

השפעתן של פרסומות הסופר בול על הווידאליזת של המפרסמות

פרויקט סיום בסטטיסטיקה



תמר שמש

עמית רבינוביץ'

אילן סיטבון

יונתן גפטר



בעבודתנו, אנו חוקרים את ההשפעה של פרסומות הסופרבויל על הוויראליות של החברות המפרסמות ובוחנים האם יש קשר בין עלות הפרסומות לרמת הוויראליות שנוצרה. את המידע אודות וויראליות החברות אספנו וביססנו על סמך google trends, פלטפורמה אשר מאפשרת להציג באופן גרפי את מגמות החיפוש בגוגל, בחתך זמנים ובהשוואה לנפח החיפוש הכולל. לצורך המחקר שלנו אספנו קבצי נתונים המכילים מידע על עלות ואורכן של הפרסומות של כל החברות שפרסמו בזמן המחצית באופן תדיר בין השנים 2004-2020 והשווינו עם התוצאות שקיבלנו בgoogle trends. למעשה, בדקנו האם יש קשר לינארי בין מחיר הפרסומת אשר נגזר מאורכה לויראליות של החברות בשבוע הסופרבויל כנגד השבוע שלפניו. המודלים הסטטיסטיים שבהם השתמשנו הם קו הריבועים הפחותים, למעשה קו הרגרסיה ובמודל הקו החסין, השתמשנו גם במבחן T מזווג ולא מזווג ובמשפט הגבול המרכזי לבניית ר"ס לתוחלת. המסקנה שאליה הגענו היא שאינו מתקיים קשר לינארי בין מחירי הפרסומות ליחס הוויראליות של החברות, שכן מקדם המתאם של פירסון בכל הבדיקות היה קרוב מאוד ל-0. לעומת זאת, מצאנו קשר בין פרסום בסופרבויל לבין עלייה בנפח החיפושים לטווח הקצר, אך לא הצלחנו להראות שהעלייה בנפח החיפושים נשמרת לטווח הארוך. אל התוצאות הגענו על ידי בדיקת השערות, על פי התוצאות מgoogle trends.



פרסומות הסופרבול התעלו מעבר לאופי המסחרי שלהן והפכו לאבני בוחן תרבותיות שמושכות מיליוני צופים ברחבי העולם. הפרסומות הללו, שמוצגות במהלך משחק האליפות של ליגת הפוטבול הלאומית, הן חלק מרכזי בתרבות המשחקים האלו. לא רק שהן ידועות בערכן הבידורי ובכוכבים המפורסמים המשתתפים בהן, הן גם מעוררות שיחות, יוצרות באז ומשאירות רושם מתמשך על החברה(1). אין להכחיש את ההשפעה התרבותית של פרסומות אלו, שכן הן מגיעות לקהל של למעלה מ-100 מיליון צופים מדי שנה, מה שהופך את הסופרבול לאחד מאירועי הטלוויזיה הנצפים ביותר בעולם. הפרסומות הללו הפכו לסמל של צרכנות אמריקאית, יצירתיות ומצוינות פרסומית.

רמת העניין ותשומת הלב שאנשים מקדישים לפרסומות המחצית יוצאת דופן, והפרסומות הפכו לחלק בלתי נפרד מחווית הסופרבול. במחקר שנעשה נמצא כי 70 אחוז מהצופים מצפים לפרסומות(4) וכי כמעט אחד מכל ארבעה צופים במשחק במיוחד עבור הפרסומות(5), ורואים בהן סוג של בידור בדומה למשחק עצמו.

פיתוי החברות לפרסם פרסומות שכאלו עולה לא מעט. בהתחשב בכמות הצופים וההשפעה התרבותית העצומה, מפרסמים מוכנים להשקיע סכומי כסף מופקעים כדי להבטיח מקום בפרק הזמן הנחשק. המחיר של הפרסום עולה משנה לשנה, כאשר העלות הממוצעת ל30 שניות פרסומת במהלך הסופרבול הגיע לשבעה מיליון דולר בשנים האחרונות. למרות המחיר הגבוה, המחקרים שנעשו עד היום לרוב מצביעים על כך שהפרסומות מייצרות תשואה חיובית ומשפיעות מאוד על פופולאריות החברה המפרסמת(2),(3).

בפרויקט זה, נחקור את הערך הכספי של מודעות אלו, ונבחון את הוויראליות הנוצרת בעקבות הפרסומות הללו. על ידי בחינת עלות הפרסומות, אנו שואפים לענות על שתי שאלות מרכזיות:

1. האם קיימת התאמה בין מחירן של הפרסומות לבין הוויראליות שנוצרת בעקיבותן.
2. נכמת את הוויראליות שנוצרת בעקבות שידור הפרסומת.

