**עבודה מסכמת - היכרות עם R**

**מאיה פוחס**

**שלב א' - הגדרת שאלת המחקר - 25 נק'**

1. **בחרו דאטאסט עם נושא שמסקרן אתכם.** למרות שבסופו של דבר אלו רק נתונים בטבלת אקסל, ברגע שהנושא יעניין אתכם העבודה תהיה מהנה יותר. כתבו כמה משפטים שמתארים את הדאטאסט ומדוע בחרתם דווקא בו (5 נק').

הדאטאסט שייך למחקר שלט החלומות, מחקר שאני מנהלת ומלווה כבר קרוב לשנה כחלק מעבודתי כמנהלת מעבדה אצל ד״ר מיכן קהן, שכולל התערבות קוגניטיבית-קלינית שמטרתה לסייע לילדים להתמודד עם סיוטי לילה. המחקר הוא מחקר אורך בו המשתתפים מוקצים באופן רנדומלי להתערבות (קבלת השלט והדרכות) או ביקורת, ויש 3 נקודות מדידה בהפרשים של שבועיים אחד מהשני- בייסליין, אמצע וסוף. לצורך העבודה אני אשתמש בשתי נקודות זמן, שאלון הבייסליין והשאלון האחרון.

בחרתי בו כי זה מחקר שקרוב לליבי ומעניין אותי במיוחד, ובנוסף רציתי לראות האם יש מגמות שניתן לזהות למרות שעוד לא סיימנו לגייס ולאסוף את הנתונים (הנתונים הללו לא עובדו אף פעם והמסקנות שייצאו מפה ישמשו אותנו גם בניתוח האמיתי במסגרת המחקר).

1. **בצעו הצגת נתונים מקדימה.** הצגת הנתונים האקספלורטיבית תאפשר לכם "לחוש" את הדאטה ולהעלות השערות על קשרים אפשריים בין הנתונים. עליכם להשתמש בחבילת ggdist או ggplot כדי להציג את המשתנים. נסו להשתמש בכמה שיותר פונקציות מתוך החבילה. (10 נק')
2. **בחרו שאלה מורכבת כך שהקשר אם וכאשר תמצאו כזה יהיה משמעותי ומלמד.** נסו לחשוב על זה כעל תרגיל לקראת העולם האמיתי. אם השאלה זניחה אז אולי לא שווה לשאול אותה מלכתחילה. נסו לחשוב על קשרים בעלי משמעות. נסחו את שאלת המחקר שלכם באנגלית. (10 נק' עבור מורכבות השאלה והניסוח הבהיר).

**Main question:**

Do psychological distress, behavioral difficulties, sleep-related problems, and participation in the intervention predict changes in the number of nights with nightmares and the nightmare intensity?

**Additional question:**

Among participants in the intervention group, how do parental and child compliance predict improvement (aka reduction) in the number of nights with nightmares?

**שלב ב' - עיבוד מקדים של הנתונים - 25 נק'**

1. **נסו להבין אילו משתנים מנבאים/מנובאים אתם צריכים עבור הניתוח.** לצורך הניתוח תרצו להשתמש לא בנתונים הגולמיים שקיבלתם, אלא באיזה שהוא עיבוד שלהם. קחו בחשבון שתצטרכו לבצע ניתוח של רגרסיה לינארית ושל רגרסיה לוגיסטית. (הגדירו את המשתנים בצורה ברורה - 5 נק')

הגדרות אופרציונליות:

משתנים מנבאים:

**Both groups:**

Sum DASS – integer summary of all DASS questions scores.

Sum SDQ – integer summary of all SDQ questions scores.

Sum CSHQ – integer summary of all CSHQ questions scores.

**Intervention group only:**

Sum parent – integer summary of all parent compliance questions scores.

Sum child - integer summary of all child compliance questions scores.

