**פרויקט 1 - הגורם האנושי באיסוף נתונים**

**ספורט ובריאות גופנית**

מאת: גיל קפלן, יעקב פרידמן והלל טריפ

תאריך הגשה: 05.01.2025

**תוכן עניינים**

[חלק 1 - מבוא 3](#_Toc186988744)

[חלק 2 - שיטה 3](#_Toc186988745)

[1. תיאור קובץ הנתונים 3](#_Toc186988746)

[2. מידע עבור המשתנים: BMIומשך הביצוע של פעילות גופנית בשבוע 3](#_Toc186988747)

[3. תוקף ומהימנות 5](#_Toc186988748)

[חלק 3 - תוצאות 6](#_Toc186988749)

[מבחן סטטיסטי וויזואליזיות 6](#_Toc186988750)

[חלק 4 – מסקנות 7](#_Toc186988751)

[משמעות התוצאות 7](#_Toc186988752)

[תוקף ומהימנות התוצאה 7](#_Toc186988753)

[ניתוחי המשך 7](#_Toc186988754)

[שיפור המדידה 7](#_Toc186988755)

[חלק 5 - ביבליוגרפיה 8](#_Toc186988756)

[חלק 6 – קובץ הנתונים 9](#_Toc186988757)

# חלק 1 - מבוא

1. **השערה אופרציונלית:** יש קשר שלילי בין משך הביצוע של פעילות גופנית בשבוע ל-BMI של אדם.
2. **השערה תיאורית:** יש קשר שלילי בין ספורטיביות לבין חוסר בריאות הגופנית.

לבריאות גופנית השלכות רבות על איכות החיים. השערתנו חשובה מכיוון שמחקרי עבר בחנו את הקשר בין ספורט למצב הבריאותי, ובפרט את הקשר בין ספורט להשמנת יתר (Cleven et al., 2020), ומצאו כי פעילות גופנית קשורה לסיכון נמוך יותר להשמנת יתר ומחלות אחרות. בנוסף, ההשערה מעניינת מכיוון שלהשמנת יתר השפעות על הבריאות הגופנית (Kopelman, 2000), ונמצא קשר בינה לבין סיכון למחלות לבביות ואף תמותה (The GBD 2015 Obesity Collaborators, 2017).

מכיוון שערכי BMI גבוהים מאפיינים עודף משקל (overweight) והשמנת יתר (obesity) בחרנו בו בתור משתנה אופרציונלי לחוסר בריאות גופנית. בגלל הקשר בין ספורט BMI ומצב בריאותי, בפרויקט אנו מעוניינים לבדוק באופן כללי יותר את הקשר בין פעילות גופנית לבין BMI.

# חלק 2 - שיטה

1. תיאור קובץ הנתונים

עבדנו עם קובץ הנתונים של The Behavioral Risk Factor Surveillance System (BRFSS) משנת 2015. ה-BRFSS הוא סקר טלפוני שנערך מדי שנה בארצות הברית, על ידי ה-CDC (Centers for (Disease Control and Prevention. הוא מכיל אוסף מידע על התנהגויות הקשורות לסיכון בריאותי והרגלים הקשורים לרפואה מונעת. בנוסף נאסף מידע על גישה לשירותי רפואה, מחלות כרוניות ופציעות. העונים על הסקר הם מעל גיל 18 (Behavioral Risk Factor Surveillance System (BRFSS) - Healthy People 2030 | Odphp.Health.Gov). קובץ הנתונים מכיל 441,456 רשומות. מקור הקובץ נמצא בנספחים.

## 2. מידע עבור המשתנים: BMIומשך הביצוע של פעילות גופנית בשבוע

BMI

מדובר במשתנה מנה. ערך ה-BMI המופיע בקובץ הנתונים חושב על ידי הסוקרים מתוך המשקל והגובה שסיפקנו הנשאלים. טווח הערכים של ה-BMI הוא בין 12.02 ל-99.95. ל-BMI יש מספר ערכים קיצוניים בריאותית הנחשבים מסוכנים, אך אין ערכים חריגים באופן חד משמעי.

