

## **1. מבוא**

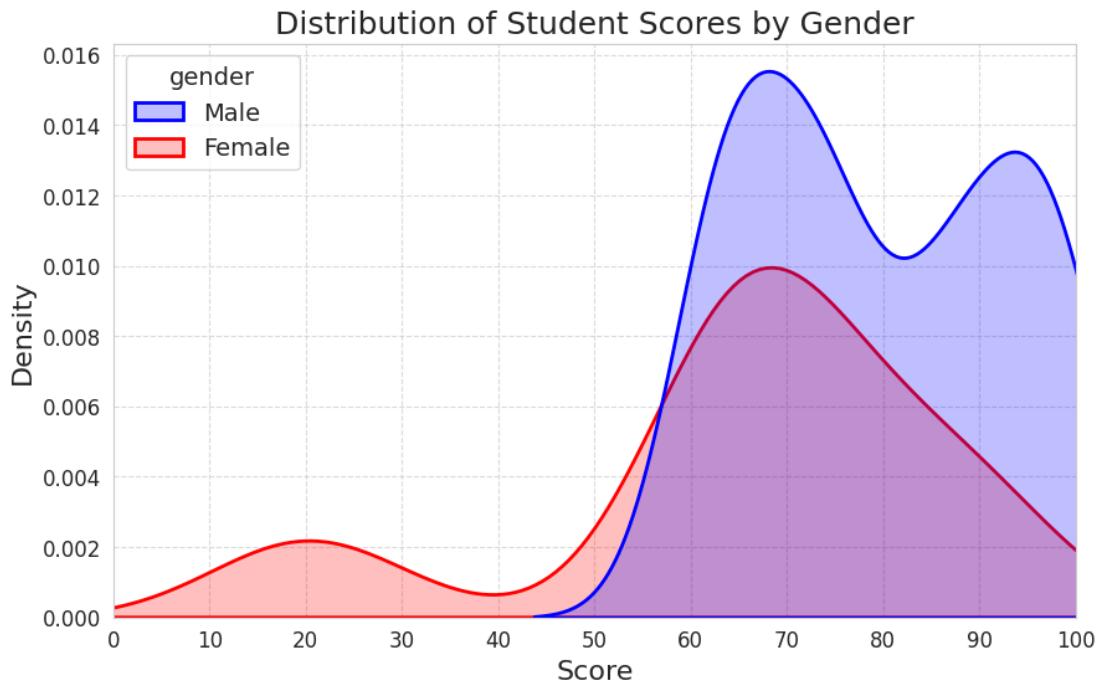
מאזור הננתונים "Higher Education Students Performance Evaluation" שנאסף בשנת 2019 מכיל מידע על 145 סטודנטים מהפקולטה להנדסה ומהפקולטה למדעים, מאוניברסיטה בקפריסן. הדadata כולל 31 משתנים, ביניהם משתנים דמוגרפיים, הרגלי לימוד והרגלי חיים. ניתוח מאזור הננתונים עשוי לסייע על הקשר בין הצלחה בלימודים (שנמדדת בדadata על ידי ציון בקורס) לבין משתנים דמוגרפיים וחברתיים ולהרגלי לימוד. השאלות שמעניינות אותנו קשורות להבדלים בין נשים וגברים בהצלחה בלימודים והרגלי הלימוד והעבודה ולאחר מכן השכלה ההורם להצלחתם של הסטודנטים.

בdatasheet המקורי, משתנה הציון הוא משתנה אורדינרי שמקבל ערכים של בין 0-7, כאשר 0 הוגדר כ"נכשל" ואילו 7 הוגדר כציון הגבוה ביותר. על מנת לשפר את איכות ניתוח הנתונים הגדרנו מנגן להמרת ציוניים איקוטיים לציוניים מסוימים. על פי שיטה זו, ציון מסופרי בטוח 0-59 מוגדר כנכשל. יתר הציוניים מחולקים לשבע קטגוריות איקוטיות (מ-1 עד 7) כאשר כל קטgorיה מייצגת טווח ציוניים שווה ככל הניתן. (כאשר 6 קטגוריות מכילות ערכים של 6 מספרים שלמים ואילו הקטגוריה של הציון הגבוה ביותר מכילה 5 מספרים שלמים). עבור כל סטודנט הגרלנו ציון לפי טווח הערכים שהגדכנו על מנת לשיר ציון מסופרי אמיתי (משתנה הציון עדין אינו רציף במלואו, מפני שמקבל רק ערכים שלמים אך מייצג ערכי ציוניים אמיתיים אשר ניתן לקבל). העבודה שטוחה הציוניים שוים וחסית קטנים מייצגת באופן דיאמין את הציוניים האמיתיים שהסטודנטים יכולים לקבל. המטרה זו מאפשרת השוואת פשטota וברורה בין ההישגים של סטודנטים, ועשוי להפוך את ניתוח הנתונים ליעיל יותר. טווחי הציוניים מוצגים להלן: 0 : .96-100 ; 1 : 0-59 ; 2 : 60-65 ; 3 : 66-71 ; 4 : 72-77 ; 5 : 78-83 ; 6 : 84-89 ; 7 : 90-95 .

