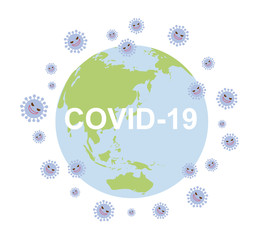
****

**פרויקט גמר**

**השפעות הקורונה בתחום הסוכרת** **–**

**מדד הגלוקוז**



ליאב ריאר

מנחה: ד"ר מיכל שאולי אהרונוב

**המחלקה להנדסת תעשיה וניהול**

**מכון טל**

**המרכז האקדמי לב**

**שנת תשפ"ג**

**תודות**

ברצוננו להודות תחילה למנחת הפרויקט שלנו, ד"ר מיכל שאולי – אהרונוב, על הליווי לאורך כל הדרך, על הפגישות בזום ועל תרומתה מניסיונה האישי והמקצועי.

תודה נוספת לד"ר שפריר, רופא משפחה בקופ"ח מאוחדת, על מאגר הנתונים.

**תקציר מנהלים**

כיום, ההשפעות של הקורונה נחקרות ברחבי העולם וטרם ידועות כל ההשלכות שהיא הביאה עימה.

"שאולי-אהרונוב, שפריר ואחרים" (2021) הראו שהסיכון לסיבוכי קורונה, גבוה באופן מובהק בקרב אנשים עם סוכר גבוה.

אולם ההשפעה ההפוכה - איך הקורונה השפיעה על ערכי הסוכר, טרם נחקרה.

במשך הקורונה נאספו נתונים של כ- מיליון מטופלי קופ"ח מאוחדת, חלקם נדבקו בקורונה וחלקם לא. המטרה הייתה להשתמש במאגר עצום זה כדי לבחון, עם תום המגפה, את השפעותיה על רמות הסוכר בגוף.

שאלות המחקר הן - האם הקורונה השפיעה על ערכי הסוכר בקרב אוכלוסיות שונות: נשים/גברים, מעשן/לא מעשן, גיל, מגזרים: אורתודוקסי, ערבי וכללי?

אלו הם המשתנים שלנו.

3 אוכלוסיות המחקר שלנו הן: מטופלים שחלו בקורונה קלה, מטופלים שחלו בקורונה חמורה ומטופלים שלא חלו בקורונה.

לאורך הפרויקט התמקדנו בשני מדדי סוכר עיקריים: HBA1C ו- (FBG) GLUCOSE.

הפרויקט התחלק כך שכל אחת התמקדה במחקר על מדד סוכר מסוים (החל מחציו השני של הפרויקט).

תחילה, ביצענו מבחני T למדגמים תלויים על 2 מאוכלוסיות המחקר: מטופלים שחלו בקורונה חמורה ומטופלים שחלו בקורונה קלה, עבור המשתנים השונים.

לאחר מכן ביצענו מבחן ANOVA על 3 אוכלוסיות הקורונה, עבור כל אחד מ-4 המשתנים שלנו.

מכאן, המשכנו לביצוע רגרסיה ליניארית, לבחינת המשתנים המסבירים את המשתנה המוסבר.

להלן המשתנה המוסבר: הפרש מדד הסוכר (ערך מדד הסוכר בתחילת תקופת המעקב פחות ערך מדד הסוכר בסיום תקופת המעקב), הנ"ל מתאר את השינוי במדד הסוכר.

בהמשך ביצענו רגרסיה לוגיסטית עם אותם המשתנים המסבירים. והתווסף משתנה מסביר, שהינו משתנה בינארי: האם המטופל חלה בסוכרת בתחילת תקופת המעקב? מרגרסיה זו נחזה את הסיכוי של מטופל לחלות בסוכרת בסיום תקופת המעקב. בנוסף, ביצענו לרגרסיה זו את יחס הסיכויים.

שימוש ברגרסיה ליניארית ולוגיסטית מאפשר לחקור את המידה בה כל משתנה משפיע על המשתנה המוסבר ולקבל ערך צפוי.

לבסוף בנינו ממשק משתמש, עבור קהילת הרופאים:

בהיבט של רגרסיה ליניארית – מביא צפי לערך מדד הסוכר בסיום תקופת המעקב.

בהיבט של רגרסיה לוגיסטית - בוחן סיכוי של מטופל לחלות בסוכרת בסיום תקופת המעקב.

הכלים שעבדנו איתם על המבחנים השונים הם: Excel ו-R studio.

סיכום ממצאים:

* ממוצע מדד ה- HBA1C בקרב המגזר האורתודוקסי שחלו בקורונה חמורה, ירד משמעותית לאחר החולי.
* ממוצע מדד ה- FBG בקרב אוכלוסיית המעשנים שחלו בקורונה חמורה, ירד משמעותית לאחר החולי.
* יש הבדל מובהק בין חולי קורונה חמורה לבין היתר בממוצע הפרשי מדדי הסוכר (AFTER-BEFORRE).
* רגרסיה ליניארית: המשתנים המסבירים המובהקים - אוכלוסיית המעשנים ומטופלים שחלו בקורונה חמורה.
* רגרסיה לוגיסטית: בהתמקדות על מדד ה- HBA1C, כל המשתנים המסבירים יצאו מובהקים ואילו בהתמקדות על מדד ה- GLUCOSE, כל המשתנים המסבירים יצאו מובהקים, מלבד המשתנה סוג הקורונה.

**Abstract**

The outcomes of Covid are yet to be fully discovered and are still under research today.

Shauly Aharonov, Shafrir et al. (2021) showed that the risk for Covid complication is higher for people with higher values of glucose in their blood. But the opposite outcome-the effect of glucose level on Covid is yet to be researched.

During Covid pandemic, data of about a million members of Meuhedet Health Services was collected. Of them, some had Covid, and some did not. The purpose of this data was to research the connection of glucose level and e Covid complications.

The research questions are-how covid affects the glucose level when focusing on these criteria: gender, smoking habits, age, and sectors.

Those are the variables we focus on.

The research populations are patients with minor Covid, severe Covid and no Covid.

The measurements we focus on are HBA1C and glucose (FBG).

The workflow was that each student focused on one measurement (starting from the second half of the project).

At the beginning we used T-test for dependent samples over 2 of the above populations: those with severe and minor Covid for the mentioned variables.

Then we performed ANOVA test over the 3 populations for each of the variables.

Following these tests, we used linear regression to examine the variables connection to the measurements.

Later we used logistic regression to the same covariates with one addition-binary variable that represents whether the patient had diabetes at the beginning of the follow up period. This regression will be used to predict the chance for a patient to have diabetes at the end of the follow up period. In addition, we performed odds ratio analysis.

Eventually we developed a user interface for the physicians:

Linear regression wise- can predict the glucose measurement at the end of the follow up period.

Logistic regression wise- can predict the odds of a patient having diabetes at the end of the follow up period.

The tools we used are Excel and R studio.

Conclusions:

•Average HBA1C measurement among the Orthodox sector with severe Covid drastically decreased.

•Average FBG measurement among the smoking population with severe Covid drastically decreased.

•There is a statistically significant difference between patients with severe Covid and other patients in terms of glucose measurements difference.

•Linear regression: the significant variables are smoking and severe Covid.

•Logistic regression: focusing on HBA1C measurement, all covariates are statistically significant. Focusing on glucose measurement, all covariates are statistically significant besides Covid type.

**תוכן עניינים**

[1. מבוא 9](#_Toc141187847)

[2. תיאור המצב הקיים 11](#_Toc141187848)

[2.1 מאמר 1 11](#_Toc141187849)

[2.2 מאמר 2 12](#_Toc141187850)

[2.3 מאמר 3 13](#_Toc141187851)

[3. תיאור העבודה 14](#_Toc141187852)

[3.1 מבחן T למדגמים תלויים 15](#_Toc141187853)

[3.2 ניתוח ANOVA+ מבחן T למדגמים בלתי תלויים 16](#_Toc141187854)

[3.3 רגרסיה ליניארית רבת משתנים 17](#_Toc141187855)

[3.4 רגרסיה לוגיסטית 18](#_Toc141187856)

[3.5 יחס הסיכויים 23](#_Toc141187857)

[3.6 ממשק משתמש 23](#_Toc141187858)

[3.7 אבני דרך 24](#_Toc141187859)

[4. תוצאות 25](#_Toc141187860)

[4.1 תוצאות מבחן T למדגמים תלויים 25](#_Toc141187861)

[4.2 תוצאות ניתוח ANOVA 26](#_Toc141187862)

[4.3 תוצאות רגרסיה ליניארית רבת משתנים 27](#_Toc141187863)

[4.4 תוצאות רגרסיה לוגיסטית 28](#_Toc141187864)

[4.5 תוצאות יחס הסיכויים 31](#_Toc141187865)

[4.6 תוצאות ממשק משתמש 32](#_Toc141187866)

[5. סיכום 34](#_Toc141187867)

[6. ביבליוגרפיה 35](#_Toc141187868)

[7. נספחים 36](#_Toc141187869)

**רשימת איורים**

[איור 1 – התפלגות לוגיסטית 18](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260972)

[איור 2 – מודל confusion matrix 20](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260973)

[איור 3 – גרף המחשה בשינוי ממוצע מדד הגלוקוז בקרב אוכלוסיית המעשנים 25](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260974)

[איור 4 – פלט מרכיבי הרגרסיה הליניארית להרכבתה 27](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260975)

[איור 5 – תוצאות סיווג confusion matrix 29](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260976)

[איור 6 – קבלת עקומת ROC 30](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260977)

[איור 7- הצגת שימוש בממשק משתמש על סמך הרגרסיה הליניארית 32](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260978)

[איור 8 - הצגת שימוש בממשק משתמש על סמך הרגרסיה הלוגיסטית 33](file:///C:\Users\97255\Desktop\מבחן%20אנובה%20ומבחני%20T\דוח%2020.07.docx#_Toc141260979)

**רשימת טבלאות**

[טבלה 1 - משתנים מובהקים שהתקבלו מניתוח ANOVA 26](#_Toc141216547)

[טבלה 2 – תוצאות מבחן T למדגמים בלתי תלויים 26](#_Toc141216548)

[טבלה 3 – הצגת המשתנים המובהקים של הרגרסיה הליניארית 28](#_Toc141216549)

[טבלה 4 – הצגת המשתנים המובהקים של הרגרסיה הלוגיסטית 29](#_Toc141216550)

[טבלה 5 – תוצאות יחס הסיכויים 31](#_Toc141216551)

# **1. מבוא**

הקורונה הינה מגפה חדשה שהגיחה לעולם ורוב המחקרים לא עוסקים בהשפעות הקורונה על הטווח הארוך.