יש %8.24 אחוז ערכים חסרים, שנובעים מערך חסר של משקל או גובה. רבים מהערכים החסרים נובעים מאנשים שסירבנו לענות על שאלת המשקל (מידע זה מתועד בקובץ הנתונים). לדעתנו דפוס הערכים החסרים הוא Missing not at Random (MNAR), מפני שלדעתנו ככול הנראה לאותם אנשים הייתה סיבה אישית לסרב לשאלה (אולי לאנשים עם משקל גבוה חשו חוסר נוחות לענות על השאלה).

**ניתוח סטטיסטי**

Mean: 28.04, Median: 26.95, Standard Deviation: 6.65, Min: 12.02, Max: 99.95,

Q1: 23.73, Q3: 30.9, IQR 7.17

A graph of distribution of bmi

Description automatically generated

בהיסטוגרמה ניתן לראות את הזנב הימני של ערכי ה-BMI הקיצוניים. רוב הערכים מרוכזים בתחום צר.

משך הביצוע של פעילות גופנית בשבוע

מדובר על משתנה מנה. ערך זה חושב על ידינו על ידי הכפלה של הנתונים על תדירות ביצוע פעילות גופנית, ומשך הביצוע של הפעילות הגופנית בכל פעם שבוצעה (תיאור מפורט מופיע בנספחים).

יש למשתנה 10.95% ערכים חסרים. לדעתנו כמו במקרה של ה-BMI, דפוס החוסרים הוא ככל הנראה Missing not at Random מכיוון שאולי חלק מהאנשים לא רצו לענות על השאלה מסיבות אישיות. בגלל כמות הרשומות הגדולה ואופי הנתונים החסרים, לא נתייחס אליהם בביצוע המבחנים.

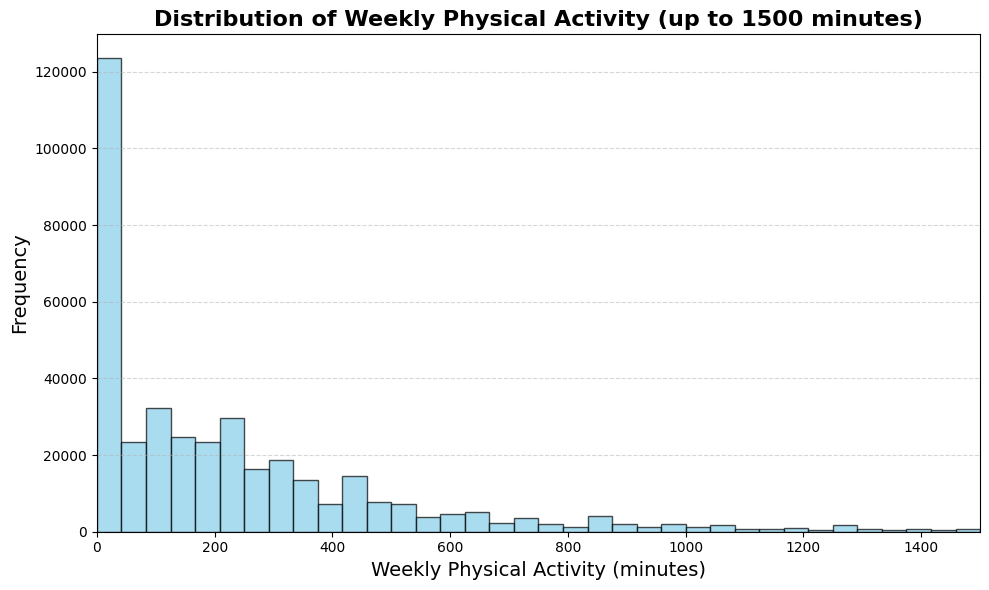
ערכים חריגים במשך ביצוע הפעילות הגופנית בשבוע נובעים מערכים חרגים בתדירות הפעילות הגופנית או במשך הפעילות. הנחנו שפעילות גופנית של יותר מ-12 שעות ביום (5040 דקות) היא ערך חריג, ותחת הנחה זו יש 189 רשומות עם ערכים חריגים. בנוסף יש 32 ערכים החורגים מעל כמות הדקות הכוללת בשבוע, ולכן אינם אפשריים.

**ניתוח סטטיסטי**

Mean: 275.88, Median: 150.00, Standard Deviation: 469.22, Min: 0.0, Max: 41670.00,

Q1 = 0.0, Q3 = 347.66, IQR = 347.66

הממוצע וסטיית התקן מושפעים מהערכים החריגים, ובלעדיהם הממוצע קטן ל-272.09 וסטיית התקן קטנה ל- 423.62.