## **2. התפלגות ציוניים כללית**

בבסיס הנתונים המקורי, בדקנו 58 נשים ו-87 גברים. מתוך 145 הסטודנטים שנבדקו, 8 נכשלו וכולן נשים. הציון החזיני בקרוב לכל הסטודנטים היה 73, בקרוב גברים 76 ואילו בקרוב נשים 68.5. התפלגות הציוניים בקרוב גברים ונשים מציגה מגמות שונות. אצל נשים ההתפלגות רחבה יותר מפני שיש קבוצה של ציוניים נמוכים (כל 8 הנכשלות הן נשים) עם שתי פסגות (20 ו-68), ואילו אצל גברים רוב הציוניים מרוכזים בטווחים גבוהים יותר עם שתי פסגות (70 ו-90). גם בטווח ציון "עובד" (60 ומעלה), ניתן לראות התפלגות שונה בין גברים לנשים.

### **תרשים 1: התפלגות הציוניים בקרוב גברים ונשים**



תרשים 1. התפלגות הציוןים בקרב גברים (בכחול) ונשים (באדום)

### 3. הרגלי למידה והבדלים בין- מגדריים

על מנת להבין טוב יותר את הבדלי הציוןים בין הסטודנטים ביחס להרגלי למידה, ואת הבדלי הציוןים בין גברים לנשים, נבדוק הבדלים בין הציוןים החזוניים ביחס להרגלי הלמידה. מתרשים 2 ניתן לראות שיש שיטות נומינטיביות שנוחים בכיתה הם בעלי ציון חזוני גבוה יותר (ק<sub>4</sub> = 0.094) ושיטות נומינטיביות שעובדים במהלך הלימודים הם בעלי ציון חזוני נמוך יותר (ק<sub>2</sub> = 0.085). בשאר הרגלי הלמידה כגון הקשבה בכיתה, סיכון החומר והרגלי שעות הלמידה לא נמצאו הבדלים בין הקבוצות. עבור משתנים בעלי יותר מ-2 ערכים אפשריים (מקשייבים, מסכמים, מספר שעות למידה) השתמשנו ב מבחן Kruskal-Wallis אשר בודק האם יש לפחות קבוצה אחת שונה באופן מובהק משאר הקבוצות.

עבור המשתנים בהם היו הבדלים בין הקבוצות, נוכחות בכיתה ועבודה במהלך הלימודים, וגם עבור המשתנה מספר שעות לימוד שבועית (אשר לו מצאנו שאין השפעה על הציון אבל הגיוני לשער שדווקא כן יש לו), בדקנו את הציוןים החזוניים בהפרדה מגדרית.

#### **nocחות בשיעורים**

מתרשים 3 עולה כי גברים שנכחו בשיעורים קיבלו ציון חזוני גבוה מאשר נשים שנכחו בשיעורים. אפקט זה לא משתחרר בתנאי "ללא נוכחות בשיעור", כאשר הציון החזוני של גברים ונשים דומה יותר עם חפיפה גדולה בטוחני הציון. בנוסף, ניתן לראות שהשפעה של נוכחות בשיעורים מוקורת בעיקר בגברים, כאשר כמעט אין הבדל בין הציון החזוני של נשים בשתי הקבוצות אך כן יש הבדל בקרב גברים שנכחו בשיעור (ציון גבוה יותר).