כיום, הוכח קשר ישיר בהשפעתו של סוכר גבוה לסיכון בהחמרת הקורונה.

אולם ההשפעה ההפוכה - איך הקורונה השפיעה על ערכי הסוכר טרם נחקרה, וזאת באנו לחקור.

משמעות מחלת הסוכרת הינה הפרעה בחילוף החומרים, הגורמת לעלייה ברמת הסוכר בדם.

אי שמירה על ערכי סוכר תקינים בדם מעלה את הסיכון ללקות בסיבוכי הסוכרת המאוחרים, הכוללים: פגיעה בכלי דם במוח עד כדי אירוע מוחי, לחסימת כלי דם בלב שמביאים לתעוקת חזה ולהתקפים, פגיעה בכלי הדם הקטנים ברשתית העין עד כדי סכנת עיוורון ועוד.

חשיבותו של פרויקט זה הינו עצום, מפני שהממצאים שנקבל יהוו המשך מחקר עתידי, אשר במידה ויתגלה כי הקורונה אכן משפיעה על ערכי הסוכר, תעלה המודעות למעקב סדיר אחר אנשים אלו ואף להתוויית תוכנית למניעת הדרדרות לסוכרת, מעבר לזה כי כל נושא העוסק בבריאות האדם, הוא נושא חשוב בפני עצמו.

במהלך הפרויקט ביצענו שימוש בכלים סטטיסטיים ובכלי כריית מידע, אשר הובילו לממצאים חדשים.

הכלים שבהם השתמשנו:

* מבחן T תלוי ובלתי תלוי.
* ניתוח ANOVA.
* רגרסיה ליניארית רבת משתנים.
* רגרסיה לוגיסטית.

מדדי הסוכר שאותם בחנו:

1. GLUCOSE BLOOD FASTING (FBG) - רמת הסוכר בדם לאחר צום.
2. HBA1C - רמת הסוכר הממוצעת בדם במשך 3 החודשים האחרונים.

החל מחציו השני של הפרויקט, כל אחת התמקדה על מדד סוכר מסוים.

ההתמקדות שלי בפרויקט התבצעה על מדד הגלוקוז בלבד, לכן בדוח זה תופיע התייחסות אך ורק למדד הגלוקוז.

תכולתו העיקרית של הדו"ח:

* סקירת ספרות על המצב הקיים
* תיאור העבודה
* תוצאות
* סיכום

# **2.** **תיאור המצב הקיים**

## **2.1 מאמר 1 - Risks and burdens of incident diabetes in long covid**

המאמר עוסק בחקר של הקשר בין אנשים שחלו בקורונה וסיכויים לחלות בסוכרת (וכן שימוש בתרופות להורדת סוכר), כשנה לאחר מכן ( מעקב לחציון של 352 ימים).

השלב הארוך של קורונה מוגדר לאחר 30 יום מהחלק האקוטי של המחלה, וכולל שלל השפעות נשימתיות וכן כשל באיברים מסוימים-כולל סוכרת. על אף הסיכוי לקשר כזה, מחקר הכולל מעל ל-1000 נחקרים על הקשר אינו קיים, ולכן חשיבות המחקר, שמתמקד בניתוח נתונים נרחב- 181,280 משתתפים.  
בסיס הנתונים של המחקר נלקח מתוך מאגר פורשים מצבא ארה"ב, מה-01/03/19 עד ל30/09/21. מתוכם בודדו אלו שחלו בקורונה לפי מאגר המידע, ותאריך גילוי הקורונה הוגדר כT0. לצורך בקרה, נלקחו גם נתונים היסטוריים מתוך אותו המאגר ובוצעה השוואה לאחוז חולי הסוכרת מתוך אותם שלא חלו בקורונה. חולי סוכרת הוגדרו ככאלה על פי סטנדרט בדיקת חלבון המעיד על סוכרת (חלבון A1c) וכן לפי אותם אלו שהונפקו להם מרשמים לתרופות נגד סוכרת.  
לצורך דיוק התוצאות, נלקחו בחשבון משתנים רבים נוספים כגון גיל, גזע, מין, עבר רפואי וכו'. כמו כן, בוצע ניתוח על קבוצת הביקורת על מנת לשערך את הסיכוי לחלות בסוכרת ללא קשר לקורונה. באופן זה ניתן לבדוק מהו הסיכוי של קבוצת החולים בקורונה לחלות בסוכרת ללא קשר לעובדת היותם חולים. חלוקות נוספות כגון חומרת הקורונה (ללא אשפוז, אשפוז או טיפול נמרץ), מתי החולים חלו, האם נרשמו שימושים בתרופות נוספות שעלולות להשפיע על סוכרת במסגרת הטיפול בקורונה וכדו' בוצעו על מנת לדייק אף יותר את התוצאות.  
נוסף על כך, על מנת לוודא את טיב התכן של המחקר, בוצעה השוואה לחיפוש קשר בין קורונה לסיכוי תמותה, הידוע כבעל קשר סטטיסטי חיובי, וכן השוואה מול גורמים הידועים ככאלה ללא השפעה הכוללים בין השאר שימוש במכשירי שמיעה, אקנה, מדדי חלבון בדם ועוד.  
תוצאות המחקר מוכיחות כי הסיכון לחלות בסוכרת גבוה יותר באנשים שחלו בקורונה בעבר (סיכוי של 1.4%), ללא קשר לגורמים כמו גיל, גזע, מין וכדו' לעומת כאלו שלא חלו. כמו כן נטל המחלות הגלובלי (Disease Burden). גם הוא גבוה יותר בכ-50% בחולים בקורונה בעבר לעומת כאלה ללא עבר. סיכוי גבוה יותר לחלות התקבל באנשים מעל גיל 65, כהי עור, בעלי BMI גבוה, בעלי עבר רפואי של מחלות כגון מחלות לב או טרום-סוכרתיים. לבסוף, התוצאות הראו נכונות גם בהערכת טיב המחקר כאשר נצפה קשר חיובי בין חולי קורונה ואחוז תמותה וקשר שלילי בין שאר הגורמים שפורטו לעיל.

הסייגים לתוצאות הם שבסיס הנתונים הורכב בעיקר מזכרים לבנים, ייתכן ובקבוצת הבקרה היו חולים שלא היו מודעים לכך (לא ביצעו בדיקה לקורונה), ולבסוף הטיפול בקורונה מתקדם ומשתנה כל הזמן, מה שגם עלול להשפיע על תוצאות המחקר.

למסקנה, המחקר מוכיח קשר ברור בין חולי קורונה בעבר לבין סיכון מוגדל לחלות בסוכרת. הגורמים לקשר עדיין לא ברורים לחלוטין, ויכולים לנבוע מקשר בין חלבונים בלבלב אותם צריך וירוס הקורונה על מנת להיכנס לגוף, או מתגובה חיסונית כוללת או עצמית (אוטואימונית) שמאפיינת חולי קורונה. לבסוף, ייתכן וגורמים נוספים הנובעים מהמגפה (בידוד, מצב כלכלי וכדו') משפיעים גם הם על הסיכוי לחלות.

## **2.2 מאמר 2 - Both high and low pre-infection glucose levels associated**

## **with increased risk for severe COVID-19**

המחקר עוסק בחיפוש הקשר בין ערכי הגלוקוז של חולה, לצד גורמים נוספים כמו משקל, גיל, גזע וכדו' והסיכוי לחלות בקורונה חמורה. קורונה חמורה מוגדרת כזו הגורמת לאשפוז של 10 ימים לפחות, הגעה לטיפול נמרץ או מוות.

לצד מחקרים רבים שהוכיחו את הקשר בין חומרת מחלת הקורונה לסוכרת, המחקר הנוכחי מתמקד בשלב לפני החולי, באנשים החולים בסוכרת ובכאלה שאינם. אוכלוסיית המחקר נלקחת מתוך קופת מאוחדת, המונה כ1.2 מיליון מבוטחים. המיקוד הוא במבוטחים מעל לגיל 18, כאשר חולה חיובי בקורונה מוגדר ככזה שאובחן בבדיקת PCR כחולה. גורמים נוספים שנלקחו בחשבון הם BMI, גיל, מגזר ומין. כמו כן, על מנת להימנע מהשפעת החלק האקוטי של המחלה על ערכי הגלוקוז, ערכים שנלקחו 30 יום לפחות לפני בדיקה חיובית לקורונה נלקחו בחשבון בלבד.

לצורך קבלת הקשר בוצעה רגרסיה לוגיסטית בין הפרמטרים שהוזכרו לעיל ובין חומרת הקורונה. כמו כן בוצעה השוואה מול אוכלוסיות בקרה (אוכלוסיות בעלות מאפיינים דומים למעט רמת הגלוקוז שלא חלו בקורונה חמורה) על מנת לוודא את מובהקות התוצאות. P-value של 0.05 הוגדר כקשר סטטיסטי מובהק בין גורם נבדק לבין הסיכוי לחלות בקורונה חמורה.

