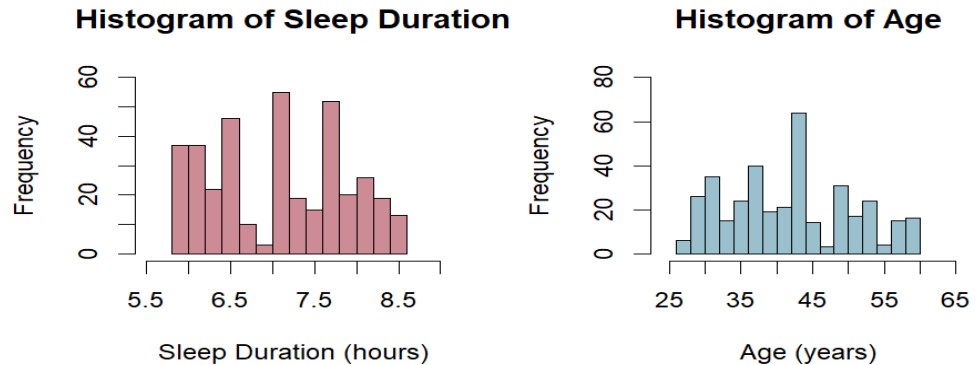


סטטיסטיקה תיאורית: הקשר בין גיל למשך השינה

בעבודה זו, נחקר כיצד גילו של אדם משפיע על מס' שעות השינה שישן ביממה. על מנת לבדוק זאת, ננתח את הקשר בין המשתנים בהתבסס על הנתונים המובאים בקובץ הלקוח מתוך האתר [Kaggle](#).

לצורך ניתוח הנתונים, הגדרנו את גיל האדם הנבדק (בשנים) להיות המשתנה המסביר (X) ואת משך השינה (בשעות) כמשתנה המוסבר (Y). השערתנו טרם ניתוח הנתונים, היא שככל שהגיל עולה, הגוף זקוק ליותר מנוחה ולכן מס' שעות השינה עולה בהתאם. כלומר, קיים מתאם חיובי בין המשתנים.

תחילה, נציג באמצעות היסטוגרמה את הפיזור עבור כל אחד מהמשתנים המצוינים לעיל.

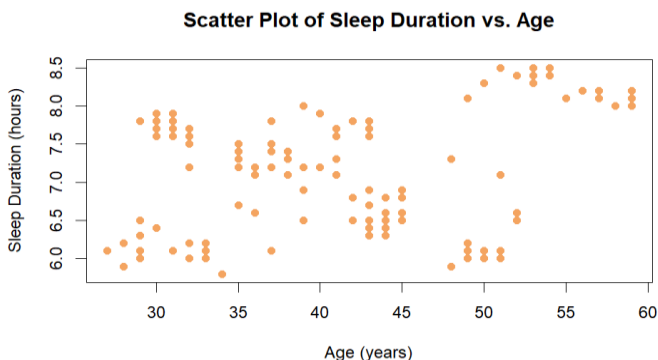


טווח הגילאים הנבדק נע בין 27-59, כאשר הייצוג המירבי במדגם הינו של גילאי 35-45. באשר להתפלגות משך השינה, זו נעה בין 5.8 ל- 8.5 שעות שינה, כאשר ניתן להבחין כי משך השינה השכיח ביותר הינו כ- 7.2 שעות. לצד זאת, ניתן לראות כי התפלגות משך השינה הינה רב-שכיחית, כלומר אין מס' שעות שינה יחיד שמאפיין את רוב הנבדקים.

כעת, נציג את מדדי המיקום והפיזור של המשתנים בטבלה שלהלן.

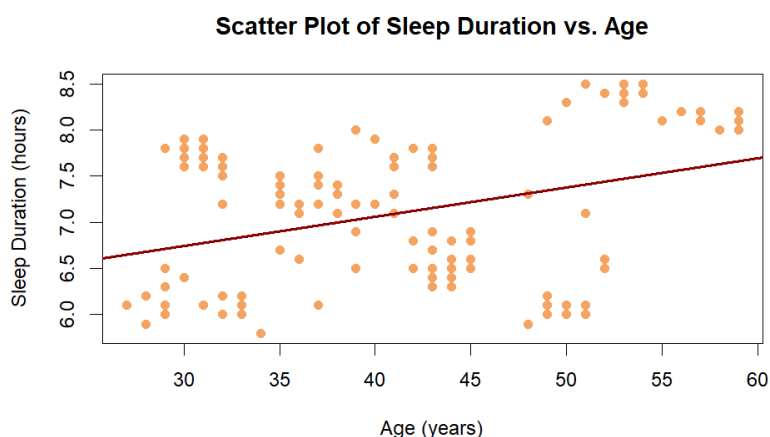
ממוצע	חציון	סטיית תקן	מינימום	מקסימום	
42.18	43	8.673	27	59	גיל (שנים)
7.132	7.2	0.796	5.8	8.5	משך השינה (שעות)