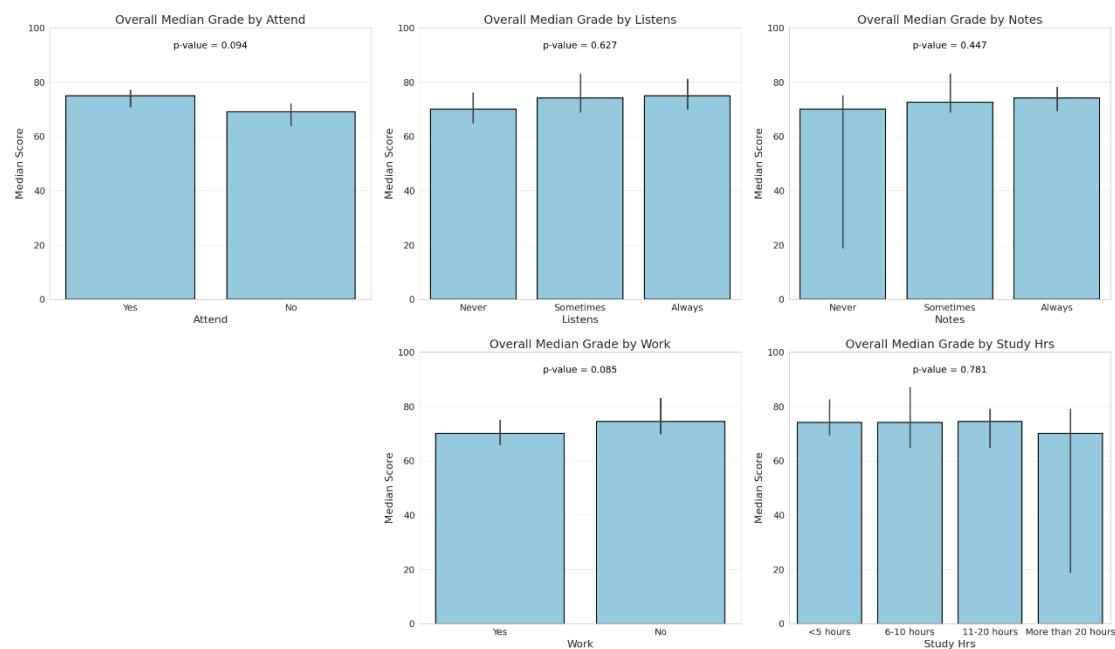
## עבודה במהלך הלימודים

מתרשים 4 עולה כי בשתי קבוצות העבודה (בן עובדים, לא עובדים) גברים קיבלו ציונים חיצוניים טובים יותר מאשר הנשים. בנוסף, נראה כי בקרב הנשים, עבודה במהלך הלימודים לא משפיעה על ההישגים בלימודים ואילו בקרב הגברים עבודה בלימודים קשורה בירידה קלה בהישגים.

### מספר שעות לימוד שבועית

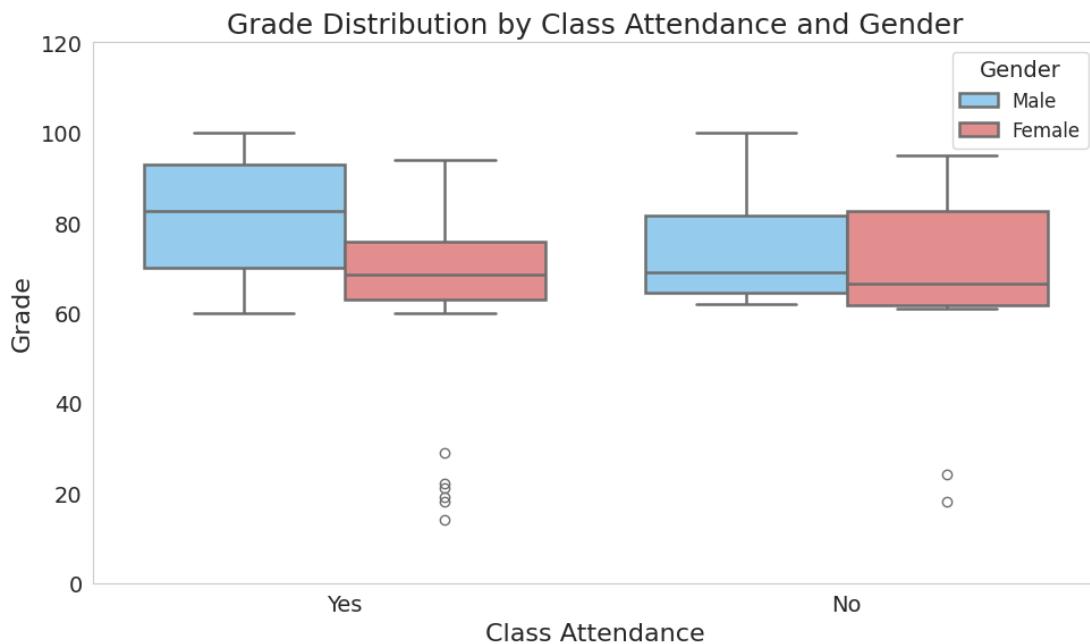
בקרב גברים, הציונים הגבוהים יותר מתקבלים דווקא בקבוצה של מספר שעות הלמידה השבועית הנמוך ביותר (5 שעות ומטה) כאשר החזון בקבוצה זאת גבוה במיוחד. בנוסף, בקרב הגברים, לא נראה שיש הבדל כל שלומדים יותר מ-5 שעות. אצל נשים נראה שלמידה של בין 10-11 שעות מפיקה את הציונים הגבוהים ביותר. בנוסף, הציונים של הנשים נמוכים משל הגברים בקבוצות "5 שעות ומטה"- "6-10", ואילו מעט גבוהים יותר כאשר מספר שעות הלימוד השבועית הוא בין 10-11 שעות. קבוצת "20 שעות ומעלה" מונה 4 סטודנטים בסה"כ, מתוכם 3 נשים וגבר אחד בלבד כך שהשוואה זאת אינה מלמדת אותנו הרבה. מתרשים 5 אנו למדים שאין קשר ברור בין מספר שעות הלמידה השבועית לבין הציון הסופי, במיוחד אצל ذكور, כאשר הציונים הגבוהים ביותר מתקבלים דווקא בקבוצות של מעט שעות למידה.