37,121 מבוטחים היו בעלי בדיקת קורונה חיובית. 15,563 מתוך החיוביים לקורונה היו עם בדיקת גלוקוז מתאימה. התוצאות הראו שחולים בעלי BMI גבוה, בגיל גבוה וגברים הם בעלי סיכון יותר גבוה לחלות בקורונה חמורה. התוצאות הציגו קשר מובהק בין סוכרת לבין בדיקת הגלוקוז ולכן בוצע ניתוח נפרד לחולי סוכרת ולכאלו שאינם חולים.  
לכאלו שאינם חולים בסוכרת התוצאות הראו קשר מובהק בין ערכי גלוקוז גבוהים לבין הסיכוי לחלות בקורונה חמורה, כאשר ככל שערך הגלוקוז בדם עולה (בקטגוריות של 105>,125>,140>) הסיכוי לחלות בקורונה חמורה עולה אף הוא. ערכי גלוקוז אלו מוגדרים כטרום-סוכרתיים. בחולי הסוכרת התוצאות היו מפתיעות, כאשר בקבוצות של מעל 125 ומעל 140 נצפה סיכוי גבוה יותר, אך גם בקבוצה של חולי סוכרת בעלי ערכי סוכר הקטנים מ105. ייתכן שתוצאה זו נובעת מכך שחולי סוכרת בעלי ערכי סוכר כאלו לרוב מטופלים באינסולין, מה שמוביל לערכי גלוקוז נמוכים, אך מעיד על מחלה קשה יותר.

לסיכום, ערכי גלוקוז גבוהים ומחלת הסוכרת מגדילים את הסיכון לקורונה חמורה. בפרט ערכים גבוהים של גלוקוז באנשים ללא אבחנה לסוכרת, וכן ערכים נמוכים בחולי סוכרת הוכחו כגורמים משמעותיים.

## **2.3 מאמר 3 - Increased Risk of New-Onset Diabetes after Recovery**

מספר מאמרים הוכיח את הקשר בין קורונה לערכי גלוקוז גבוהים ולחלות בקורונה. מטרת המאמר היא לסקור את המאמרים האלה ולאפיין את גורמי הסיכון העיקריים לחלות בסוכרת לאחר החלמה מקורונה.

במחקרים שחקרו השפעות כלליות של מחלת הקורונה לטווח ארוך שבוצעו בארה"ב וכן באנגליה נמצא כי באופן מובהק הסיכון לחלות בסוכרת לאחר החלמה מקורונה מתגבר. בפרט נמצא שהסיכון גדול בחולים מתחת לגיל 70 ושאינם לבנים. לצד סוכרת גם מחלות נוספות כגון מחלות לב וסיבוכים נשימתיים נמצאו בסיכון מוגדר אצל מחלימי קורונה. באופן מפתיע, דווקא בחולים שחלו בדלקת ראות בעקבות המחלה, רק סוכרת נמצאה עדיין בסיכון גבוה.

במחקרים שהתמקדו רק בסיכון לחלות בסוכרת (לצד גורמים נוספים כגון משקל, מין וכו') נמצא ב2 מחקרים שנעשו על הפורשים מצבא ארה"ב כי הסיכון גדל. במאמר השני נמצא כי הסיכון הוא רק בגברים, אם כי מאגר המידע הכיל בעיקר גברים ולכן לא ניתן לומר זאת באופן מובהק סטטיסטית. המאמר הראשון שצוטט הוא המאמר הראשון לעיל. מחקר נוסף שנשען על בסיס נתונים פדרלי מצא גם הוא קשר סטטיסטי מובהק, אם כי לאחר השמטה של חולים שקיבלו טיפול סטרואידי הקשר נחלש (אם כי עדיין מובהק סטטיסטית).

מחקר יחיד נמצא החוקר את הקשר בין קורונה לסוכרת סוג 1 בלבד. מחקר זה עמד על הקשר בנוסף לפרמטרים גיל, מין ועבודת הנחקר. מחקר זה לא מצא קשר מובהק, אך מדובר במחקר שלא בוצע עבורו ביקורת עמיתים, לא פורסם מאגר המידע שלו ומאפיינים נוספים היו חסרים.

מחקר שבוצע בילדים מצא גם הוא קשר מובהק בין הסיכון לחלות בסוכרת, אך מאמר זה לא מתמקד בסוכרת חדשה ולכן לא ניתן להצביע ממנו על קשר כזה.

המנגנון בגללו קורונה מעלה את הסיכון לסוכרת אינו ברור במלואו, אך מספר סיבות אפשריות לכך. קורונה גורמת לשחרור מוגבר של ציטוקינים, ושחרור זה יכול לגרום לעמידות לאינסולין שמובילה לסוכרת. מנגנון נוסף קשור למבנה התאים בלבלב שעלול להיות מושפע מקורונה ולהוביל לייצור מופחת של אינסולין.

לסיכום, מספר מאמרים מראים סיכון מוגדל לחלות בסוכרת לאחר קורונה, בפרט בסוכרת סוג 2. מרבית המאמרים אינם מספיק מדויקים, בעיקר בגלל העובדה שעבר מעט מדי זמן ממגפת הקורונה בשביל לייצר מעקב מספק. כמו כן, ייתכן ולא כלל הגורמים נבדקו ונעשה בידוד משתנים מספק על מנת לאפיין את הקורונה כגורם העיקרי. בין כה וכה, הקשרים שנמצאו מספקים כדי להוות גורם מתריע בפני רופאים ומטופלים כאחד.

# **3.** **תיאור העבודה**

על מנת לחקור את נושא השפעת הקורונה על מדדי הסוכר, התמקדתי תחילה בקריאת חומרים ומאמרים על מחלת הסוכרת – מהי מחלת הסוכרת, מה ההשלכות שהיא גוררת עימה, ההבדלים בין סוכרת מסוג 1 לסוכרת מסוג 2, ממה היא נובעת, מהם התסמינים העיקריים שלה ואיך מאבחנים אותה, מדדי הסוכר העיקריים וכו'.

כמו כן, קראתי על מאמרים העוסקים בקשר בין קורונה לסוכרת ועל מחקרים שהתבצעו עד עתה בתחום.

בנוסף, למדתי באופן עצמאי את תוכנת ,R Studio תוכנה סטטיסטית המאפשרת יכולת של ניתוח נתונים ומסקנות סטטיסטיות.

הלמידה של תוכנה זו כללה את למידת שפת התכנות, כתיבת פקודות ופונקציות, כתיבת רגרסיות והוצאת פלטים וגרפים.

לאחר הבנה עמוקה את הנ"ל, קיבלתי את מאגר הנתונים.

מאגר הנתונים כולל בתוכו 221 עמודות ו- 1,072,521 שורות.

מאגר זה מתייחס למיליוני מטופלי קופ"ח מאוחדת, אשר חלקם נדבקו בקורונה וחלקם לא.

העמודות מציגות את סוג המידע על אותו מטופל והשורות את ערכי העמודות פר מטופל.

מאגר זה מכיל 3 אוכלוסיות: מטופלים שחלו בקורונה חמורה, מטופלים שחלו בקורונה קלה ומטופלים שלא חלו בקורונה כלל.

תחילה ביצעתי סיווג ראשוני על הנתונים, כאשר ביצעתי הפרדה למאגר אשר יכלול בתוכו מטופלים שחלו בסוג קורונה כלשהי ומאגר למטופלים שלא חלו בקורונה.

סיווג זה התבצע דרך תוכנת ,R Studioולאחר מכן ייבאתי את הקבצים שהתקבלו לאקסל (מובא בנספח 1).

לאחר מכן, ביצעתי סיווג נוסף על גבי הסיווג הראשון, סיווגתי את מאגר המידע שכולל בתוכו את המטופלים שחלו בסוג קורונה כלשהי – כאשר מאגר אחד כולל בתוכו מטופלים שחלו בקורונה חמורה בלבד ומאגר שני הכולל בתוכו מטופלים שחלו בקורונה קלה בלבד (מובא בנספח 2).

הסיווג התבצע באותה הדרך של הסיווג הראשוני, תוך אבחנה של מטופלים שחלו בקורונה חמורה, המקיימים את אחד מ-3 התנאים הבאים:

1. מטופל שאושפז יותר מ-10 ימים בביה"ח.

2. מטופל שאושפז בטיפול נמרץ.

3. מטופל שקיים לו תאריך פטירה במהלך תקופת המעקב.

שני מדדי הסוכר לבדיקה:

FASTING BLOOD GLUCOSE ( FBG) .1 - רמת הסוכר בדם לאחר צום.

HBA1C .2 – רמת הסוכר הממוצעת בדם במשך 3 החודשים האחרונים.

טווח ערכים FBG : (מיליגרם לדציליטר)

70-100= תקין

100-125= טרום סוכרת

126+= סוכרת

טווח ערכים HBA1C : (מילימול גלוקוז למול דם)

4-5.7 = תקין

5.7-6.4 = טרום סוכרת

6.5+= סוכרת

התנאי הוא– עבור כל אחד מהמטופלים חייב להיות את ערכי הסוכר לפני ואחרי.

עבור מטופלים שחלו בקורונה:

ערך לפני= ערך סוכר שהתקבל לפני התחלואה בקורונה.

ערך אחרי= ערך סוכר שהתקבל לאחר התחלואה בקורונה.

עבור מטופלים שלא חלו בקורונה:

ערך לפני= ערך סוכר שהתקבל לפני תקופת המעקב.

ערך אחרי= ערך סוכר שהתקבל במהלך תקופת המעקב.

במהלך הפרויקט, קראתי לערך המדד לפני "BEFORE" ואחרי נקרא "AFTER".

## **3.1 מבחן T למדגמים תלויים**

נשתמש במבחן זה כאשר אנו מעוניינים להשוות בין שתי קבוצות שיש ביניהן תלות.

מדובר על השוואת המדד בתחילת תקופת המעקב לסיום תקופת המעקב, של **אותו מטופל**.

במבחן זה רציתי לבחון האם מדד הגלוקוז של המטופל השתנה בעקבות הקורונה (בממוצע).

לשם כך יצרתי משתנה המהווה את הפרש מדדי הגלוקוז של אותו מטופל: AFTER-BEFRE ומתאר את השינוי שחל באותו מדד סוכר.

חישוב משתנה זה בוצע על כל אחד מהמטופלים שחלו בקורונה חמורה או קורונה קלה.

משתני מבחן זה, הם ארבעת המשתנים הללו[[1]](#footnote-1):

1. גיל: לפי עשורים.
2. מגזר: אורתודוכסי, ערבי וכללי.
3. מין (משתנה בינארי): 0=זכר, 1=נקבה.
4. סטטוס עישון (משתנה בינארי): 0=לא מעשן, 1=מעשן.

מבחן זה התבצע על 2 אוכלוסיות המחקר בלבד מתוך השלוש ונבחן עבור כל אחד מהמשתנים הנ"ל.