משתנים מנובאים:

Change in night count – change observed in night count.   
Calculated is the following way: Baseline night count score minus the final night count

Change in intensity score – change observed in nightmare intensity score.   
Calculated is the following way: Baseline intensity score minus the final intensity score

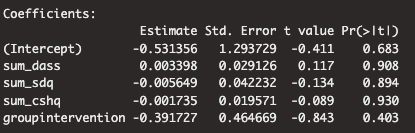
1. **השתמשו בכלים שלמדנו בחלק הראשון של הקורס על מנת לעבד את המשתנים.** נסו להשתמש בכמה שיותר פונקציות שלמדנו (עד גבול הטעם הטוב) מתוך חבילת dplyr. (עבור העיבוד 15 נק')
2. **השתמשו בפונקציה שיצרתם בעצמכם.** צרו פונקציה משלכם והפעילו אותה על הדאטה שלכם. (5 נק' עבור מימוש הפונקציה)
3. **בונוס: השתמשו בפונקציה מחבילה שלא למדנו במהלך הקורס.** חפשו בגוגל (או בChatGPT) דרך יעילה לממש פעולה מסוימת. (5 נק' בונוס)

**שלב ג' - ניתוח הנתונים - 50 נק'**

**כעת, כשהנתונים מוכנים, החליטו מהם הניתוחים הסטטיסטיים שאתם צריכים לבצע על מנת לענות על שאלת המחקר שלכם.**

1. בצעו **שני** ניתוח רגרסיה - רגרסיה לינארית מרובה ורגרסיה לוגיסטית. שימו לב להתאים את הניתוח להשערה ולנתונים. (15 נק' עבור השימוש בניתוח מתאים)
2. עבור כל אחד מהניתוחים, **פרשו את התוצאות שקיבלתם**. התייחסו למקדמים שהתקבלו ולפרשנות שלהם. (15 נק')

**Linear regression Summary:**

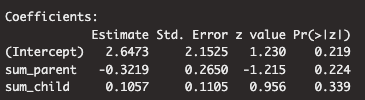


פרשנות:

הניתוח של הרגרסיה הליניארית לא מצא אפקטים מובהקים עבור אף אחד מהמשתנים. אף אחד מהמשתנים המנבאים (DASS, SDQ, CSHQ) לא ניבא באופן משמעותי את השינוי במספר לילות הסיוטים. הקבוצה (התערבות מול ביקורת) הראתה ירידה מעט גדולה יותר במספר הלילות עם סיוטים, אך גם השפעה זו לא הייתה מובהקת סטטיסטית (b = -0.392, p = 0.403).

כלומר, לא מצאנו עדות לכך שהמשתנים שנבדקו משפיעים באופן משמעותי על השינוי במספר לילות הסיוטים.

**Logistic regression Summary:**



פרשנות:

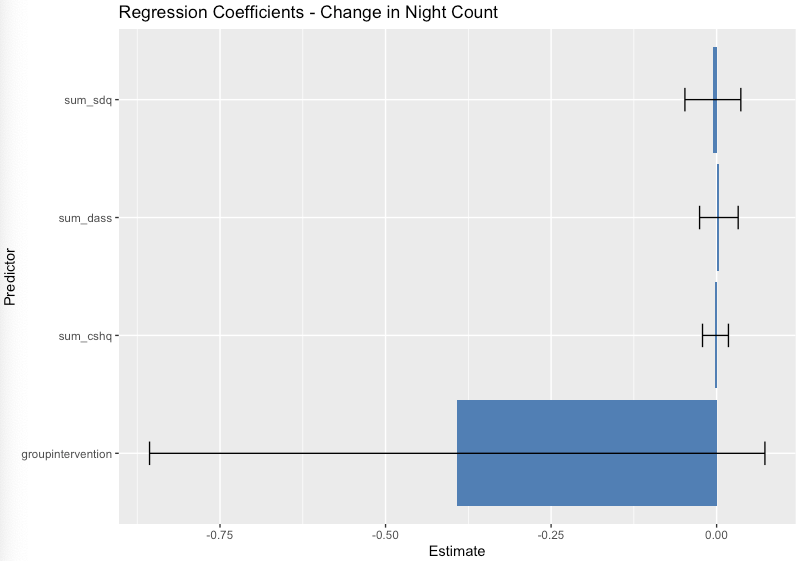
בניתוח הרגרסיה הלוגיסטית נמצא כי אף אחד מהמשתנים (סך ציוני הציות של ההורה וסך ציוני הציות של הילד) לא ניבא באופן מובהק את ההצלחה בהתערבות ( p < 0.05 עבור שני המשתנים המנבאים).