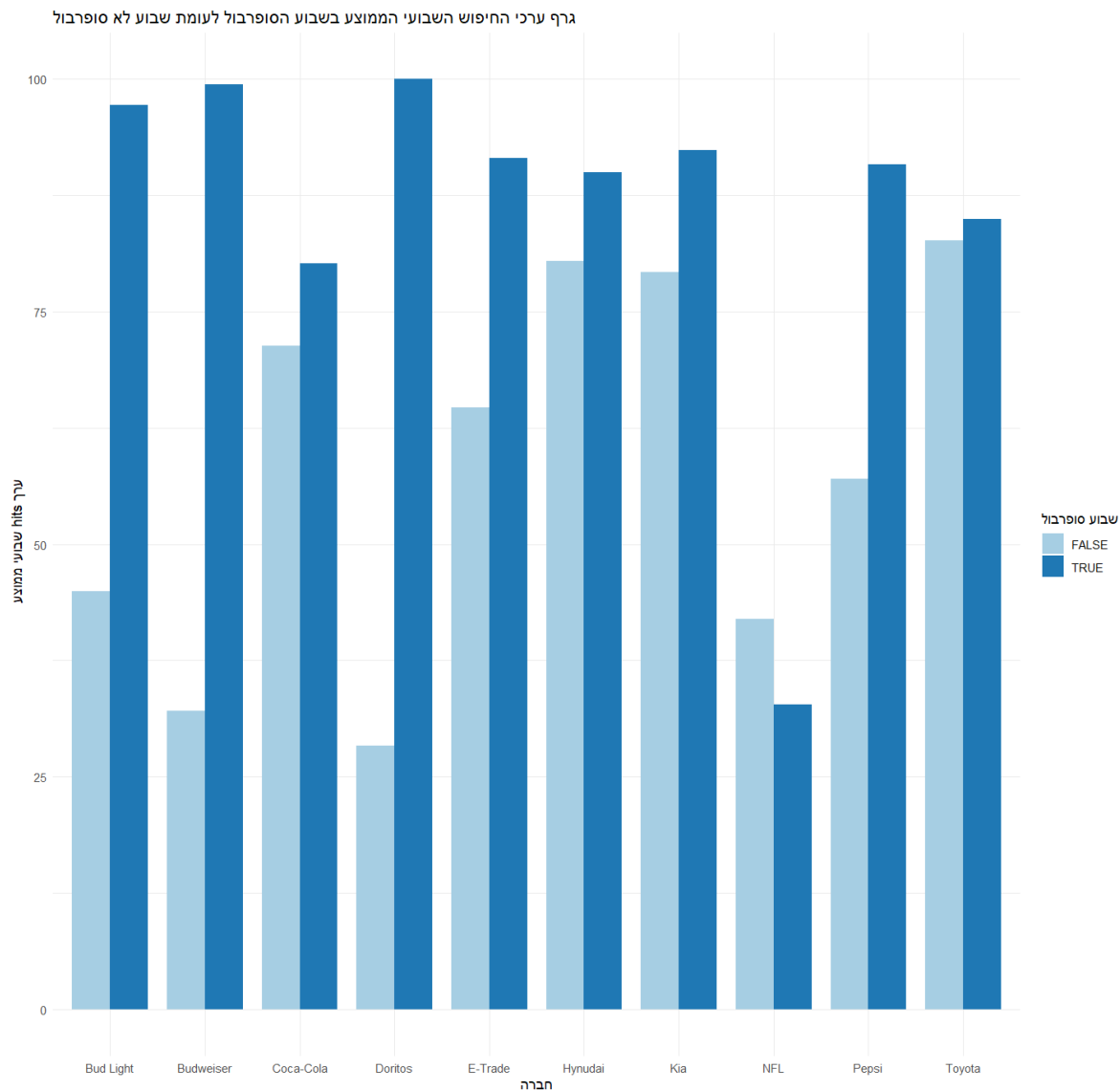


איסוף הנתונים

כלל נתוני הוויראליות בפרויקט נאספו בעזרת google trends. המידע המתקבל הנו מנורמל (7). הערכים המתקבלים נקראים hits והם בקנה מידה של 0 עד 100.

השתמשנו ב google trends בשביל להפיק ולאגד את המידע הדרוש לפי הנושאים הרלוונטיים לנו (8). הסתכלנו על כל השבועות בשנה עבור כל החברות שפרסמו בין השנים 2004-2020. כאשר כל שנה מנורמלת בפני עצמה, כלומר ללא תלות בשנים האחרות¹

את המידע על החברות ושנות הפרסום הפקנו ממסד נתונים (6) אלו החברות שניתחנו את המידע עבורן:



¹ כל השבועות נבדקו ביחס לארצות הברית

בנוסף, את זמני הפרסומת הפקנו מסרטוני היוטיוב שקישוריהם מופיעים
באותו מסד נתונים (6)

ההשערות הנבדקות

לאורך כל הפרויקט נעבוד עם רמת מובהקות של חמישה אחוזים.
כמו כן, אנו מניחים שבועות משנים שונות לא תלויים אחד בשני
כי הם מנורמלים לחוד, ובנוסף אנו מניחים שבועות של חברות שונות בלתי
תלויים.