### תרשים 2: ציונים החיצוניים לפי נוכחות, הקשה, סיכון, עבודה בזמן לימודים ומספר שעות לימוד שבועיות



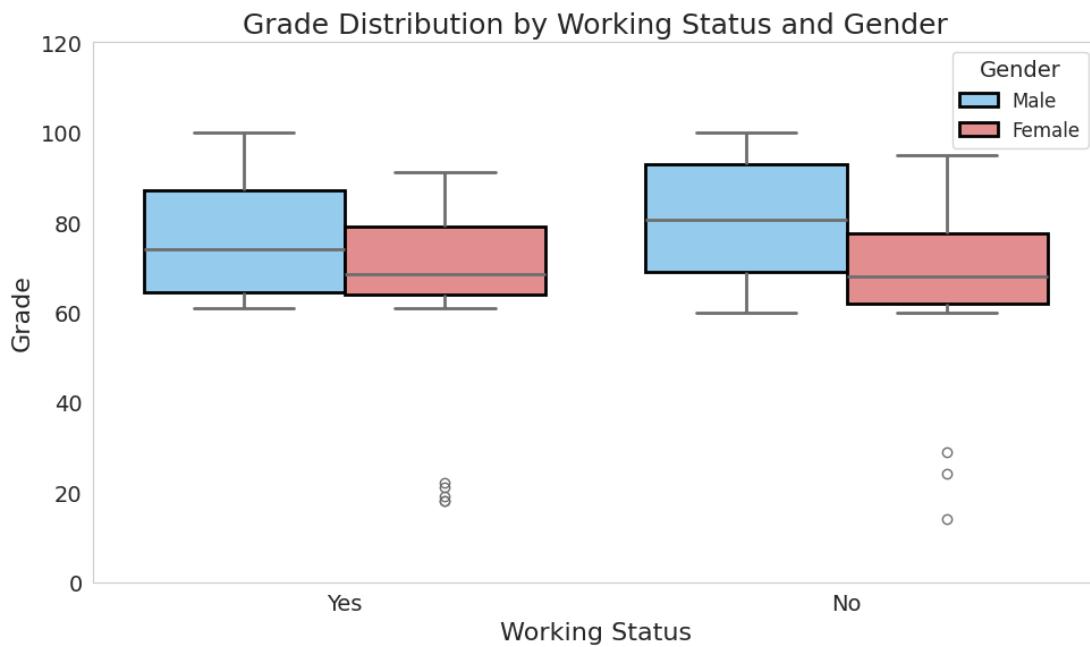
תרשים 2. הציונים החיצוניים לפי הרגeli למידה שונות וsegiaot התקן.

### **תרשים 3: התפלגות הציונים לפי נוכחות בכיתה ומגדר**



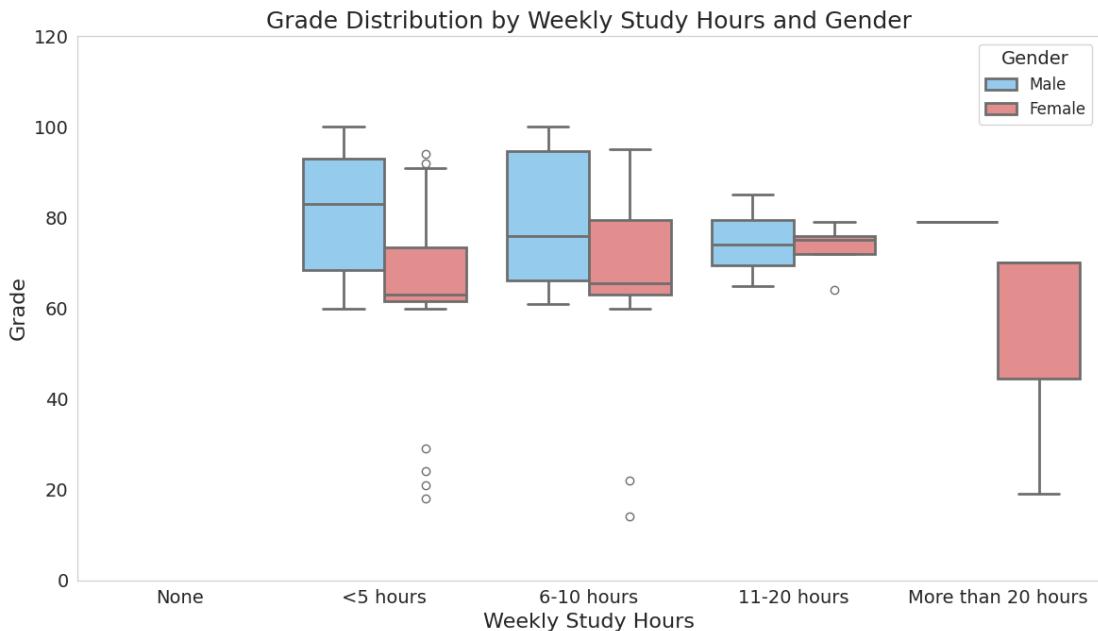
תרשים 3. התפלגות הציונים בקרב גברים (בכחול) ונשים (באדום) לפי נוכחות בкласс (nocחות – Yes, לא nocחת – No)

### **תרשים 4: התפלגות הציונים לפי סטטוס עבודה במהלך הסמסטר ומגדר**



תרשים 4. התפלגות הציונים בקרב גברים (בכחול) ונשים (באדום) לפי סטטוס עבודה במהלך הסמסטר (עובדות – Yes, לא עובדות – No)