המשתנה**sum\_parent**  הראה קשר שלילי עם הצלחה בירידת בלילות בהם הילד חווה סיוטים, כלומר ככל שהציות של ההורה היה גבוה יותר, כך הסיכוי להצלחה היה מעט גבוה יותר, אך האפקט לא היה מובהק סטטיסטית (b = -0.322 , p = 0.224).

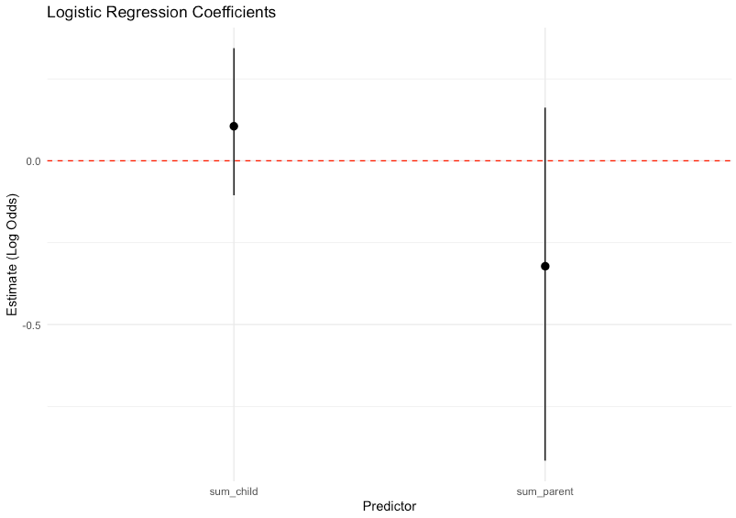
מנגד **sum\_child** הראה קשר חיובי חלש ולא מובהק עם הצלחה בירידת בלילות בהם הילד חווה סיוטים (b = -0.106, p = 0.339).

1. הוסיפו גרף אשר מתאר כל אחד מן האפקטים שקיבלתם. (10 נק')

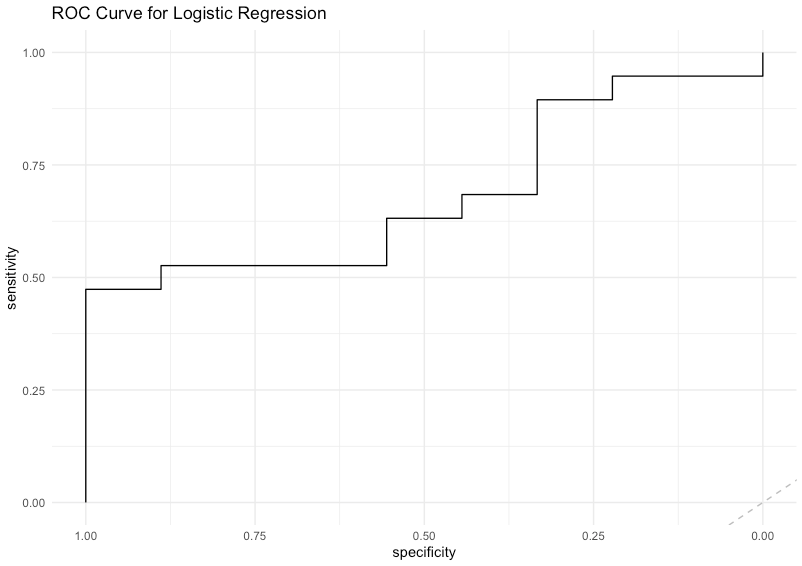
**Linear regression plot:**



**Logistic regression Summary plot:**

****

1. הוסיפו גרף ROC עבור ניתוח הרגרסיה לוגיסטית. (10 נק')



AUC: 0.6842105