השערות המבחן הנ"ל הינן:

H0: מדד FBG של המטופל לא השתנה בעקבות הקורונה (בממוצע).

H1: מדד FBG של המטופל השתנה בעקבות הקורונה (בממוצע).

במידה ונקבל מובהקות של 5% P.value<, נדחה את השערת H0 ונקבל את השערת H1.

במידה ונקבל מובהקות של 5% ,P.value> נקבל את השערת H0 ונדחה את השערת H1.

במידה ואכן נראה את מובהקות המבחן, נרצה לראות מהי המשמעות של השינוי במדד הסוכר- האם חלה ירידה בערך הסוכר או שמא עלייה.

לשם כך, חישבתי עבור אותם המשתנים שיצאו לי מובהקים, את ממוצע מדד הגלוקוז בתחילת תקופת המעקב ובסיום תקופת המעקב והצגתי זאת על גבי גרף מתאים.

בפרק התוצאות אראה בנוסף לגרף שהתקבל, את מובהקות המבחן (P.value) – ככל שהמובהקות קרובה יותר לערך 0, כך המבחן יותר מובהק.

## **3.2 ניתוח ANOVA+ מבחן T למדגמים בלתי תלויים**

בניתוח זה ישנה השוואה בין 3 האוכלוסיות: קורונה חמורה, קורונה קלה וללא קורונה.

המשתנה שעליו ביצעתי את המבחן הוא תוחלת של AFTER-BEFORE, עליו כבר הסברתי לעיל.

במבחן זה בחנתי איזו אוכלוסייה שונה בתוחלתה משאר תוחלות האוכלוסיות.

כאן לא ביצעתי הפרדה של כל אוכלוסייה לארבעת המשתנים, כלומר כאן המשתנים לא באים לידי ביטוי.

תוצאה מובהקת מתרחשת כאשר: הערך הקריטי קטן מהערך הסטטיסטי.

השערות המבחן הינן:

H0: µ1= µ2= µ3

H1: At least one µ is different.

השערת H0 טוענת כין אין הבדל בין התוחלות ואילו השערת H1 טוענת כי ישנה לפחות תוחלת אחת ששונה משאר התוחלות.

אסביר את משמעות התוחלות:

µ1= E(After COVID19 - Before COVID19)

מתייחסת לממוצעי הפרשי מדד הגלוקוז בקרב מטופלים שחלו בקורונה קלה.

µ2 = E(After severe COVID19 - Before severe COVID19)

מתייחסת לממוצעי הפרשי מדד הגלוקוז בקרב מטופלים שחלו בקורונה חמורה.

µ3 = E(After research started - Before research started)

מתייחסת לממוצעי הפרשי מדד הגלוקוז בקרב מטופלים שלא חלו בקורונה כלל.

ערכי תוחלות אלו שהתקבלו, יובאו בפרק התוצאות.

לאחר ניתוח זה, הוספתי למבחן גם את 4 המשתנים.

עבור כל תוצאה מובהקת שיצאה, ביצעתי את מבחן T למדגמים בלתי תלויים, שכן מדובר ב-3 אוכלוסיות שלא תלויות זו בזו, וזאת על מנת להצביע על האוכלוסייה שגורמת לשוני בתוחלתה מבין שאר האוכלוסיות.

במידה ונקבל מובהקות של 5% ,P.value< נדחה את השערת H0 ונקבל את השערת H1.

במידה ונקבל מובהקות של 5% ,P.value> נקבל את השערת H0 ונדחה את השערת H1.

התוצאות שהתקבלו, מופיעות במפורט בפרק התוצאות.

## **3.3 רגרסיה ליניארית רבת משתנים**

רציתי כעת לבחון את הקשר בין משתנים בלתי תלויים למשתנה המוסבר התלוי, ואת מידת הרלוונטיות של המשתנה המסביר.

היתרון של רגרסיה זו: נטרול השפעות מתערבות (יותר חותר לקשר סיבתי).

כאשר המשתנים המסבירים הם 4 המשתנים הידועים, השייכים ל-3 אוכלוסיות הקורונה, והמשתנה המוסבר הוא ה-AFTER-BEFORE.

ישנם 2 שלבים מקדימים שביצעתי תחילה:

1. סיווג המגזרים למשתנים בינאריים.
2. סיווג 3 האוכלוסיות שלנו למשתנים בינאריים.

המשוואה שהתקבלה תיתן את ערכו הצפוי של שינוי מדד הגלוקוז, פר מטופל, ועד כמה כל משתנה משפיע על ערך צפוי זה.

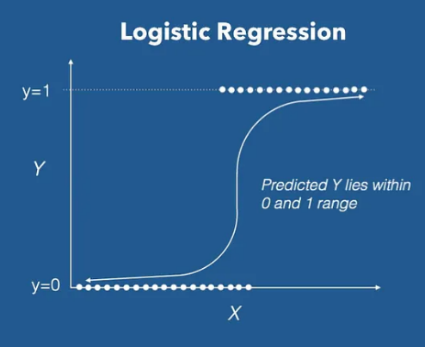
ארחיב על כך בפרק התוצאות.

## **3.4 רגרסיה לוגיסטית**

כעת בחנתי משתנה מוסבר AFTER DIABETE, כלומר ברצוני לחזות האם למטופל תהיה סוכרת בסיום תקופת המעקב.

משתנה מוסבר זה הינו משתנה בינארי, מה שמתאים לבחון את משקלם של המשתנים המסבירים את המשתנה המוסבר בעזרת רגרסיה לוגיסטית- בה המשתנה התלוי הינו בינארי.

זהו מודל הסתברותי שהפלט שלו הוא תוצאה של נוסחת הרגרסיה, המביאה תוצאה מספרית בין 0 ל-1 ועל פי ערך הסתברותי זה, נדע את הסיכוי של מטופל לחלות בסוכרת אחרי (ראה איור 1).



איור 1 – התפלגות לוגיסטית

לפני ביצוע הרגרסיה, הצטרכתי לבצע כמה שלבים מקדימים[[2]](#footnote-2):

1. הפיכת 3 האוכלוסיות למשתנים קטגוריאליים (מובא בנספח 3).
2. הפיכת 3 המגזרים למשתנים קטגוריאליים (מובא בנספח 3).
3. המרת משתנה Glucose\_Bef למשתנה בינארי BEF DIABATE.
4. המרת משתנה Glucose\_Aft למשתנה בינארי AFT DIABATE.

שלב 1:

המרתי את סוגי הקורונה למשתנים קטגוריאליים, כאשר: ללא קורונה=1, קורונה חמורה=2, קורונה קלה=3.

שלב 2:

באותה מידה ביצעתי זאת עבור המגזרים הקיימים, כאשר הקטלוג התבצע כך:

מגזר ערבי=1, מגזר כללי=2 ומגזר אורתודוכסי=3.

שלב 3:

הוספתי משתנה מסביר נוסף לרגרסיה זו, שמבטא סוכרת לפני- כלומר, האם למטופל קיימת בתחילת תקופת המעקב סוכרת.

גם כאן הביטוי "לפני" או "אחרי", דומה במשמעותו לערך הגלוקוז "לפני" ו"אחרי, שהסברתי בתחילה.

לשם כך, המרתי את משתנה Glucose\_Bef ל - ,BEF DIABATEבעזרת סיווג, דרך טווחי הגלוקוז המוזכרים לעיל.

כל ערך גלוקוז של מטופל שעומד על יותר מ-126 (מיליגרם לדציליטר) , מעיד על מטופל החולה בסוכרת ויסווג כ-1, ובמידה וקטן מ-126 (מיליגרם לדציליטר), יסווג כ-0, מפני שמטופל זה אינו חלה בסוכרת לפני.

שלב 4:

בתחילת ההסבר הזכרתי כי רציתי שכעת המשתנה המוסבר יהיה AFTER DIABETE.

ולכן פעלתי באותה הדרך של שלב 3, רק על ערך Glucose\_Aft, וקיבלתי מהסיווג את משתנה

AFT DIABATE.

לאחר סיום ביצוע 4 שלבים אלו, בניתי את הקוד בתוכנת R studio, לקבלת הרגרסיות הלוגיסטיות (מובא בנספח 3).

2 סוגי רגרסיות לוגיסטיות שבניתי:

1. רגרסיה לוגיסטית עם כל המשתנים שהיו עד כה, עם הוספת משתנה מסביר Glucose\_Bef.
2. רגרסיה לוגיסטית עם כל המשתנים שהיו עד כה, עם הוספת משתנה מסביר bef\_diabetes.

כאשר בשתיהן המשתנה המוסבר הינו aft\_diabetes.

ביצוע כל אחת מהרגרסיות, התבצע על כל 3 האוכלוסיות (אוכלוסיות הקורונה), כך שסה"כ מדובר על 6 רגרסיות שהתקבלו.

התוצאות שהובאו בפרק התוצאות, הינן תוצאות הרגרסיה הלוגיסטית בעלת מדד הדיוק הגבוה ביותר, שאסביר עליו בהמשך - שם יהיה ניתן לראות את מודל הרגרסיה הלוגיסטית שהתקבל ואת המשתנים המובהקים שהתקבלו ממנו.

לאחר מכן, רציתי להתעמק בטיב המודל, עד כמה המודל הוא מדויק- כיוון שמדובר במודל הסתברותי, הוא אינו יכול להיות מדויק ב-100%.

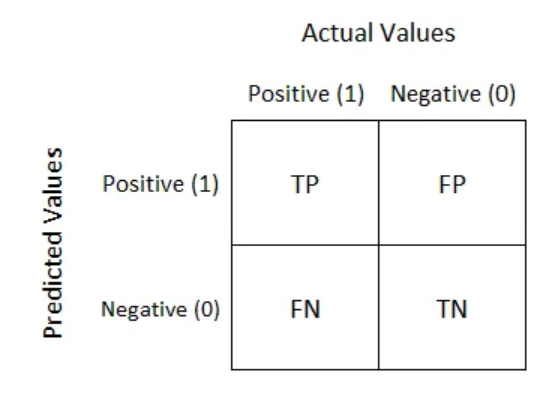
בחנתי את טיב המודל באמצעות:

* בניית CONFUSION MATRIX
* עקומת ROC
* חישוב שטח AUC (השטח שמתחת לעקומה)
* חישוב מדד הדיוק Accuracy
* P.value של המודל

**CONFUSION MATRIX**

עבור כל אחת מהרגרסיות התעניינתי לקבל את מטריצת הבלבול.