בהיסטוגרמה ניתן לראות את התפלגות זמני הפעילות הגופנית. מכיוון שלערכים יש זנב ימני ארוך מאוד הצגנו רק ערכים קטנים מ-1500 דקות (אלו 97.8% מהערכים). 0 הוא הערך השכיח, שמהווה כ-27%.

## 3. תוקף ומהימנות

**תוקף ה-BMI**

תוקף פנים: לדעתנו ל- BMI יש תוקף גבוה עבור חוסר בריאות גופנית, מכיוון שמחקרי עבר מצאו קשר בין BMI לבין מגוון מחלות (Kopelman, 2000). למרות זאת, בעיה שפוגעת בתוקף של BMI היא שערכי BMI נמוכים (תת-משקל) גם אינם טובים לבריאות. בנוסף, BMI לא תמיד משקף מצב בריאותי חיובי או שלילי עבור בן אדם בודד. למרות שה-BMI נובע מערכים מדווחים לגובה ומשקל, לדעתנו רוב האנשים ידווחו את הערכים הנכונים, ולכן ערכי ה-BMI המדווחים יהיו קרובים לערכי ה-BMI הנכונים.

תוקף מתכנס: מכיוון שבקובץ הנתונים שלנו יש משתנים שונים שיכולים להוות משתנים אופרציונאליים לחוסר בריאות גופנית, אנו יכולים לבדוק את התוקף המתכנס של BMI. בחרנו לבדוק את הקשר בין BMI לבין בעיות לבביות, שגם הן משתנה אופרציונאלי לחוסר בריאות גופנית . שני משתנים רלוונטיים שמכיל קובץ הנתונים הם, האם הנשאל חווה התקף לב, והאם הוא סובל מאנגינה או ממחלה לבבית. הגדרנו שנשאל סובל מבעיות לבביות אם לפחות אחד התנאים מתקיים לגביו (שמות העמודות מופיעים בנספחים). ביצענו מבחן רגרסיה לוגיסטית כדי לבדוק את הקשר בין BMI לבין בעיות לבביות. קיבלנו ו-. (, וסטטיסטי המבחן *). קיבלנו קשר מובהק סטטיסטית בין* BMI *לבין בעיות לבביות, למרות שערך ה- נמוך. הדבר מחזק במידה מסוימת את התוקף המתכנס של ה-* BMI *בתור משנה אופרציונלי לחוסר בריאות גופנית.*

**מהימנות ה-BMI**

מהימנות test-retest: אנשים רבים יודעים את הגובה והמשקל שלהם בדיוק יחסית גבוה, והם קלים למדידה, לכן לדעתנו למשתנה זה מהימנות גבוהה. בנוסף, עבור אנשים בוגרים (הנבדקים מעל גיל 18), הגובה אינו משתנה באופן ניכר, וככל הנראה השינויים במשקל אינם קורים בטווחים זמן קצרים.

**תוקף הזמן המוקדש לביצוע פעילות גופנית**

תוקף פנים: לדעתנו למדד זה תוקף פנים בינוני. מדד זה לא מתחשב במאמץ הפעילות הגופנית, ששונה מאוד בין פעילות, ושונה מאוד בין אנשים שונים המבצעים את אותה פעילות. בנוסף, מכיוון שהוא נובע מדיווח עצמי על תדירות ומשך פעילות גופנית, התוקף שלו נפגע. אין משתנים אופרציונאליים אחרים שונים מספיק עבור ספורטיביות, ולא ניתן לבדוק סוגי תוקף אחרים.

**מהימנות הזמן המוקדש לביצוע פעילות גופנית**

מהימנות test-retest: בשאלון עליו מתבסס קובץ הנתונים, השאלות בנושא הפעילות גופנית מתייחסות לחודש האחרון בלבד. לכן עבור אנשים שאינם מבצעים פעילות גופנית באופן עקבי (למשל אימונים בזמנים מוגדרים בשבוע), התשובות לשאלות אלו יכולות להשתנות. בנוסף, ייתכן שלא כל הנשאלים, בעיקר אלו שלא מצבעים פעילות קבועה, יודעים בדיוק את כמות הפעילות הגופנית שהם ביצעו, או את הזמן שהם מבצעים אותה. לכן למשתנה זה מהימנות נמוכה.