אנו נבחן שתי שאלות עיקריות:

1. האם קיימת התאמה לינארית בין מחירן של הפרסומות לבין
הוויראליות שנוצרת בעקבותן. הוויראליות תבחן כיחס הוויראליות בין
השבוע של הסופרבול לבין השבוע שלפניו. הוויראליות היא
המשתנה המוסבר של המודל. את הוויראליות הגדרנו להיות היחס בין
משתנה hits בשבוע של הסופרבול לבין המשתנה בשבוע שלפניו. זאת
מכיוון שכך אנו יכולים להתעלם מהשפעת הנרמול של הנתונים, שכן
אם $x_1 =$ מספר החיפושים בשבוע הסופרבול, $x_2 =$ מספר החיפושים
בשבוע שלפני הסופרבול, $y =$ קבוע הנרמול של גוגל עבור נתונים
מאותה שנה אז מתקיים $\frac{x_1}{x_2} = \frac{y}{x_2}$. כך אנחנו מסוגלים למדוד באופן ישיר
את השינוי במספר החיפושים של החברה לפני ואחרי שידור
הפרסומות.

בנוסף, מכיוון שלחלק מן החברות יש כמה פרסומות שונות באותה
שנה, המשתנה המסביר של המודל הוא מחיר הפרסומות שמחושב
כסכום עלות השידור של הפרסומות באותה שנה, כאשר מחיר כל
פרסומת חושב על ידי הכפלת זמן הפרסומת במחיר הממוצע בדולרים
לחצי דקת פרסום באותה שנה.

לשם חקירת שאלת המחקר הגדרנו את ההשערות הבאות:
H0: לא קיימת התאמה לינארית בין מחיר פרסומת לבין הוויראליות
שנגרמת בעקבותיה.
H1: קיימת התאמה לינארית בין מחיר פרסומת לבין הוויראליות
שנגרמת בעקבותיה.
עבור שאלה זו השתמשנו במודל הריבועים הפחותים ובקו החסין.

נסמן את מקדם המתאם של פירסון ב- r
נסמן את מקדם השיפוע של קו הריבועים הפחותים ב- b_{LS}
נסמן את מקדם השיפוע של הקו החסין ב- b_{RL}

2. הערכת רמת הוויראליות

בהערכת רמת הוויראליות נתייחס לארבעה גורמים מרכזיים:

1. נבדוק האם קיים קשר בין שבוע השיא של חברה לבין שבוע הסופרבוול. נבדוק זאת באמצעות מבחן USP.

לשם כך נשתמש בהשערות הבאות:

H0: לא קיים קשר בין שבוע השיא של חברה לבין שבוע הסופרבוול.

H1: קיים קשר בין שבוע השיא של חברה לבין שבוע הסופרבוול.

2. אמידת r - התצפיות שלנו עבור חברות פרסמו הנו: X_1, \dots, X_{114}
כאשר x_i - ערך hits בשבוע (של חברה שפרסמה). נשים לב

שכל $X_i, X_j : i \neq j$

בלתי תלויים מההנחות שלנו כי אם הם שייכים לאותו סופרבוול אז הם שייכים לחברות שונות ואם הם שייכים לאותה חברה אז הם משנים שונות. השתמשנו במשפט הגבול המרכזי² בשביל להגיד כי \bar{X} בקירוב μ , בסטיית תקן המדגמית והתפלגות t (כי השונות לא ידועה). על מנת לבצע אמידה לתוחלת בשבוע הסופרבוול ובשבועות האחרים, דגמנו 100 תצפיות והנחנו נורמליות³, ביצענו את אותו תהליך על מנת לבנות להם רווח סמך.

3. נחשב הפרש תוחלות לhits בין שבועות בהם הייתה פרסומת לבין שבועות שבהם לא הייתה. דגמנו 300 שבועות בהם לא הייתה פרסומת והנחנו כי ההפרש מתפלג נורמלי⁴ לכן השתמשנו במבחן welch.

לשם כך נשתמש בהשערות הבאות:

H0: הפרש התוחלות קטן או שווה ל25.

H1: הפרש התוחלות גדול מ25.

4. נבחן באמצעות מבחן t מזווג⁵ האם ניתן לקבוע כי בשנה שבה חברה מפרסמת בסופרבוול נפח החיפוש השנתי של החברה יגדל.

לשם כך נשתמש בהשערות הבאות:

² תרשים 1 ברפרנסים

³ תרשים 2 ברפרנסים

⁴ תרשים 3 ברפרנסים

⁵ תרשים 4 ברפרנסים

H0: לא ניתן לקבוע כי בשנה שבה חברה מפרסמת בסופרבול נפח החיפושים השנתי יגדל.

H1: ניתן לקבוע כי בשנה שבה חברה מפרסמת בסופרבול נפח החיפושים השנתי יגדל.

