

## Wilcoxon Rank-Sum Test

התרגיל עוסק בקובץ הנתונים "weightliftingScores.csv". בקובץ זה מופיעות תוצאות הספורטאים של נבחרת ארה"ב ושל נבחרת רוסיה בתחרות הרמת משקולות (בק"ג). עליכם לברר האם ניתן להסיק על סמך תוצאות אלו כי הביצועים של נבחרת רוסיה טובים יותר מהביצועים של נבחרת ארה"ב.

לצורך כך:

- קראו את קובץ הנתונים לטבלת מידע (dataframe).
- הציגו את התוצאות ב-boxplot, ודאו שאין ביניהם ערכים חריגים.
- ערמו את הנתונים לעמודה אחת תחת הכותרת "values" ובעמודה נוספת הציגו את הקטגוריה "ind".
- הסירו שורות בעלות ערכים חסרים.
- צרו עמודה ובה הערכים מוסטים להתפלגות משותפת סביב 0, על ידי החסרת הממוצע הקבוצתי מכל ערך נתון.
- בדקו האם ההתפלגות המשותפת היא נורמלית? (איזה מבחן משמש בצורה מהימנה לבדיקת נורמליות? מהו הסיכוי לקבלת התפלגות נורמלית עפ"י מבחן זה? האם ניתן לשמר טענת נורמליות ברמת בטחון של 95%?)
- באם ההתפלגות איננה נורמלית, בנו מבחן Wilcoxon להשוואת תוחלות באופן הבא:
  - נסחו את השערת האפס למבחן
  - חשבו את הדרוגים של ספורטאים (1 יהיה המשקל הגבוה ביותר, גודל המדגם, N, יהיה המשקל הנמוך ביותר)
  - חשבו את סכום הדרוגים של כל קבוצה (ניתן להשתמש בפונקציה "aggregate").
  - חשבו את גודל המדגם של כל קבוצה
  - חשבו את התוחלת הצפויה של נבחרת רוסיה
  - חשבו את השונות הצפויה באוכלוסיות (זכרו שבהנחת המודל השונות אחידה בשתי הקבוצות)
  - חשבו את האומדן הסטטיסטי, Z, לתוחלת הדרוג הממוצע של נבחרת רוסיה.
  - חשבו את הסיכוי שהדרוג הממוצע של נבחרת רוסיה קטן או שווה לתוחלת שחישבתם.

בצעו מבחן Wilcoxon בעזרת הפונקציה הרלוונטית ב-R והשוו את התוצאות שקיבלתם עם Z ו- $P_V$  המחושבים.

בצעו מבחן t-test, לבדיקת שוויון תוחלות בין הקבוצות. (בין אם מותר או אסור בהתאם לדרישות הסף של המבחן).

בחנו את ההבדלים באומדנים הסטטיסטיים של מבחן t ושל המבחן הלא פרמטרי.  
איזה מבחן יותר מתירני? איזה מבחן יותר שמרני?

תזכורת לנוסחאות רלוונטיות (לקוח מהרצאה 10):

$$\mu(T_1) = \frac{n_1}{2} (N + 1)$$

$$\sigma(T_1) = \sigma(T_2) = \sqrt{\frac{n_1 \cdot n_2 (N + 1)}{12}}$$

$$Z = \frac{T - \mu(T)}{\sigma(T)}$$