לדעתנו המשתנים מספיקים, למרות שתוקף והמהימנות של משך הפעילות הגופנית נמוכים. את תוקף הניבוי נבדוק לאחר המבחן הסטטיסטי.

# חלק 3 - תוצאות

## מבחן סטטיסטי וויזואליזיות

כדי לבדוק קשר בין הזמן המוקדש לביצוע פעילות גופנית בשבוע לבין ה-BMI, ביצענו רגרסיה לינארית (ללא התייחסות לערכים חסרים). כפי ששיערנו, נמצא קשר שלילי מובהק בין משך הביצוע של פעילות גופנית בשבוע לבין ה-BMI. המודל מסביר 0.7% מהשונות (*). כאשר מבצעים את מבחן הרגרסיה ללא הערכים החריגים, התוצאה נשארת מובהקת ואחוז השונות המוסברת גדל מעט. (). בנוסף, השיפוע נהיה שלילי יותר.*

|  |  |
| --- | --- |
|  |  |

התרשים הימני מציג הנתונים ואת קווי הרגרסיה עבור כלל הנתונים. התרשים השמאלי הוא הגדלה של התרשים הימני בתחום בו משך פעילות גופנית קטן מ-1500 דקות (97.8% מהערכים). קווי הרגרסיה מוצגים באדום – הקו האדום המלא הוא עבור רגרסיה שבוצעה על כלל הנתונים, והקו האדום המקווקו הוא עבור הרגרסיה שבוצעה ללא ערכים חריגים. בגלל השיפוע הקטן, בתחום עד 1500 דקות כמעט אין הבדל בין הקוים.

# חלק 4 – מסקנות

## משמעות התוצאות

המבחן שביצענו מאשש את ההשערה שלנו, כלומר ניתן להסיק שיש קשר שלילי מובהק בין משך הביצוע של פעילות גופנית בשבוע לבין ה-BMI של אדם. במונחי ההשערה התאורטית, ניתן להסיק שיש קשר שלילי מובהק בין ספורטיביות לחוסר בריאות גופנית. מסקנה זו מתיישבת עם מחקרים קודמים (Cleven et al., 2020). למרות שהקשר מובהק אחוז השונות המוסברת נמוך מאוד.

כעת נוכל להסיק את תוקף הניבוי של משך הביצוע של פעילות גופנית ביחס ל-BMI. מכיוון שאחוז השונות המוסברת נמוך, ככל הנראה תוקף הניבוי חלש.

למרות הקשר, לא ניתן לקבוע קשר סיבתי, מאחר וייתכנו הסברים חלופיים. לדוגמה, ייתכן שלאנשים בעלי BMI נמוך קל יותר לבצע פעילות גופנית ולכן הם עושים יותר ממנה.

## תוקף ומהימנות התוצאה

תוקף ומהינות התוצאה שלנו תלויים בתוקף ובמהימנות של המשתנים האופרציונאליים. למרות שקיבלנו p-value נמוך מאוד, תוקף ומהימנות המשתנה של משך זמן ביצוע הפעילות הגופנית בשבוע נמוך (כפי שהוסבר בחלק השיטה). לכן למרות שהתוצאה שלנו מתיישבת עם מחקרי עבר (Erlichman et al., 2002), אנו לא חושבים שיש לה תוקף ומהימנות גבוהים.

## ניתוחי המשך

קיבלנו אחוז שונות מוסברת מאוד נמוך. לכן ייתכן שאם נוסיף למודל עוד משתנים מסבירים, נוכל להבין בצורה טובה את הקשר בין המשתנים השונים לבין ערך ה-BMI. לדוגמה, בקובץ הנתונים יש מידע לגבי צריכת פירות וירקות, שלפי מחקרים יש קשר בינה לבין השמנת יתר (המוגדרת על ידי BMI גבוה) (Arnotti & Bamber, 2020).

## שיפור המדידה

למרות שלדעתנו ערך ה-BMI הנובע מדיווח עצמי הוא בעל מהימנות גבוהה, אם הוא ייבדק במהלך המחקר (על ידי שקילה של הנדבקים ומדידת הגובה שלהם) הוא יהיה בעל מהימנות גבוהה אף יותר.