חבילות R שהיו בשימוש:

- tidyverse ○
- Usp ○
- GtrendsR ○
- graphics ○

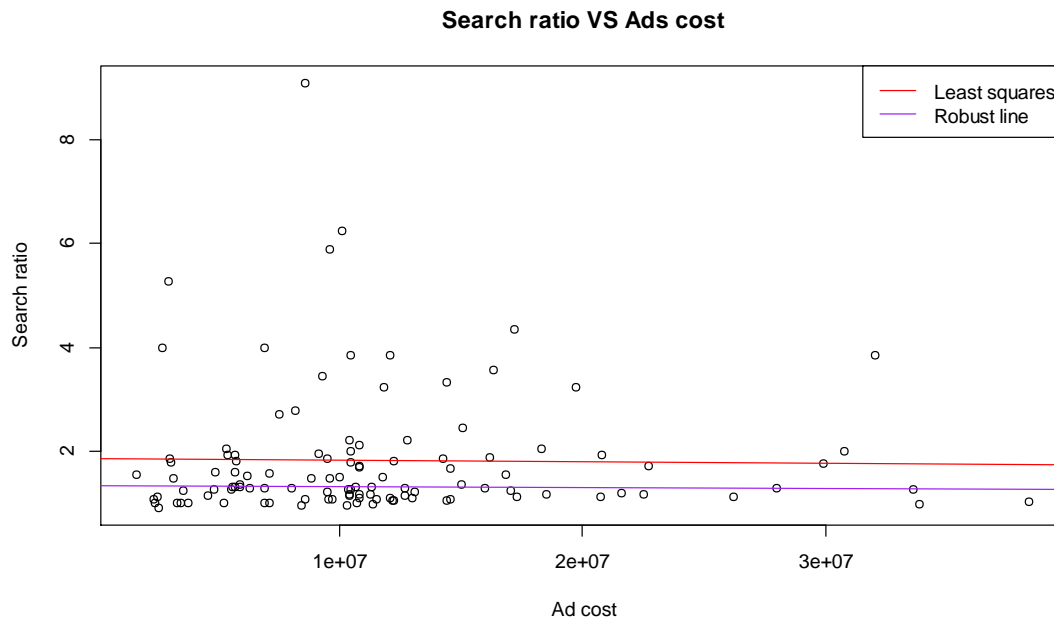


תוצאות

1. האם קיימת התאמה לינארית בין מחירן של הפרסומות לבין הוויראליות שנוצרת בעקיבותן.

לאחר איסוף הנתונים בנינו מודל רגרסיה לינארית ליחס החיפוש של החברה לפני ואחרי הסופרבול כפונקציה של מחיר הפרסומות של החברה עבור אותה שנה.

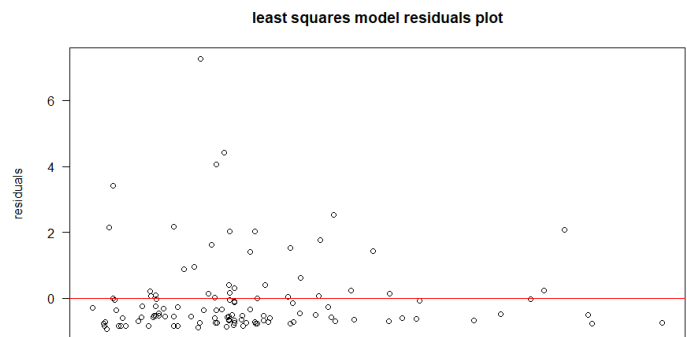
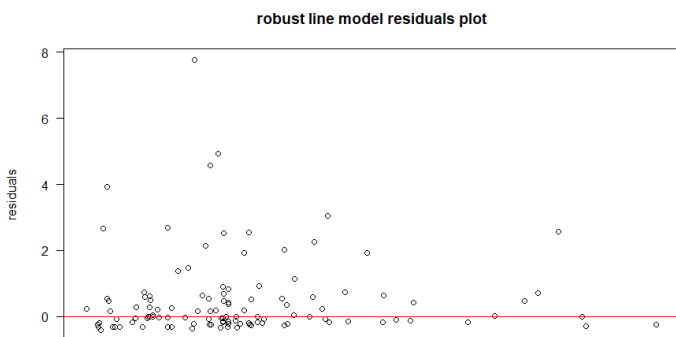
להלן גרף המודל:



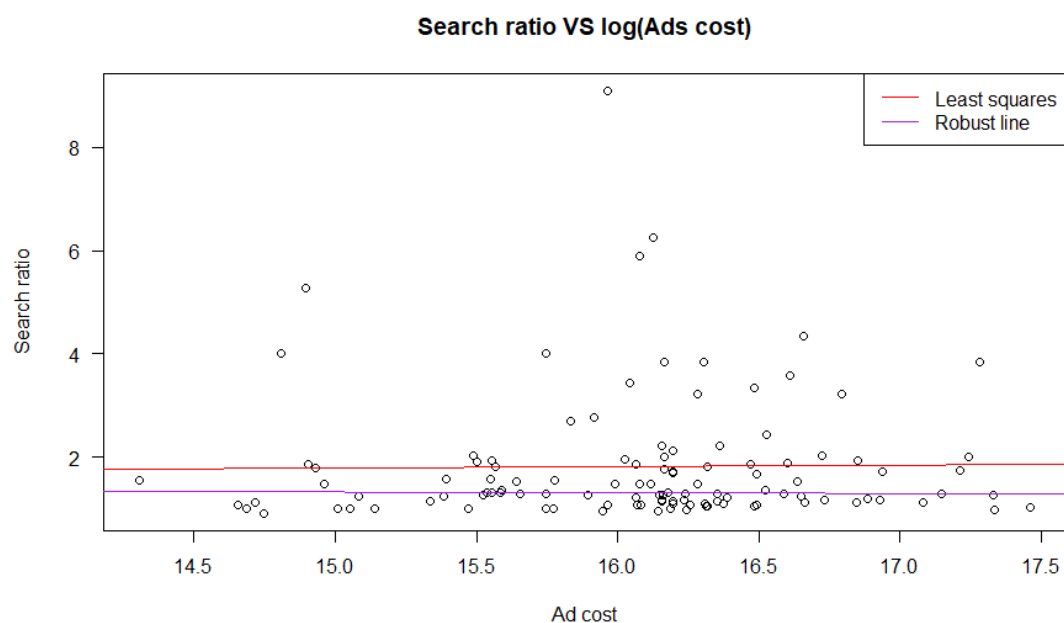
לאחר שראינו כי יש מספר של תצפיות חריגות החלטנו להשוות את מודל הריבועים הפחותים אל מול הקו החסין. כאשר לאחר חישוב קיבלנו $b_{LS} = -2.4754 \times 10^{-9}$, $b_{RL} = -1.34735 \times 10^{-9}$

ניתן לראות עבור קו הריבועים הפחותים ועבור הקו החסין ששיפועי הגרף קרובים מאוד ל-0, ניתן להסביר זאת על ידי מקדם המתאם של פירסון שערכו $r = -0.01472$, מכאן הסקנו שאין קשר לינארי בין המשתנים.

כדי לבדוק את טיב ההתאמה של המודלים לנתונים הסתכלנו בתרשימי השאריות של המודלים, ניתן לראות כי אין פיזור אחיד סביב קו ה-0 שהיינו מצפים לראות אם הנתונים היו מתאימים למודל הלינארי. ניתן להבחין כי גרף השאריות של כל מודל דומה מאוד בצורתו לגרף המקורי, זאת מכיוון ששיפועי הקווים קרובים מאוד ל-0.



בגרף ניתן לראות טרנספורמציית לוגריתם על מחיר הפרסומות:



עבור המודל אחרי טרנספורמציית לוגריתם התקבל: $b_{LS} = r = 0.016896$
 $b_{RL} = -0.013554$ לכן הסקנו כי גם לאחר הטרנספורמציה לא קיים קשר לינארי בין המשתנים.

2. הערכת הוויראליות

מבחן 1: בדיקת קשר

ראשית על מנת לקבל הבנה בסיסית על הנתונים בדקנו האם יש קשר בין שבוע השיא של חברה, כלומר השבוע בשנה עם הכי הרבה חיפוש, לבין שבוע הסופרבול. שכן מטרת המחקר היא לבדוק את הקשר בין פרסום בסופרבול לאחוזי החיפוש באותה שנה.
האם שבוע סופרבול X -

האם שבוע שיא Y -

השערת האפס: אין תלות בין המשתנים.

השערת המחקר: יש תלות בין המשתנים.

השתמנו במבחן USP שראינו בכיתה.

התוצאה: $P - Value = 1 \times 10^{-4} < 0.05$

ערך הסטטיסטי: $T = -0.000261121$

לכן נדחה את השערת האפס ברמת מובהקות 5% ונסיק כי אכן יש תלות בין שבוע השיא לבין פרסום בסופרבול

בנינו את טבלת השכיחות הבאה:

Observed		X		Total
		not Superbowl week	Superbowl week	
Y	Hits == 100	59	127	186
	Hits != 100	55	8659	8714
Total		114	8786	8900

Expected		X		Total~
		not Superbowl week	Superbowl week	
Y	Hits == 100	2.382	183.6175	186
	Hits != 100	111.617	8602.382	8714
Total~		114	8786	8900

מבחן 2: אמידת התוחלת ע"י רווח סמך

אחרי שראינו שיש קשר נבדוק אם תוחלת הhits בשבוע הסופרבול גבוהה אצל חברות שפרסמו בסופרבול. לשם כך בנינו רווח סמך לתוחלת הhits בשבועות בהם החברות פרסמו בסופרבול ברמת סמך 95%, ורווח סמך לתוחלת הhits בשבועות בהם חברות לא פרסמו ברווח סמך 95%. השתמשנו בפונקציה t.test על מנת לבנות את הרווח סמך והתוצאות: רווח נסמך שקיבלנו לתוחלת הhits בשבועות שבהם חברות פרסמו הנו

$$CI_{95\%}(\mu_1) = [84.94, 91.97]$$

הרווח סמך שקיבלנו לתוחלת הhits בשבועות שבהם החברות לא פרסמו הנו

$$CI_{95\%}(\mu_2) = [55.98, 65.43]$$

מכאן אנחנו מסיקים כי אכן יש פער משמעותי בין ערכי הhits בשבועות בהם החברות פרסמו בסופרבול לבין שבועות שבהם לא פרסמו בסופרבול.