מטריצה זו תשקף את מצב סיווג המודל, כאשר התגית החיובית היא מי שיחלה בסוכרת אחרי (ראה איור 2).



איור 2 – מודל confusion matrix

העמודות מציגות את התגיות שסיווג המודל, והשורות את הסיווג בפועל.

הטבלה היא מטריצה של 2X2 כיוון שבמקרה שלנו קיימות רק שתי אפשרויות – יחלה בסוכרת אחרי או שלא.

הסבר:

True Positives (TP) - המודל חזה נכונה את מי שחולה בסוכרת אחרי.

True Negatives (TN) - המודל חזה נכונה את מי שאינו חולה בסוכרת אחרי.

False Positives (FP) - המודל שגה וזיהה בטעות מטופל שאינו חולה בסוכרת אחרי כמי שחולה בסוכרת אחרי.

False Negatives (FN) - המודל שגה וזיהה בטעות מטופל שחלה בסוכרת אחרי כמי שלא חלה בסוכרת אחרי.

לשם כך הצטרכתי ב:

1. סט אימון.
2. סט בדיקה.

הסבר:

1. סט אימון: סט הבדיקה שלי הם הנתונים האמיתיים של בסיס המידע שאותו קיבלתי, והוא בעצם ה- aft\_diabetes שעליו הסברתי לעיל.
2. סט בדיקה: מבוסס על הרגרסיה הלוגיסטית. מפלט הרגרסיה הלוגיסטית הרכבתי את התפלגות רגרסיה זו, כלומר את ההסתברות של אותו מטופל לחלות בסוכרת אחרי. התפלגות זו נראית כך:

סיווגתי את תוצאת ההסתברות בעזרת משתנה סף של 0.5.

כלומר, במידה וההסתברות גבוהה מחצי, אזי מדובר במטופל שיחלה בסוכרת אחרי ויקבל את הערך 1, במידה והתוצאה נמוכה מחצי, המטופל לא צפוי לחלות בסוכרת אחרי ויקבל ערך של 0.

לאחר מכן הכנסתי לתוכנת R studio את 2 המשתנים, הלא הם סט האימון וסט הבדיקה (מובא בנספח 3). וקיבלתי בתמורה את המטריצה, שעליה ארחיב בפרק התוצאות.

**עקומת ROC**

עקומה זו מתארת את היחס בין ה-TP כנגד ה-.FP

כלומר היחס בין מדד הרגישות לבין 1-Specificity.

כאשר:

Sensitivity - היחס של סיווגי אמת עבור דוגמאות שהם חיוביות במציאות (כלומר חולי סוכרת אמיתיים): TP / (TP + FN)

- Specificity הוא היחס של סיווגי אמת עבור דוגמאות שהם שליליות במציאות:

TN / (TN + FP)

נרצה להגביר דווקא את רגישות המודל, כי נרצה לאתר מטופלים שחלו בסוכרת אחרי.

כדי לקבל את עקומת ה-ROC , הכנסתי גם כן את הפרמטרים סט אימון וסט בדיקה וכתבתי קוד מתאים בתוכנה (מובא בנספח 3).

ניתן לראות את עקומת ה- ROC שהתקבלה, בפרק התוצאות.

עקומה אידיאלית תשיק לדפנות השמאלית והעליונה של הגרף כי היא תייצג מודל שהוא ספציפי ורגיש.

**AUC**

חישבתי את השטח שמתחת לעקומת ROC שבניתי, המכונה AUC - כלל שערכו גבוה יותר כך המודל נחשב לטוב יותר, כשהערך הגבוה ביותר של AUC הוא 1.

ככל שהמודל המסווג הוא טוב יותר, כך השטח מתחת לעקומה יהיה גדול יותר כי העקומה תשיק לדפנות השמאלית והעליונה של הגרף, הודות לרגישות וספציפיות גבוהות.

כתבתי פקודה לתוכנה שתפלוט את שטח זה וקיבלתי בתמורה את גודל השטח שמתחת לעקומה (מובא בנספח 3).

עם קבלת שטח זה, קיבלתי גם את ערכו של P.value

בפרק התוצאות ניתן לראות את ערכו של שטח זה, שהתקבל וכן את ערך P.value.

**ACCURACY**

זהו המדד הבסיסי והאינטואיטיבי ביותר, הוא מודד כמה סיווגים היו נכונים מתוך סך הסיווגים, להלן נוסחת הדיוק:

Accuracy = (TP +TN)/(TP+TN+FP+FN)

את מדד זה הפקתי מתוך המטריצה הנ"ל שנבנתה, ניתן לראות בפרק התוצאות.

## **3.5 יחס הסיכויים**

על רגרסיה זו חישבתי את יחס הסיכויים, עבור כל המשתנים המסבירים שלה.

יחס הסיכויים בוחן כיצד משתנה ההסתברות במעבר בין קבוצה אחת לשנייה.

הנוסחה לחישוב הינה:

כאשר גודל יחס הסיכויים יוצא 1 אזי אין קשר עם המשתנה התלוי, כי אם המעבר בין קבוצה אחת לשנייה הוא אותו דבר, אזי משתנה זה אינו משפיע על המשתנה התלוי.

כאשר גודל יחס הסיכויים קטן מ-1 אזי יש קשר שלילי עם המשתנה התלוי.

כאשר גודל יחס הסיכויים גדול מ-1 אזי יש קשר חיובי עם המשתנה התלוי.

ארחיב על כך, בפרק התוצאות.

## **3.6 ממשק משתמש**

לאחר מכן, בניתי כלי ממשק משתמש לרופאים שמסתמך על 2 הרגרסיות- הן על הרגרסיה הליניארית והן על הרגרסיה הלוגיסטית.

הכלי משמש לכלל המטופלים ולא רק למטופלים שקיים להם ערך גלוקוז לפני ואחרי.

זהו כלי שנבנה באקסל בשפת תכנות VBA (מובא בנספח 4).

כאשר הרופא יזין את פרטי המטופל בממשק שמסתמך על הרגרסיה הליניארית, הוא יקבל את ערך הגלוקוז הצפוי בסיום תקופת המעקב, עם משמעות ערך זה, כלומר האם הוא נמצא בטווח הנורמאלי של ערך הגלוקוז או בטווח של ערך שמהווה טרום סוכרת או שמא סוכרת גמורה.

כאשר הרופא יזין את פרטי המטופל בממשק שמסתמך על הרגרסיה הלוגיסטית, הוא יקבל את ערך ההסתברות של אותו מטופל לחלות בסוכרת אחרי.

ניתן לראות את הדוגמא לשימוש בממשק זה, בפרק התוצאות.

## **3.7 אבני דרך**

* ביצוע אינטגרציה בין 2 המדדים ולחקור את ההשפעה של 2 המדדים על ערכי הסוכר.
* בחינת משמעות ירידת מדד הסוכר אם היא מצביעה על איזושהי בעיה בעקבות הקורונה החמורה.

נרצה לאחר הבאת הממצאים, לבצע איחוד בין 2 מדדי הסוכר, כך שלכל המטופלים יהיה קיים את 2 מדדי הסוכר, הן בבדיקה שבתחילת תקופת המעקב והן בסיום תקופת המעקב.

נרצה לראות מאיחוד זה את ההשפעה שתנבע עכשיו על השינוי במדד הסוכר, אם קיים.

בנוסף, ירידת מדד הסוכר יכולה להוות גם משהו שלילי, ניתן דוגמא לשם כך: מטופל שנמצא בטווח הנורמאלי של ערך הגלוקוז וחלה לו ירידה במדד זה, עקב הקורונה החמורה, וכעת הוא סטה מערך הנורמל, אזי הוא נמצא במצב של היפוגליקמיה המהווה משמעת שלילית ולא דווקא חיובית- כלומר כעת, נרצה לבדוק את משמעות ירידת מדד הסוכר על טווח ערכי הסוכר ולא רק להתמקד בשינוי מדד הסוכר, אלא על משמעות ירידה זו או משמעות עלייה.

# **4.** **תוצאות**

## **4.1 תוצאות מבחן T למדגמים תלויים**

תמונה שמכילה טקסט, צילום מסך, מספר, גופן

התיאור נוצר באופן אוטומטי

איור 3 – גרף המחשה בשינוי ממוצע מדד הגלוקוז בקרב אוכלוסיית המעשנים

ניתן לראות כי בקרב אוכלוסיית המעשנים שחלו בקורונה חמורה, חלה ירידה משמעותית בממוצע הגלוקוז שלהם, לאחר החולי, וזאת לעומת אוכלוסיית הלא מעשנים שלא ניכר שום שינוי במדד:

ממוצע גלוקוז המעשנים לפני שחלו בקורונה: 121 מיליגרם לדציליטר.

ממוצע הגלוקוז המעשנים לאחר שחלו בקורונה: 109.5 מיליגרם לדציליטר.

חלה ירידה ממוצעת של: 11.5 מיליגרם לדציליטר.

מובהקות מבחן זה עומדת על P.value=0.0419.

סיכום הממצא שהתקבל: **אוכלוסיית הקורונה החמורה והן אוכלוסיית המעשנים, גורמים לשינוי במדד הגלוקוז וחשודים בהשפעתם על מדד הסוכר.**

## **4.2 תוצאות ניתוח ANOVA**

|  |  |
| --- | --- |
| **אוכלוסייה** | **p.value** |
| כל האוכלוסיות יחד | 0.0029 |
| זכר | 0.003523 |
| אורתודוכסי | 0.002868 |
| טווח גיל: 71-80 | 0.0000812 |
| טווח גיל: 81-90 | 0.04024 |
| טווח גיל: 91-100 | 0.031111 |
| מעשן | 0.0000439 |

טבלה 1 - משתנים מובהקים שהתקבלו מניתוח ANOVA

בטבלה זו ניתן לראות את כל המשתנים שקיבלו תוצאה מובהקת מניתוח זה, מובהקות זו מעידה על קיום של לפחות תוחלת אחת ששונה משאר התוחלות של אותה קבוצה מובהקת.