## **תרשים 5: התפלגות הציונים לפי נוכחות בכיתה ומגדר**



**תרשים 5. התפלגות הציונים בקרב גברים (בכחול) ונשים (באדום) מספור שעות לימוד שבועית**

## **4. השכלה של ההורים**

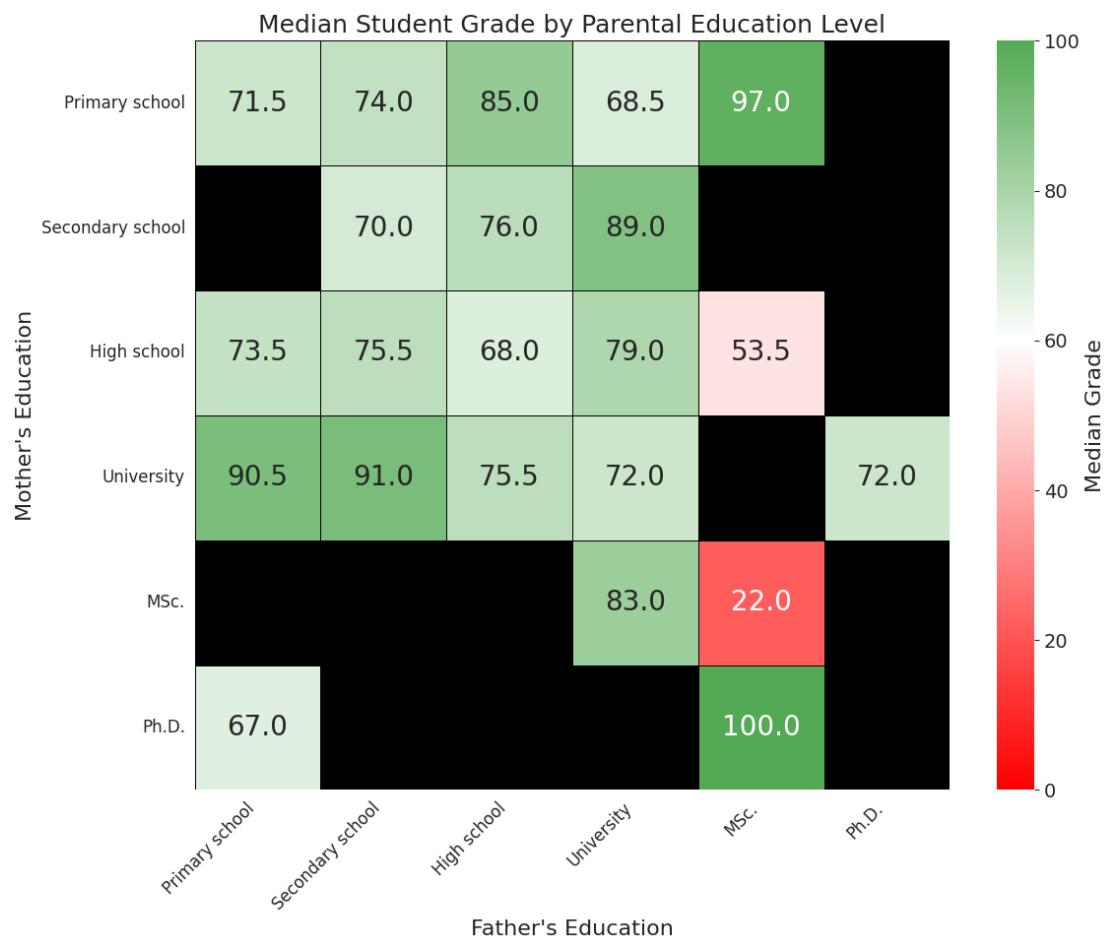
אחד הדברים המעניינים הוא הקשר בין השכלה הסטודנטית לבין השכלה ההורם. בדקנו האם יש קשר בין ההישגים בלימודים לבין השכלה ההורם. בנתונים יש את פירוט השכלה ההורם מבית ספר יסודי ועד לתואר דוקטור. לפי תרשים 6 לא ניתן להגיד שהקשר בין השכלה הסטודנטית לבין ההישגים בלימודים ישן.

תרשים 6 מציג את הציון החזיני של סטודנטים לפי רמת ההשכלה של האם והאב. מהתרשימים עולה שאין דפוס עקבי שבו עליה ברמת ההשכלה של אחד ההורם או שניהם מובילה בהכרח לציונים גבוהים יותר אצל הסטודנטים. ניתן לראות דוגמאות לציונים גבוהים גם כאשר ההורם השכלה נמוכה, ולהפך – יש מקרים שבהם להורם השכלה גבוהה במיוחד, אך הציוניים החזוניים של ילדיהם דואק נמוכים יחסית. בנוסף, ישנים שילובים מסוימים של השכלה הורם שבהם הציוניים החזוניים גבוהים במיוחד, אך אין בכך כדי להצביע על מגמה ברורה.

תרשים 7, המציג את התפלגות הציונים לפי קבוצות השכלה של ההורם (שני הורים בעלי השכלה גבוהה, שני הורים בעלי השכלה נמוכה, או שילובים ביניהם), מחזק את הממצא זהה. הਪערים בין הקבוצות אינם משמעותיים, ופיזור הציונים רחב בכל אחת מהקבוצות. לעומת זאת, גם כאשר שני ההורם משכילים, יש שונות רבות בציוני הסטודנטים, ואין עדות לכך שלהשכלה ההורם יש השפעה ברורה ועקבית על הישגי הילדים.