מבחן 3: הפרש תוחלת

מהמבחן הקודם קיבלנו אינטואיציה לערך הממוצע בשבועות בהם הייתה פרסומת ובשבועות שבהם לא הייתה, כעת ננסה לעגן קביעה זו באופן יותר מדויק ע"י כלים סטטיסטיים.

נסמן:

- ערכי השבועות שבהם הייתה פרסומת סופרבול. X

- ערכי השבועות שבהם לא הייתה פרסומת סופרבול. Y

נרצה לבחון את ההשערות הבאות:

$$H_0: \mu_x - \mu_y \leq 25, H_1: \mu_x - \mu_y > 25$$

התוצאות שקיבלנו ממבחן Welch:

$$t = 2.8741, df = 118.46, p\text{-value} = 0.002402$$

לכן נסיק ברמת מובהקות 5% כי הפרש התוחלות גדול מ-25 ובפרט התוחלת בשבוע שהייתה פרסומת גדולה בלפחות 25 מתוחלת שבוע שלא הייתה פרסומת, מכאן קיבלנו אומדן מדויק יותר להפרש התוחלות ונסיק כי אכן קיים הבדל משמעותי בערכי hits בין השבועות.

מבחן 4: מבחן מזווג

כעת נתבונן בחברות השונות וניצור מדגם מזווג מהצורה (X, Y) כאשר X - ערך הhits הממוצע השנתי בשנה שבה לחברה הייתה פרסומת, Y - ערך הhits הממוצע השנתי בשנה שבה לחברה לא הייתה פרסומת.

$$H_0: \mu_x - \mu_y = 0, H_1: \mu_x - \mu_y \neq 0$$

תוצאות:

$$t = -1.8078, df = 49, p\text{-value} = 0.07678$$

מה-p-value נסיק כי לא נדחה את השערת האפס, ולכן לא נוכל לקבוע באופן חד משמעי כי פרסום בסופרבול משפיע על נפח החיפוש השנתי של חברה מסוימת.



דיון ומסקנות

מסקנות ומענה על שאלת המחקר

הפרויקט שלנו עוסק בהשפעת פרסומות הסופרבויל על וויראליות החברות המפרסמות. הנתונים שלנו התקבלו מ-google trends. בפרויקט עסקנו בשתי שאלות מרכזיות. השאלה הראשונה היא האם קיים קשר לינארי בין מחיר הפרסומת לווראליות שנגרמה בעקבותיה. עבודה מדדנו את השינוי בחיפוש בעקבות הפרסום והגענו למסקנה כי אין התאמה בין הווראליות של הפרסומות לעלות שלהן.

השאלה השנייה היא האם פרסום בסופרבויל משפיע על הווראליות של החברות. על מנת לענות על שאלה זו פירקנו אותה לכמה מבחנים שביחד מרכיבים את התמונה. הצלחנו להראות בעזרת המבחן הראשון כי יש קשר בין שבועות שיא לפרסום בסופרבויל ובעזרת המבחן השני והשלישי הצלחנו לקבוע כי אכן קיים פער בין הווראליות הממוצעת בשבועות שהייתה פרסומת סופרבויל לבין שבועות שלא הייתה. לעומת זאת, במבחן הרביעי לא הצלחנו להראות כי נפח החיפוש מושפע מפרסום בסופרבויל, כלומר פרסום בסופרבויל לא נותן השפעת ארוכת טווח על הווראליות בשנה. מכך, המסקנה הסופית שלנו היא שפרסום בסופרבויל אכן משפיע לטובה על ויראליות החברות, אם כי ההשפעה הנה עבור טווח קצר ואינה בהכרח נשמרת לטווח הארוך. נסתייג ונאמר כי ככל הנראה גם לאופי החברה יש השפעה על הווראליות. למשל, עבור קוקה-קולה ככל הנראה אין סיבה לחפש אותה באופן תדיר בגוגל שכן מדובר בחברת משקאות בניגוד ל E-TRADE שמתעסקת במסחר אינטרנטי.

התיישבות התוצאות עם התחזיות

במחקר המקדים קראנו כי פרסומות הסופרבויל משתלמות לחברות המפרסמות (3) למרות שמחירן עולה משמעותית בכל שנה. בעקבות כך האמנו כי יתגלה קשר לינארי בין מחיר הפרסומות לווראליות שהיא יוצרת, אך גילינו כי אין קשר לינארי בין השניים.