להלן תוחלות 3 אוכלוסיות הקורונה שהתקבלו ממבחן ANOVA(בנטרול חלוקה למשתנים):

µ1=1.143, **µ2=-0.8**, µ3=1.23

אם נסתכל על ממוצעים אלו שהתקבלו, ניתן לראות כי אכן קיים הבדל מובהק בין חולי קורונה חמורה לבין היתר.

נראה כי בקרב חולי קורונה חמורה חלה ירידה בערכי FBG ובקרב האחרים לא נמצאה ירידה כזו.

על מנת לבחון תוחלת של איזו אוכלוסייה היא השונה מהשאר, בקרב המשתנים, נראה את תוצאות מבחן T למדגמים בלתי תלויים שהתקבלו:

|  |  |  |  |
| --- | --- | --- | --- |
| **אוכלוסייה** | **מובהקות השוני בין התוחלות:**  **קורונה חמורה לללא קורונה** | **מובהקות השוני בין התוחלות:**  **קורונה חמורה לקורונה קלה** | **מובהקות השוני בין התוחלות:**  **קורונה קלה לללא קורונה** |
| זכר | 0.001384 | 0.000459 | - |
| אורתודוכסי | 0.003356 | 0.000221 | - |
| טווח גיל: 71-80 | 0.025381 | 0.018335 | - |
| טווח גיל: 81-90 | 0.014674 | 0.015022 | - |
| טווח גיל: 91-100 | 0.00998 | - | - |
| מעשן | 0.020064 | 0.017503 | - |

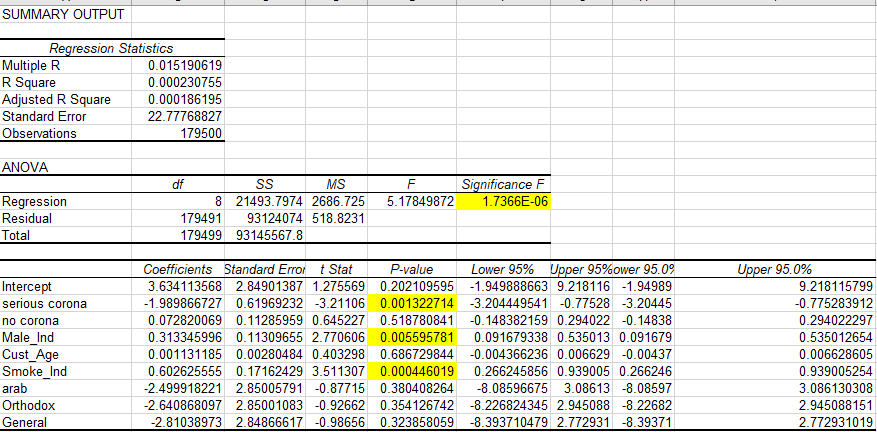
טבלה 2 – תוצאות מבחן T למדגמים בלתי תלויים

ניתן לראות כי המובהקות מתבטאת בהשוואתה של תוחלת הקורונה החמורה לשאר התוחלות, עבור כל אחד מהמשתנים, מלבד טווח הגילאים 91-100, ששם לא חלה מובהקות בין הקורונה החמורה לקורונה הקלה.

סיכום הממצא שהתקבל: **הקורונה החמורה חוזרת כאן בשנית בהשפעתה על מדד הסוכר, ומהווה כממצא שחוזר על עצמו ויכול להעלות איזושהי השערה על סוג קורונה זו.**

## **4.3 תוצאות רגרסיה ליניארית רבת משתנים**

חלק מפלט הרגרסיה שהתקבל[[3]](#footnote-3):



איור 4 – פלט מרכיבי הרגרסיה הליניארית להרכבתה

מפלט זה הרכבתי את משוואת הרגרסיה שנראית כך:

3.6-2\*X1+0.1\*X2+0.3\*X3+0.001\*X4+0.6\*X5-2.5\*X6-2.64\*X7-2.8\*X8=

כאשר:

X1= מטופל שחלה בקורונה חמורה

X2= מטופל שלא חלה בקורונה כלל.

X3= מטופל ממין זכר.

X4= גיל.

X5= מעשן.

X6= מגזר ערבי.

X7= מגזר אורתודוכסי.

X8= מגזר כללי.

רגרסיה זו מקבלת מובהקות גמורה של (ערכו של P.value) מה שמוסיף לדיוק המודל.

להלן המשתנים המובהקים של הרגרסיה הליניארית:

טבלה 3 – הצגת המשתנים המובהקים של הרגרסיה הליניארית

|  |  |
| --- | --- |
| **אוכלוסייה** | **P.value** |
| קורונה חמורה | 0.001 |
| זכר | 0.005 |
| מעשן | 0.0004 |

ניתן לראות כי המשתנים המסבירים שיצאו לנו מובהקים מרגרסיה זו, הינם: מטופלים שחלו בקורונה חמורה, מטופל ממין זכר ומטופל מעשן כלומר, אלו המשתנים שמשפיעים באופן מובהק על שינוי במדד הגלוקוז.

הממצא החשוב שהתקבל כאן - **אוכלוסיית המעשנים והקורונה חמורה חוזרות בהשפעתן על מדד הגלוקוז גם כאן.**

## **4.4 תוצאות רגרסיה לוגיסטית**[[4]](#footnote-4)

רגרסיה זו, שקיבלה את מרכיב הדיוק הגבוה ביותר, מתייחסת להוספת המשתנה המסביר bef\_diabetes.

להלן מודל הרגרסיה הלוגיסטית שהתקבל:



כאשר:

X1= מטופל שלא חלה בקורונה כלל.

X2= מטופל שחלה בקורונה חמורה.

X3= מטופל שחלה בסוכרת לפני.

X4= גיל.

X5= מין זכר.

X6= מעשן.

X7= מגזר כללי.

X8= מגזר אורתודוכסי.

להלן המשתנים המובהקים של רגרסיה זו:

|  |  |
| --- | --- |
| **אוכלוסייה** | **P.value** |
| מטופל שחלה בסוכרת לפני | 2\*10^-16 |
| גיל | 2\*10^-16 |
| זכר | 2\*10^-16 |
| מעשן | 3.23\*10^-12 |
| מגזר כללי | 2\*10^-16 |
| מגזר אורתודוכסי | 2\*10^-16 |

טבלה 4 – הצגת המשתנים המובהקים של הרגרסיה הלוגיסטית

ניתן לראות כי כל המשתנים יצאו לנו מובהקים, מלבד המשתנים המבטאים "סוג קורונה", דבר זה מפתיע, כיוון שעד כה הממצאים הראו איך הקורונה החמורה באה לידי ביטוי בהשפעתה על מדד הסוכר, ואילו כאן הקורונה החמורה אינה מסבירה באופן מובהק את הסיכוי לחלות בסוכרת אחרי.

ממצא מעניין זה יכול לשמש כהמשך מחקר עתידי, להסברת תופעה זו.

בחינת טיב המודל- מורכב מ-4 רכיבים:

1. CONFUSION MATRIX

תמונה שמכילה טקסט, צילום מסך, תוכנה, קו

התיאור נוצר באופן אוטומטינראה כעת את ה- confusion matrixשהתקבלה עבור רגרסיה זו:

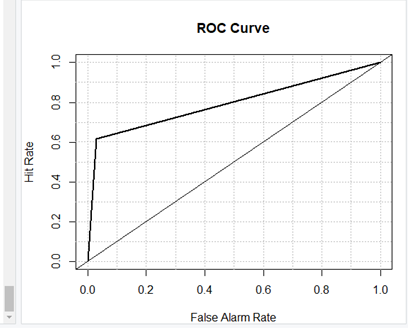
איור 5 – תוצאות סיווג confusion matrix

True Positives (TP) - המודל חזה נכונה (10,914) את מי שחולה בסוכרת אחרי.

True Negatives (TN) - המודל חזה נכונה (157,078) את מי שאינו חולה בסוכרת אחרי.

False Positives (FP) - המודל שגה וזיהה בטעות (4,662) מטופל שאינו חולה בסוכרת אחרי כמי שחולה בסוכרת אחרי.

False Negatives (FN) - המודל שגה וזיהה בטעות (6,846) מטופל שחלה בסוכרת אחרי כמי שלא חלה בסוכרת אחרי.

2.עקומת ROC

איור 6 – קבלת עקומת ROC

ניתן לראות, כי עבור רגישות של 0.6, נאלץ לשלם ב-0.01 בשגיאות מסוג FP.

3. שטח AUC

תמונה שמכילה טקסט, צילום מסך, תוכנה, קו

התיאור נוצר באופן אוטומטי

נראה כי השטח עומד על 0.8 וקרוב לערך 1, דבר המעיד על טיב המודל.

4. Accuracy



מדד הדיוק עומד על כ- 94%.

5. P.VALUE

P.value של מודל הרגרסיה:

**תמונה שמכילה טקסט, צילום מסך, תוכנה, קו

התיאור נוצר באופן אוטומטי**

נראה כאן כי ה-P.value הוא ממש 0 , מה שמעיד על מובהקות המודל ומחזק את נכונות המודל.