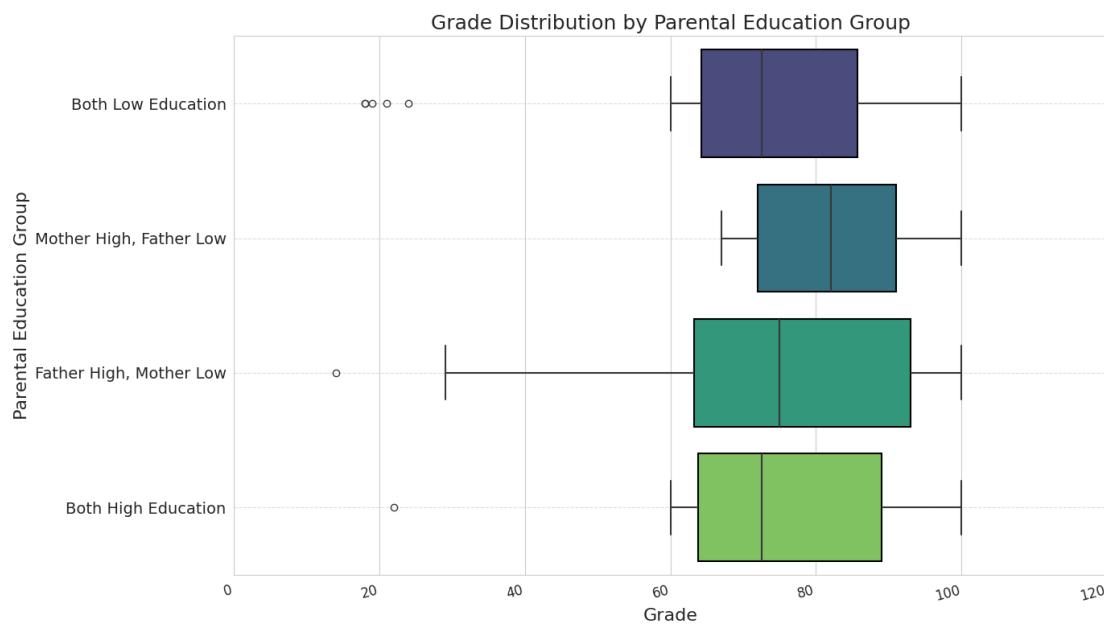
שני התרשיים ייחד מצביעים על כך שבקרוב הסטודנטים שנבדקו, לרמת השכלה ההורים קשור להישגים אקדמיים.

#### **תרשים 6: ציון חצאיוני של סטודנטים לפי רמת השכלה של האם והאב**



תרשים 6. מפת חום המציג את הציון החצאיוני של סטודנטים בהתאם לרמת השכלה של האם (שורות) והאב (עמודות). ערך הבינים (בצבע לבן) הוא ציור עובי (60), כאשר ציונים גבוהים יותר מצביעים בצבע ירוק ואילו ציונים נמוכים יותר מצביעים בצבע אדום.

#### **תרשים 7: התפלגות הציונים לפי קבוצות השכלה של ההורים**



תרשים 7. התפלגות הציוןים בקרב קבוצות ההשכלה של ההורים.