בעקבות אותה קביעה שהפרסומות משתלמות למרות מחירן הגבוה, האמנו כי נראה שהווראליות הנגרמת בעקבותן גבוהה ביותר. תוצאות המבחנים איששו את ההשערה הזאת.

מגבלות באיסוף וניתוח הנתונים

1. אין לנו שום דרך לדעת באמת כמה מדויקת המדידה של גוגל. כאשר בודקים כמות גדולה של שנים, google trends מציג את המידע בחודשים. בנוסף, כל המידע מנורמל יחד, מה שהקשה עלינו להסיק מסקנות עבור השנים השונות.

בעקבות כך התמודדנו עם בעיה זאת בכך שהסתכלנו על כל שנה בנפרד, ואז קיבלנו שכל השנים מנורמלות בנפרד. בכך המידע ניתן בשבועות ולא בחודשים, מה שהגדיל את הרזולוציה של המידע שלנו. יתראה מזאת, מחירי הפרסומות עבור הסופרבול לא לוקחים בחשבון את האינפלציה בין השנים השונות, לכן המחירים לא מהווים מדד אחיד עבור כל השנים.

שיפור המחקר וכיווני מחקר חדשים

במחקר הגענו לכך שלא קיים קשר לינארי בין מחיר הפרסומות לווראלית אותה היא יוצרת. דרך אפשרית לשיפור המחקר היא לבצע בדיקות שונות של קשרים אחרים במטרה למצוא אם קיים קשר בין מחיר הפרסומות לשינוי בווראליות שהן יוצרות, ואם הוא קיים, איזה קשר הוא. דרך נוספת לשיפור המחקר היא להתחשב באינפלציה בחישוב מחירי הפרסומות על מנת ליצור השוואה מדויקת יותר בין הכספים שהושקעו בפרסומות.

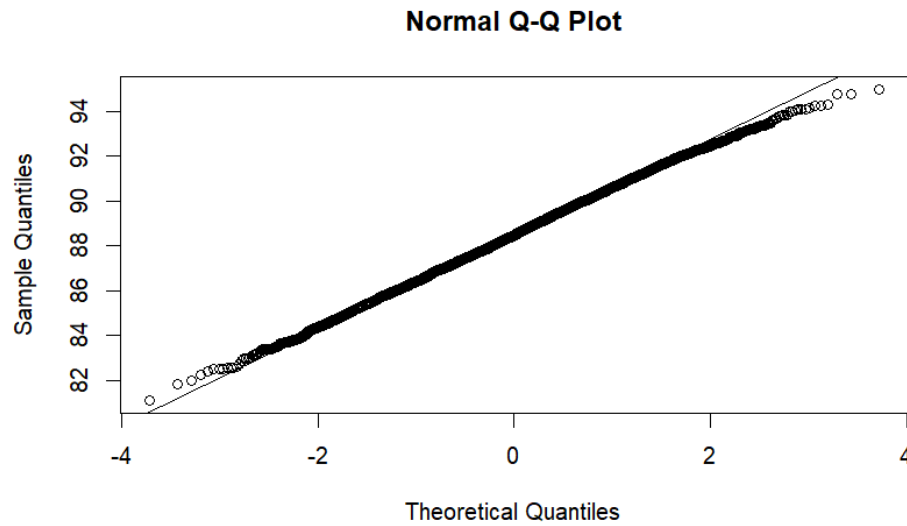
דרך להפחית את ההטיה של המחקר היא לדגום נתונים גם מרשתות חברתיות. ערך ה hits-מבוסס בעיקר על חיפושים מגוגל ויוטיוב, בעוד צעירים רבים כיום מבצעים את החיפושים שלהם ברשתות החברתיות. מכך, כיוון טוב לשיפור המחקר יהיה לקחת בחשבון גם את הרשתות הללו.

מחקר המשך שניתן לבצע בעקבות המחקר שלנו הוא על אילו שווקים (שתייה, רכבים...) הפרסומות משפיעות ברמה מקסימלית.

