טובה יותר עם בעיות של חוסר איזון בין התאים ([Aiken, et al., 1991, pp. 127](https://books.google.co.il/books?hl=iw&lr=&id=LcWLUyXcmnkC); [Singmann & Kellen, 2017, pp. 24](http://singmann.org/download/publications/singmann_kellen-introduction-mixed-models.pdf)).

איך עושים את זה (ב־R)? מריצים בתחילת הסקריפט את השורה הבאה:

options(contrasts=c('contr.sum', 'contr.poly')) [[5]](#footnote-5)

|  |
| --- |
| אם אתם מבצעים ניתוחי שונות בעזרת חבילת afex, [אין צורך לעשות את זה](https://github.com/singmann/afex/issues/63), כי afex עושה את זה עבור המודל המחושב מאחורי הקלעים – אבל זה לא יזיק לאמץ את זה כהרגל. אבל… afex [לא ממרכז משתנים רציפים](https://github.com/singmann/afex/issues/59) בניתוחי ANCOVA (וכיוון שJASP עושה ANOVA/ANCOVA עם afex, זה נכון גם עבור JASP).  כאשר עושים ניתוח שונות ב־Statistica וב־SPSS, המשתנים הקטגוריאליים עוברים אוטומטית את ה"מרכוז" הזה עבור משתנים קטגוריאליים. אבל… בדיקה העלתה ש־Statistica לא ממרכז משתנים רציפים בANCOVA. (הדבר טרם נבדק בSPSS.) |

# הרחבה

**הרחבה:** כשאנחנו מחשבים טבלאות ANOVA, השונות של כל משתנה (אם לדייק, הSS של האפקט) מחושב כאשר שאר המשתנים מוחזקים קבועים ב0 (שוב, כמו שיפועים במודל רגרסיה). נניח את המודל הבא: מודל מלא (אפקטים עיקריים + אינטרקציה) עם משתנים קטגוריאליים A ו־B, כאשר משתנה B מקודד למשתני דמה המקודדים *ביחס לקבוצת בסיס* אחת – b1. כאשר אנחנו מחשבים את SSA, הדבר נעשה כאשר כל משתני הדמה של B מוחזקים ב0. אבל… במצב כזה, SSA יחושב לא סביב הממוצע מעבר לB, אלא סביב הממוצע של b1! ("קל להוכיח" שלדבר יש משמעות רק כאשר מידלנו את האינטרקציה בין A ו־B; "קל להוכיח" גם שהדבר נכון גם עבור משתנים רציפים).

**הרחבה 2:** "אבל," אני שומע אתכם זועקים, "כשלמדנו ניתוח שונות לא למדנו שצריך למרכז! או לקודד משתני דמה, או שום דבר כזה!" ואתם כמובן צודקים – לא למדתם את זה בצורה מפורשת, אבל אם תסתכלו על המשוואות של ניתוח שונות, אפשר לראות שהSS של כל גורם מחושב מעבר לממוצע של שאר הגורמים. למשל, בחישוב SSAB, אנחנו מחסירים קודם כל את התרומה של A ואת התרומה של B (לפי הממוצעים השוליים של A ו־B) – תהליך זה שווה ערך למרכוז!:

… גם בחישוב של SS של אפקט עיקרי אנחנו עושים זאת מעבר לממוצעים המותנים על ידי B, ולמעשה היה ניתן לכתוב מחדש את נוסחת SSA כך שתייצג באופן פורמלי את המרכוז של B (את ההפחתה של התרומה של B ושל התרומה של A בתנאי B – כלומר אחרי מרכוז של B), אבל זה מסובך מאד ומיותר עוד יותר:

פועל יוצא הוא שהמרכוז של משתנה מסויים משפיע רק על הSS של האפקטים העיקריים של משתנים שבאים איתו באינטרקציה, ולא על הSS של משתנה עצמו.

1. לא מדובר כאן על טבלאות ANOVA להשוואה בין מודלים, אלא על מה שקוראים לו בדרך כלל "ניתוח שונות". [↑](#footnote-ref-1)
2. אם כי הדבר יהיה מיותר טבלת ANOVA עבור מודל עם משתנים רציפים בלבד יהיה שווה בדיוק לבחינת מובהקות השיפועים של כל משתנה. [↑](#footnote-ref-2)
3. אגב, מרכוז של המשתנה התלוי לא משפיע על התוצאות של טבלאות ANOVA/ANCOVA (כלומר, הוא כן, אבל רק על הintercept, ונראה שלאף אחד לא אכפת ממנו). [↑](#footnote-ref-3)
4. על אף שמומלץ למרכז משתנים עבור מודלים של מיתון ברגרסיה, בחישוב טבלאות ANOVA מדובר בחובה של ממש ([Dalal & Zickar, 2011](https://doi.org/10.1177%2F1094428111430540)). [↑](#footnote-ref-4)
5. יש לציין כי זה יהפוך את מקדמי הרגרסיה של משתני הדמה לקשים לפרשנות, אבל בטבלאות ANOVA אנחנו לא עוסקים כאן במידול שיפועים של משתני דמה, אלא מעניין אותו האפקט העיקרי למשתנה עצמו. [↑](#footnote-ref-5)