נשים לב כי הממוצע והחציון של משך השינה קרובים זה לזה. הדבר מעיד על התפלגות סימטרית, בה אין הטיה משמעותית כלפי ערכים נמוכים או גבוהים של שעות שינה בקרב הנבדקים. כמו כן, סטיית התקן יחסית גבוהה, מה שמצביע על כך שהגילאים במדגם מפוזרים בצורה יחסית רחבה סביב הממוצע. לאחר שהתבוננו במשתנים, נרצה לבחון את אופי הקשר ביניהם באמצעות דיאגרמת פיזור דו-משתנית.



מהסתכלות ראשונית, ניתן לראות כי התצפיות עבור רוב טווחי גילאים נפרסות על מס' שעות שינה מגוון. כלומר, לא קיים קשר משמעותי בין הנתונים, אם כי כן ניתן להבחין בקשר מסוים, בעיקר בגילאי 55 ומעלה. נמשיך ונבחן את טיב הקשר באמצעות רגרסיה ליניארית ומתאם פירסון.

ראשית, נציג את ישר הרגרסיה בין המשתנים: $y = 0.03162x + 5.79809$



כמו כן, מקדם המתאם הינו: $r(x, y) = 0.3447094$

מהנתונים עולה כי בין המשתנים קיים קשר לינארי חיובי בינוני חלש. הדבר מעיד על כך שהשונות המוסברת של משך השינה ע"י הגיל היא יחסית קטנה. כלומר, גילו של אדם אינו משפיע באופן משמעותי על כמות שעות השינה שלו. עם זאת, נוכל להסיק (משיפוע הרגרסיה) כי ככל שהגיל עולה, משך השינה החזוי עולה בממוצע בכ-0.032 שעות (כ-2 דקות), מה שמעיד על עלייה מתונה של משך השינה עם הגיל.

מבדיקת התצפיות, לא איתרנו תצפיות חריגות (מעל 2 סטיות תקן). נוכל להסביר זאת ע"י כך שישנה שונות רבה בנתונים בין התצפיות השונות בכל קטגוריית גיל ואין איזושהי מגמה שניתן לחרוג ממנה.

כעת, נבחר תצפיות ונציג עבורן את הערך החזוי עפ"י קו הרגרסיה, הערך האמיתי והשארית ביניהן.

השארית ¹	ממוצע הערכים האמיתיים	הערכים האמיתיים	הערך החזוי	אופן החישוב	
-0.664	6.02	5.9, 6.2	6.68	$y = 0.03162 \cdot 28 + 5.79809 = 6.68345$	גיל 28
0.312	7.22	6.7, 7.2-7.5	6.9	$y = 0.03162 \cdot 35 + 5.79809 = 6.90479$	גיל 35
0.307	7.43	6.5, 6.8, 7.8	7.13	$y = 0.03162 \cdot 42 + 5.79809 = 7.12613$	גיל 42
0.002	7.41	6, 6.1, 7.1, 8.5	7.41	$y = 0.03162 \cdot 51 + 5.79809 = 7.41071$	גיל 51
0.368	8	8	7.63	$y = 0.03162 \cdot 58 + 5.79809 = 7.63205$	גיל 58

ניתן לראות כי הערך של השאריות נע בטווח קטן יחסית, מה שמעיד שהמודל מצליח לתאר את הנתונים בצורה די טובה.

לסיכום, מניתוח הנתונים עולה כי הגיל אינו משפיע באופן משמעותי על משך השינה. עם זאת, הקשר הינו חיובי וישנה עליה מסוימת במס' שעות השינה עם עלייה בגיל. כמו כן, מממצאי השאריות ניתן להסיק כי ישנם גורמים נוספים מעבר לגיל המשפיעים על משך השינה וייתכן שהוספתם תשפר את יכולת הניבוי.

¹ השארית המחושבת היא בין הערך החזוי לממוצע הערכים האמיתיים.