לדעתנו ניתן להשיג מדדים אופרציונאליים בעלי תוקף ומהימנות גבוהים עבור ספורטיביות אם ניתן למדוד באופן מדויק יותר את הפעילות הגופנית של הנדבקים למשך טווח זמן ארוך יותר. לדוגמה, בעזרת שעון חכם ניתן למדוד באופן מדויק את זמן הביצוע של פעילות גופנית, ואת הדופק של הנבדקים בעת ביצוע הפעילות.

# חלק 5 – ביבליוגרפיה

Arnotti, K., & Bamber, M. (2020). Fruit and Vegetable Consumption in Overweight or Obese Individuals: A Meta-Analysis. In *Western Journal of Nursing Research* (Vol. 42, Issue 4, pp. 306–314). SAGE Publications Inc. https://doi.org/10.1177/0193945919858699

*Behavioral Risk Factor Surveillance System (BRFSS) - Healthy People 2030 | odphp.health.gov*. (n.d.). Retrieved January 3, 2025, from https://odphp.health.gov/healthypeople/objectives-and-data/data-sources-and-methods/data-sources/behavioral-risk-factor-surveillance-system-brfss

Cleven, L., Krell-Roesch, J., Nigg, C. R., & Woll, A. (2020). The association between physical activity with incident obesity, coronary heart disease, diabetes and hypertension in adults: A systematic review of longitudinal studies published after 2012. *BMC Public Health*, *20*(1). https://doi.org/10.1186/s12889-020-08715-4

Erlichman, J., Kerbey, A. L., & James, W. P. T. (2002). Physical activity and its impact on health outcomes. Paper 2: Prevention of unhealthy weight gain and obesity by physical activity: An analysis of the evidence. In *Obesity Reviews* (Vol. 3, Issue 4, pp. 273–287). https://doi.org/10.1046/j.1467-789X.2002.00078.x

Kopelman P. G. (2000). Obesity as a medical problem. Nature, 404(6778), 635–643. https://doi.org/10.1038/35007508

The GBD 2015 Obesity Collaborators. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *New England Journal of Medicine*, *377*(1), 13–27. https://doi.org/10.1056/NEJMoa1614362

# חלק 6 – קובץ הנתונים

את קובץ הנתונים לקחנו מתוך ה-kaggle:

<https://www.kaggle.com/datasets/cdc/behavioral-risk-factor-surveillance-system>

קישור לקבוץ CSV:

<https://www.kaggle.com/datasets/cdc/behavioral-risk-factor-surveillance-system/data?select=2015.csv>

הסברים על קובץ הנתונים מופיעים כאן:

<https://www.cdc.gov/brfss/annual_data/annual_2015.html>

קובץ ה-<https://www.cdc.gov/brfss/annual_data/2015/pdf/codebook15_llcp.pdf> מתאר את המשתנים, והקובץ <https://www.cdc.gov/brfss/annual_data/2015/pdf/2015_Calculated_Variables_Version4_08_10_17-508c.pdf> מתאר חישובים שנעשו על ידי עורכי הסקר.

# נספחים

## טרנספורמציות

את ערך ה-BMI לקחנו מתוך העמודה \_BMI5, שבה הוא מופיע ערך ה-BMI כפול 100.

הערך EXERANY2 מכיל עמודה בינארית של האם הנשאל מבצע פעילות גופנית או לא. עבור נבדקים שענו לא, השמנו משך פעילות גופנית שבועית של 0.

הסקר מכיל שאלות על שתי הפעילויות הגופניות הנפוצות ביותר. עמודות PAFREQ1\_ ו- PAFREQ2\_ מכילות את כמות הפעמים בשבוע בהן מבוצעות הפעילויות (כפול 1000). עמודות PADUR1\_ו- PADUR2\_ מכילות את כמות הזמן שכל פעילות לוקחת בכל פעם שהיא מבוצעת. הכפלו בין הערכים לכל פעילות וסכמנו אותם.

עמודה CVDINFR4 מכילה מידע בינארי של האם הנבדק חווה התקף לב, ועמודה CVDCRHD4 מכילה מידע בינארי של האם הנבדק סובל מאנגינה או מחלות לבביות.