## **4.5 תוצאות יחס הסיכויים**

|  |  |
| --- | --- |
| **המשתנה הבלתי תלוי** | **ערך יחס הסיכויים** |
| זכר | 1.55 |
| מעשן | 1.23 |
| קורונה חמורה | 1.04 |
| גיל | 1.03 |
| ללא קורונה | 1.02 |
| אורתודוכסי | 0.7 |
| מגזר כללי | 0.65 |

טבלה 5 – תוצאות יחס הסיכויים

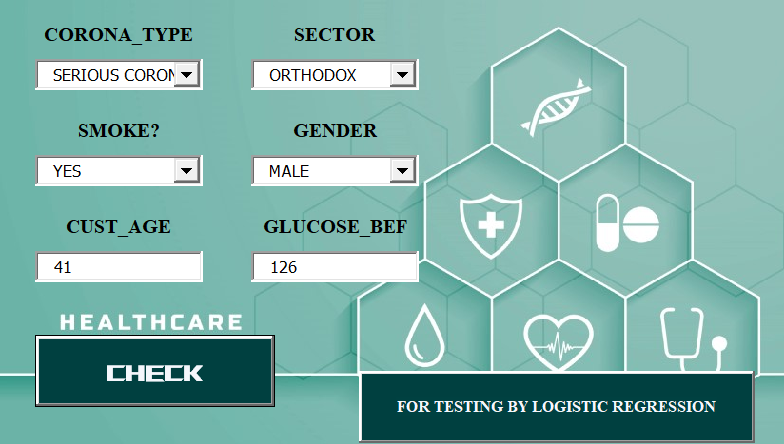
הסבר תוצאות:

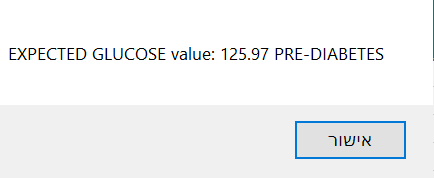
1. מטופל זכר: מטופל ממין זכר סיכויו לחלות בסוכרת אחרי, גדול פי 1.55 ממטופלת מהמין הנשי.
2. מטופל מעשן: מטופל שהוא מעשן, סיכויו לחלות בסוכרת אחרי, גדול פי 1.23 ממטופל שאינו מעשן.
3. מטופל שחלה בקורונה חמורה: מטופל שחלה בקורונה חמורה, סיכויו לחלות בסוכרת אחרי גדול פי 1.04 ממטופל שחלה בקורונה רגילה.
4. גיל: כל שנת גיל מעלה את הסיכוי לחלות בסוכרת אחרי, פי 1.03.
5. מטופל שלא חלה בקורונה: מטופל שלא חלה בקורונה, סיכויו לחלות בסוכרת אחרי גדול פי 1.02 ממטופל שחלה בקורונה רגילה.
6. מטופל מהמגזר האורתודוכסי: מטופל מהמגזר האורתודוכסי, סיכוי לחלות בסוכרת אחרי, קטן פי 0.7 ממטופל מהמגזר הערבי.
7. מטופל מהמגזר הכללי: מטופל מהמגזר הכללי, סיכוי לחלות בסוכרת אחרי, קטן פי 0.65 ממטופל מהמגזר הערבי.

## **4.6 תוצאות ממשק משתמש**

להלן הממשק שמסתמך על הרגרסיה הליניארית – קבלת ערך גלוקוז צפוי בסיום תקופת המעקב + משמעות ערך זה.

להלן דוגמא הממחישה שימוש בכלי זה:



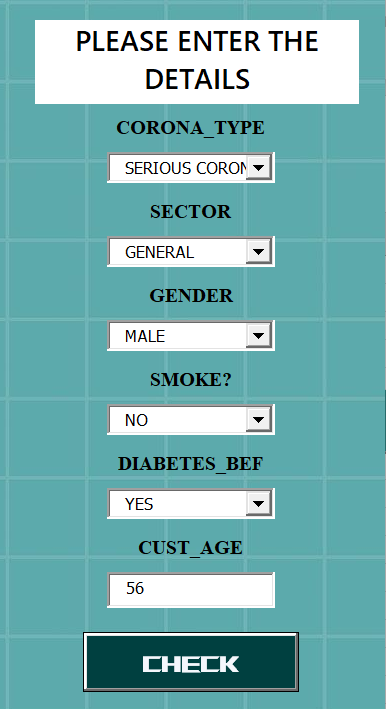


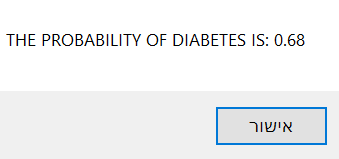
איור 7- הצגת שימוש בממשק משתמש על סמך הרגרסיה הליניארית

הזנת הרופא את פרטי המטופל, תיתן את התוצאה שתיארתי לעיל:

1. סוג הקורונה שהייתה למטופל.
2. מגזר המטופל.
3. מין המטופל.
4. ערך גלוקוז המטופל לפני.
5. האם המטופל מעשן.
6. גיל המטופל.

להלן הממשק שמסתמך על הרגרסיה הלוגיסטית – קבלת ערך הסתברותי לחלות בסוכרת אחרי.

להלן דוגמא הממחישה שימוש בכלי זה:



איור 8 - הצגת שימוש בממשק משתמש על סמך הרגרסיה הלוגיסטית

הזנת הרופא את פרטי המטופל, תיתן את התוצאה שתיארתי לעיל:

1. סוג הקורונה שהייתה למטופל.
2. מגזר המטופל.
3. מין המטופל.
4. האם המטופל מעשן.
5. האם הייתה למטופל סוכרת לפני.
6. גיל המטופל.

# 

# **5. סיכום**

מטרת הפרויקט הייתה לחקור את השפעות הקורונה על הטווח הארוך בתחום הסוכרת, לשם כך במהלך הפרויקט השתמשתי במספר כלים סטטיסטיים ובכלי כריית מידע בכדי לחקור זאת.

הגעתי לממצאים שיכולים להוות כהמשך מחקר עתידי, היכול למנוע אנשים להתדרדרות לסוכרת, אחר מעקב מתאים.

להלן הממצאים שהתקבלו מ-2 מדדי הסוכר:

* ממוצע מדד ה- HBA1C בקרב המגזר האורתודוקסי שחלו בקורונה חמורה, ירד משמעותית לאחר החולי.
* ממוצע מדד ה- FBG בקרב אוכלוסיית המעשנים שחלו בקורונה חמורה, ירד משמעותית לאחר החולי.
* יש הבדל מובהק בין חולי קורונה חמורה לבין היתר בממוצע הפרשי מדדי הסוכר (AFTER-BEFORRE).
* רגרסיה ליניארית: המשתנים המסבירים המובהקים - אוכלוסיית המעשנים ומטופלים שחלו בקורונה חמורה.
* רגרסיה לוגיסטית: בהתמקדות על מדד ה- HBA1C, כל המשתנים המסבירים יצאו מובהקים ואילו בהתמקדות על מדד ה- GLUCOSE, כל המשתנים המסבירים יצאו מובהקים, מלבד המשתנה סוג הקורונה.

לאחר קבלת ממצאים אלו, הגעתי למספר רעיונות להמשך ואלו ישמשו לאבני דרך של הפרויקט הנ"ל:

* ביצוע אינטגרציה בין 2 המדדים ולחקור את ההשפעה של 2 המדדים על ערכי הסוכר.
* בחינת משמעות ירידת מדד הסוכר אם היא מצביעה על איזושהי בעיה בעקבות הקורונה החמורה.

# **6. ביבליוגרפיה**

* NASSER Mikhail. (2022). Increased Risk of New-Onset Diabetes after Recovery from COVID-19
* Shauly-Aharonov M, Shafrir A, Paltiel O, Calderon-Margalit R, Safadi R, Bicher R, et al. (2021) Both high and low pre-infection glucose levels associated with increased risk for severe COVID-19: New insights from a population-based study. PLoS ONE 16(7): e0254847. <https://doi.org/10.1371/journal.pone.0254847>.
* Yan Xie, Ziyad Al-Aly. (2022). Risks and burdens of incident diabetes in long COVID a cohort study.

# **7. נספחים**

נספח 1- סיווג הדאטה למטופלים שחלו בקורונה ולמטופלים שלא חלו בקורונה כלל ויצירת קובץ אקסל.

corona<-covid\_may\_22\_2[covid\_may\_22\_2$Covid\_Ind>0 ,]

all<-covid\_may\_22\_2[covid\_may\_22\_2$Covid\_Ind>=0 ,]

View(all)

View(corona)

as.data.frame(corona)

as.data.frame(ff)

nocorona<-covid\_may\_22\_2[covid\_may\_22\_2$Covid\_Ind==0 ,]

View(nocorona)

as.data.frame(nocorona)

library(writexl)

getwd()

setwd("C:/Users/97255/Documents/final project")

write\_xlsx(all, "all.xlsx")

write\_xlsx(corona, "corona.xlsx")

write\_xlsx(nocorona, "nocorona.xlsx")

נספח 2- סיווג הדאטה עבור מטופלים שחלו בקורונה חמורה ויצירת קובץ אקסל.

serious<-corona[corona$Icu\_Clinic > 0 | corona$Hosp\_Days > 10 |corona$Death\_Date>1 ,]

View(serious)

as.data.frame(serious)

write\_xlsx(serious, "serious.xlsx")

**נספח 3 – רגרסיה לוגיסטית**

#המרת משתנים לקטגוריאליים.

aftdiabetes<-factor(make\_to\_logistic\_$`aft\_diabetes`, levels = c("1","0" ))

nocorona<-factor(make\_to\_logistic\_$`TYPE\_CORONA`, levels = c("1","2","3" ))

seriouscorona<-factor(make\_to\_logistic\_$`TYPE\_CORONA`, levels = c("2","1","3" ))

corona<-factor(make\_to\_logistic\_$`TYPE\_CORONA`, levels = c("3","1","2" ))

typesex<-factor(make\_to\_logistic\_$`Male\_Ind`, levels = c("0","1"))

typesexfemale<-factor(make\_to\_logistic\_$`Male\_Ind`, levels = c("1","0"))

typesector<-factor(make\_to\_logistic\_$`TYPE\_SECTOR`, levels = c("1","2","3","4" ))

typesmoke<-factor(make\_to\_logistic\_$`Smoke\_Ind`, levels = c("0","1" ))

typenosmoke<-factor(make\_to\_logistic\_$`Smoke\_Ind`, levels = c("1","0" ))

#חישוב רגרסיה לוגיסטית עבור משתנה מסביר Glucose\_Bef.