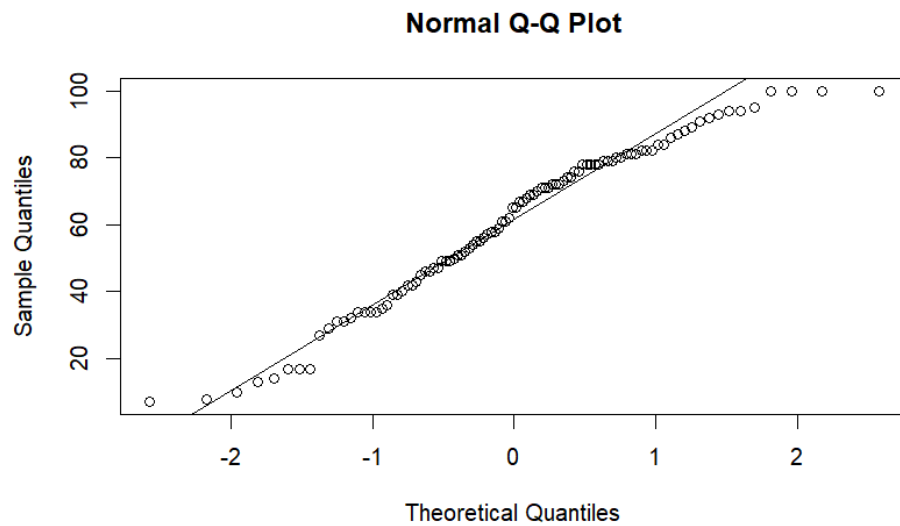


רפרנסים

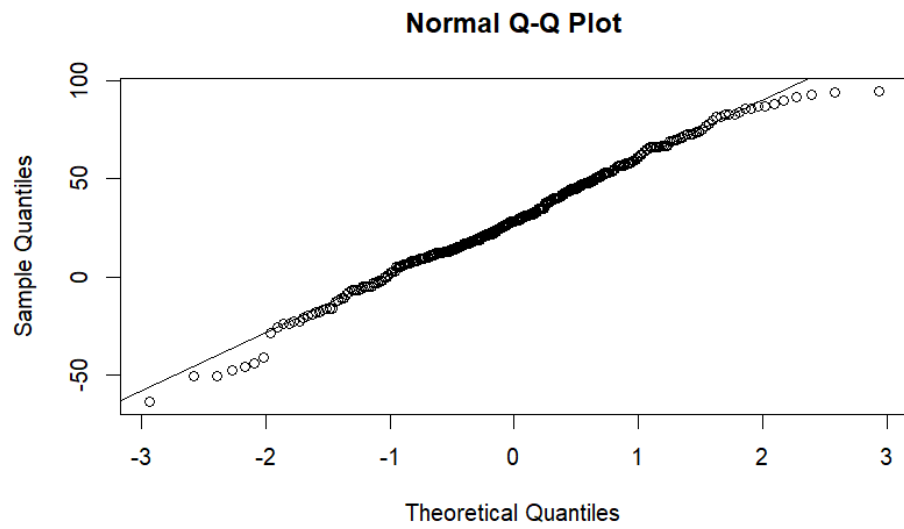
1. <https://twin-cities.umn.edu/news-events/impact-super-bowl-ads>
2. <https://www.mdpi.com/2071-1050/13/13/7127>
3. <https://www.kantar.com/north-america/inspiration/sport-leisure/how-much-is-a-super-bowl-commercial-and-is-it-worth-it>
4. <https://www.statista.com/statistics/668666/excited-super-bowl-ads/>
5. <https://www.statista.com/chart/4325/reasons-to-watch-super-bowl/>
6. [Superbowl Ads | Kaggle](#)
7. [FAQ about Google Trends data - Trends Help](#) -section "How is Google Trends data normalized?"
8. <https://support.google.com/trends/answer/4359550#zippy=> - הסבר על איגוד הטרנדס לפי נושאים



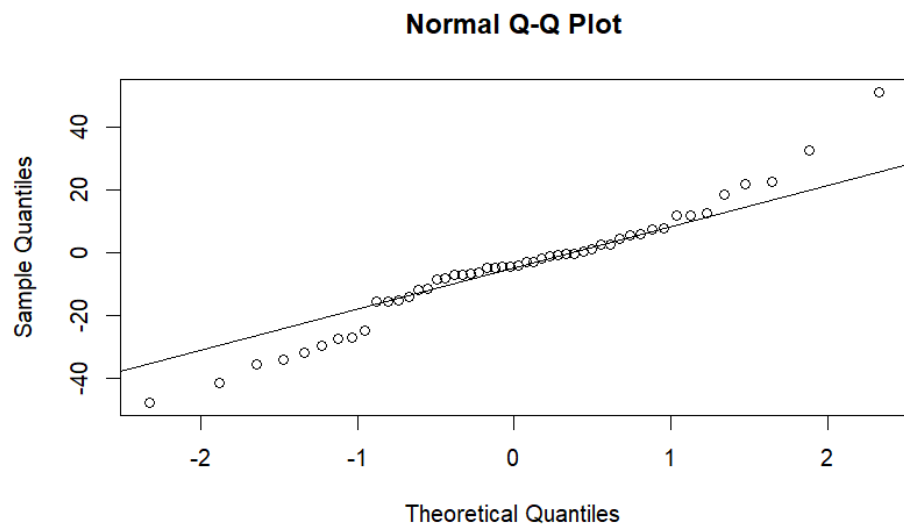
1. סימולציה המראה כי הנתונים מקיימים את משפט הגבול המרכזי



2. גרף שמראה התפלגות נורמלית של 100 שבועות שדגמנו



3. 3. גרף שמראה התפלגות נורמלית של הפרש השבועות.



4. ההנחה שלנו לקירוב נורמליות של הפרש השבועות המבחן המזווג