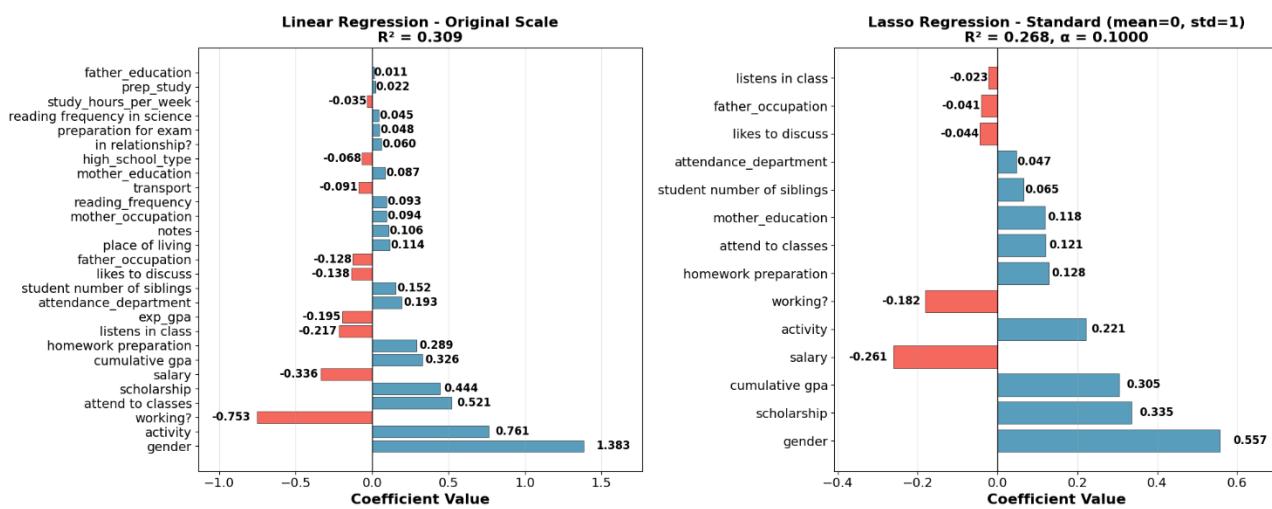
## 5. אמידת פרמטרים

בחלקים הקודמים בעבודה שאלנו האם יש מאפיינים ספציפיים מסוימים אשר יכולים להיות קשורים להצלחה בלימודים. עמדנו על ההבדלים בין מגדרים, הרגלי למידה ורקע השכלתי משפחתי. בסיס הנתונים כולל משתנים נוספים אשר יכולים להיות קשורים להצלחה בלימודים וכן בחלק זה נבצע אמידת פרמטרים בעזרת רגסיה לינארית. כאמור, הדטה כוללת 31 משתנים (מתוכם, לאחר סינון נשארו 27) ועל מנת לנסוט להימנע מ-overfitting, ביצענו גם רגסיה Lasso. ישנו כמה משתנים אשר מוגבלים לא נכללו ברגסיה Lasso לעומת רגסיה הלינארית הרגילה, למשל, מספר שעות הלמידה השבועיות (אשר ראיינו כבר היעדר קשר), השכלה האב, סוג בית הספר התיכון בו למד. כמו כן, אחוז השונות המוסבר ברגסיה הלינארית היה 0.309. ברגסיה Lasso (עם פרמטר אלפא שווה ל-0.1), נשארו 14 משתנים רלוונטיים, כאשר מחצית מתוכם מוגבלים סטטיסטיות ברמת מובהקות של  $<0.05$ . המשנה עם רמת השפעה הגדולה ביותר על הציון הסופי, וברמת המובהקות הגבוהה ביותר הוא משתנה המגדר. כמובן שאין לחשב שגברים מצלחים יותר בלימודים מאשר נשים, אלא זהה תוצאה של המדגם הספציפי זהה.

שתי תוצאות מעניינות עלות מהמשתנים: 1) האם הסטודנטית עובדת?; גובה המלגה; תוצאות האמידה מצביעות על השפעה שלילית על הציון הסופי של עבודה במהלך הלימודים, ובמקביל על השפעה חיובית של גובה המלגה על הציון הסופי. ישנה מחלוקת לגבי התרומה של עבודה במהלך הלימודים לסטודנטים. מחד, סטודנטים שעובדים משקיעים פחות זמן בלימודים וכך הציון שלהם יכול לרדת, ומайдך, עבודה עשויה לתרום הרבה לאספקטים אחרים שלא קשורים בהכרח לציון אבל כן משפיעים על הסטודנטים. בדוגמא הזאת מזמין

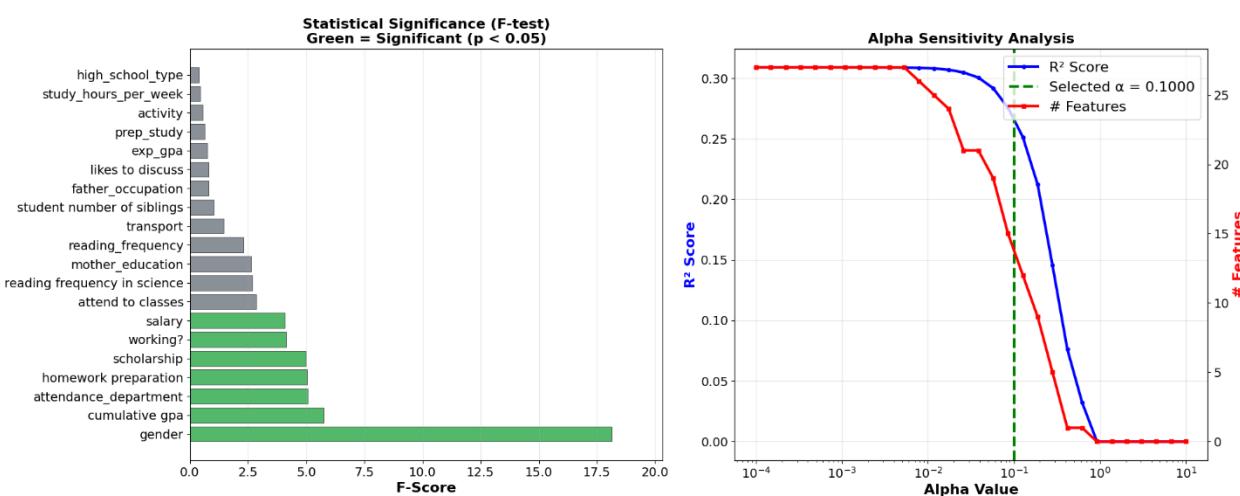
שלubitoda b'mahal hilemودim yesh shafua shelilit u'zion hoso'i b'korot. Baofen la' maftei'ut, gam meshituna ha'sher (muto'am baofen chivo'i um ubudah b'mahal hilemودim) muto'am shelilit um zion hoso'i um shafu'u ulio le'reua. Achuz ha'shuvot ha'mosberat b'regretiyt Lasso nemo' yoter u'vemd ul hoso'i um shafu'u ulio le'reua. Choshov lo'mer, she ha'mashutnim shiz'ao mobekim stutiyetit (b'remat mobekot mikoblat shel 0.268), hem ha'manavim ha'tobim bi'otar le'zion hoso'i rak ul pi ha'ntotim ha'alo, v'ihiya ksha le'hcalil 5%), hem ha'manavim ha'tobim bi'otar le'zion hoso'i rak ul pi ha'ntotim ha'alo, v'ihiya ksha le'hcalil v'lo'ha'gad she'alo hem ha'manavim ha'tobim bi'otar baofen kalli'.