logaftseriouscorona<- glm(aft\_diabetes~seriouscorona+Glucose\_Bef+Cust\_Age+typesex+typesmoke+typesector, data = make\_to\_logistic\_, family =binomial() )

summary(logaftseriouscorona)

logaftnocorona< glm(aft\_diabetes~nocorona+Glucose\_Bef+Cust\_Age+typesex+typesmoke+typesector, data = make\_to\_logistic\_, family =binomial() )

summary(logaftnocorona)

logaftcorona<- glm(aft\_diabetes~corona+Glucose\_Bef+Cust\_Age+typesex+typesmoke+typesector,

data = make\_to\_logistic\_, family =binomial() )

summary(logaftcorona)

#חישוב רגרסיה לוגיסטית עבור משתנה מסביר bef\_diabetes.

logaft1seriouscorona<- glm(aft\_diabetes~seriouscorona+bef\_diabetes+Cust\_Age+typesex+typesmoke+typesector, data = make\_to\_logistic\_, family =binomial() )

summary(logaft1seriouscorona)

logaft1nocorona<- glm(aft\_diabetes~nocorona+bef\_diabetes+Cust\_Age+typesex+typesmoke+typesector, data = make\_to\_logistic\_, family =binomial() )

summary(logaft1nocorona)

logaft1corona<- glm(aft\_diabetes~corona+bef\_diabetes+Cust\_Age+typesex+typesmoke+typesector, data = make\_to\_logistic\_, family =binomial() )

summary(logaft1corona)

# הרצת קונפישין מטריקס עבור משתנה בסיס של קורונה קלה עם משתנה מסביר Glucose\_Bef.

testcorona1<- factor(c(train\_test$`aft\_diabetes-test5`))

traincorona1<- factor(c(train\_test$`logaftcorona-train`))

confusionMatrix(testcorona1, traincorona1)

# הרצת קונפישין מטריקס עבור משתנה בסיס של קורונה חמורה עם משתנה מסביר Glucose\_Bef.

testcoronase1<- factor(c(train\_test$`aft\_diabetes-test4`))

traincoronase1<- factor(c(train\_test$`logaftseriouscorona-train`))

confusionMatrix(testcorona1, traincorona1)

# הרצת קונפישין מטריקס עבור משתנה בסיס של ללא קורונה עם משתנה מסביר Glucose\_Bef.

testnocorona1<- factor(c(train\_test$`aft\_diabetes-test6`))

trainnocorona1<- factor(c(train\_test$`logaftnocorona-train`))

confusionMatrix(testcorona1, traincorona1)

# הרצת קונפישין מטריקס עבור משתנה בסיס של קורונה קלה עם משתנה מסביר bef\_diabetes.

testcorona<- factor(c(train\_test$`aft\_diabetes-test3`))

traincorona<- factor(c(train\_test$`logaft1corona-train`))

confusionMatrix(testcorona, traincorona)

# הרצת קונפישין מטריקס עבור משתנה בסיס של קורונה חמורה עם משתנה מסביר bef\_diabetes.

testcoronase<- factor(c(train\_test$`aft\_diabetes-test1`))

traincoronase<- factor(c(train\_test$`logaft1seriouscorona-train`))

confusionMatrix(testcorona, traincorona)

# הרצת קונפישין מטריקס עבור משתנה בסיס של ללא קורונה עם משתנה מסביר bef\_diabetes.

testnocorona<- factor(c(train\_test$`aft\_diabetes-test2`))

trainnocorona<- factor(c(train\_test$`logaft1nocorona-train`))

confusionMatrix(testcorona, traincorona)

#חישוב עקומת ROC לרגרסיה לוגיסטית עבור משתנה בסיס של קורונה קלה, עם משתנה מסביר Glucose\_Bef[[5]](#footnote-5).

library(verification)

test3<- c(train\_test$`aft\_diabetes-test1`)

train3<-c(train\_test$`logaftcorona-train`)

roc.plot(test3, train3)

#חישוב עקומת ROC לרגרסיה לוגיסטית עבור משתנה בסיס של קורונה חמורה, עם משתנה מסביר Glucose\_Bef.

library(verification)

test2<- c(train\_test$`aft\_diabetes-test1`)

train2<-c(train\_test$`logaftseriouscorona-train`)

roc.plot(test2, train2)

#חישוב עקומת ROC לרגרסיה לוגיסטית עבור משתנה בסיס של ללא קורונה, עם משתנה מסביר Glucose\_Bef.

library(verification)

test4<- c(train\_test$`aft\_diabetes-test1`)

train4<-c(train\_test$`logaftnocorona-train`)

roc.plot(test4, train4)

#חישוב עקומת ROC לרגרסיה לוגיסטית, עם משתנה מסביר bef\_diabetes.

library(verification)

test1<- c(train\_test$`aft\_diabetes-test1`)

train1<-c(train\_test$`logaft1seriouscorona-train`)

roc.plot(test1, train1)

#חישוב עקומת AUC לרגרסיה לוגיסטית עבור משתנה בסיס של קורונה קלה, עם משתנה מסביר Glucose\_Bef.

roc.area(test3,train3)

#חישוב עקומת AUC לרגרסיה לוגיסטית עבור משתנה בסיס של קורונה חמורה, עם משתנה מסביר Glucose\_Bef.

roc.area(test2,train2)

#חישוב עקומת AUC לרגרסיה לוגיסטית עבור משתנה בסיס של ללא קורונה, עם משתנה מסביר Glucose\_Bef.

roc.area(test41,train4)

#חישוב עקומת AUC לרגרסיה לוגיסטית, עם משתנה מסביר bef\_diabetes.

roc.area(test1,train1)

**נספח 4 – ממשק משתמש**

#כלי ממשק משתמש המסתמך על רגרסיה ליניארית.

Private Sub CommandButton1\_Click()

Intercept = 3.634113568

Intercept = Intercept + TextBox1 \* 0.001131185

If ComboBox1 = "NO CORONA" Then Intercept = Intercept + 0.072820069

If ComboBox1 = "SERIOUS CORONA" Then Intercept = Intercept - 1.989866727

If ComboBox2 = "GENERAL" Then Intercept = Intercept - 2.81038973

If ComboBox2 = "ORTHODOX" Then Intercept = Intercept - 2.640868097

If ComboBox2 = "ARAB" Then Intercept = Intercept - 2.499918221

If ComboBox3 = "MALE" Then Intercept = Intercept + 0.313345996

If ComboBox4 = "YES" Then Intercept = Intercept + 0.602625555

Intercept = Intercept + TextBox2

Sum = Round(Intercept, 2)

If Intercept <= 100 And Intercept >= 70 Then

MsgBox ("EXPECTED GLUCOSE VALUE: " & Sum & " NORMAL")

ElseIf Intercept >= 126 Then

MsgBox ("EXPECTED GLUCOSE VALUE: " & Sum & " DIABETES")

Else

MsgBox ("EXPECTED GLUCOSE value: " & Sum & " PRE-DIABETES")

End If

End Sub

Private Sub CommandButton2\_Click()

UserForm2.Show

End Sub

Private Sub TextBox1\_Change()

CUST\_AGE = UserForm1.TextBox1.Value

End Sub

Private Sub TextBox2\_Change()

GLUCUSE\_BEF = UserForm1.TextBox2.Value

End Sub

Private Sub UserForm\_Initialize()

ComboBox1.AddItem "CORONA"

ComboBox1.AddItem "SERIOUS CORONA"

ComboBox1.AddItem "NO CORONA"

ComboBox2.AddItem "GENERAL"

ComboBox2.AddItem "ORTHODOX"

ComboBox2.AddItem "ARAB"

ComboBox2.AddItem "UNKNOWN"

ComboBox3.AddItem "FEMALE"

ComboBox3.AddItem "MALE"

ComboBox4.AddItem "YES"

ComboBox4.AddItem "NO"

End Sub

#כלי ממשק משתמש המסתמך על רגרסיה לוגיסטית.

Private Sub CommandButton1\_Click()

Intercept = -4.8120277

Intercept = Intercept + TextBox1 \* 0.0320385

If ComboBox1 = "GENERAL" Then Intercept = Intercept - 0.4348721

If ComboBox1 = "ORTHODOX" Then Intercept = Intercept - 0.3637783

If ComboBox1 = "UNKNOWN" Then Intercept = Intercept + 0.7800159

If ComboBox3 = "MALE" Then Intercept = Intercept + 0.4371398

If ComboBox2 = "YES" Then Intercept = Intercept + 0.2129679

If ComboBox4 = "YES" Then Intercept = Intercept + 3.7191675

If ComboBox5 = "NO CORONA" Then Intercept = Intercept + 0.0182983

If ComboBox5 = "SERIOUS CORONA" Then Intercept = Intercept + 0.0414144

Intercept = Exp(Intercept)

reg = Intercept / (Intercept + 1)

Sum = Round(reg, 2)

MsgBox ("THE PROBABILITY OF DIABETES IS: " & Sum)

End Sub

Private Sub UserForm\_Initialize()

ComboBox1.AddItem "GENERAL"

ComboBox1.AddItem "ORTHODOX"

ComboBox1.AddItem "UNKNOWN"

ComboBox3.AddItem "FEMALE"

ComboBox3.AddItem "MALE"

ComboBox2.AddItem "YES"

ComboBox2.AddItem "NO"

ComboBox4.AddItem "YES"

ComboBox4.AddItem "NO"

ComboBox5.AddItem "CORONA"

ComboBox5.AddItem "SERIOUS CORONA"

ComboBox5.AddItem "NO CORONA"

End Sub

1. בכל פעם שאזכיר את המשתנים במהלך הדוח, הכוונה ל-4 משתנים אלו. [↑](#footnote-ref-1)
2. שלב 1 ו-2 נועדו כדי שתוכנת R Studio תבין כי מדובר במשתנים שאינם כמותיים. [↑](#footnote-ref-2)
3. ישנו משתנה בסיס שלא בא בפלט לידי ביטוי, אך הרגרסיה לוקחת אותו בחשבון והוא: מטופלים שחלו בקורונה קלה. [↑](#footnote-ref-3)
4. נתייחס בתוצאות אלו רק על הרגרסיה הלוגיסטית בעלת מרכיב הדיוק הגבוה ביותר. [↑](#footnote-ref-4)
5. הורץ רק פעם אחת ולא 3, כיוון שהנתונים מרגרסיה זו, עבור משתנה מסביר זה, יצאו זהים. [↑](#footnote-ref-5)