#### תרשים 8: תוצאות רגression השונות (لينארית ו-lasso)



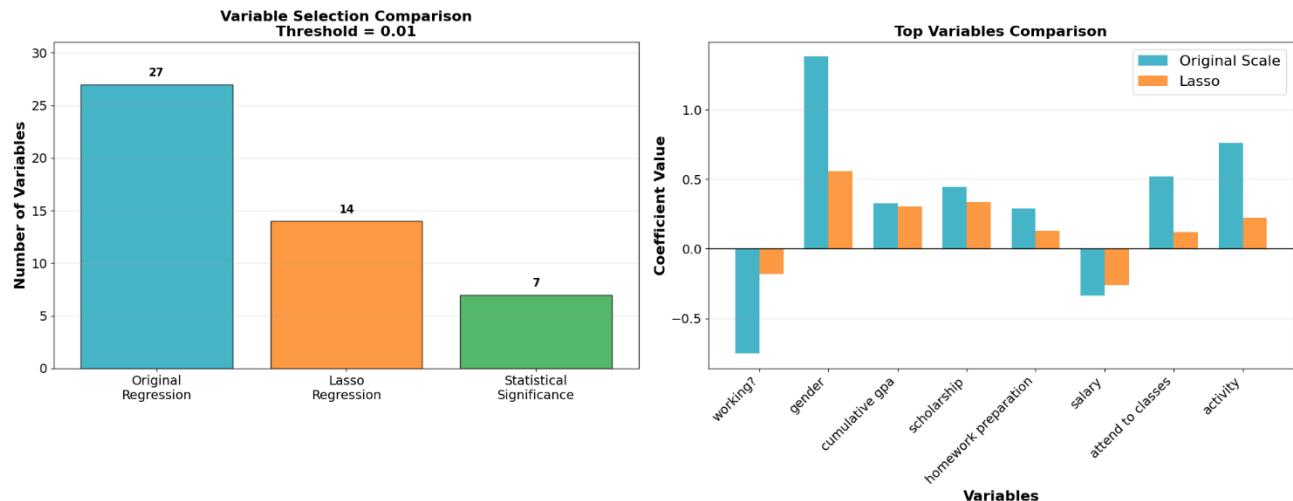
תרשים 8. התרשימים מציגים אמידה של הציון הסופי אל מול מגוון משתנים מסבירים. התרשימים השמאלי מציג רגession לינארית ואילו התרשימים הימני מציג רגession lasso אשר מסמן הממשתנים המסבירים בה ירד ל-14%.

#### תרשים 9: מובהקותות התוצאות ורגישות רגession lasso לגודל הפרמטר אלףא



תרשים 9. התרשימים מציגים את מובהקותות הפרמטרים לפי מבחן F (להתאמת פרמטרים) ואת רגישות כמוות המשתנים שנכללים ברגession lasso ואת טיב ההתאמה לפי גודל הפרמטר אלףא.

## תרשים 10: השוואת בין רגסיה לינארית לרגסיה לasso



תרשים 10. ברגסיה הלינארית כלנו 27 משתנים מסבירים, אך ברגסיה *lasso* כמהות המשתנים המסבירים ירדה ל-14 כאשר מותכם 7 מובהקים סטטיסטיות ברמת מובהקות של 5%. בתרשים הימני ניתן לראות את השוואת ערכי המקדים (המנורמלים) בין הרגסיה הלינארית לרגסיה *lasso*.

## 6. סיכום

בפרויקט זה ניסינו לבדוק אילו גורמים משפיעים על הצלחה בלימודים, אשר נמדדה, בסיסו הנוטוניים הזה, על ידי ציון סופי ב מבחן. מצאנו תוצאות די מפתיעות, ולדעתי, לא מייצגות את האוכלוסייה הכללית. ככלומר התוצאות שמצאנו מייצגות לדעתי רק את המדגם אותו בדקנו. מצאנו את כל המשתנים שמנבאים בצורה הטובה ביותר את הציון הסופי בקורס כהיגויים ביותר, אך עם זאת, מצאנו גם شيء נאם גם משתנים שנמצאו שלא משפיעים על הציון הסופי אך זה פחות עוזה היגיון עם המציאות, ולכן לדעתי התוצאות שמצאנו מייצגות רק את המדגם הזה.