

האם עושרן של קבוצות כדורגל משפיע על הצלחתן בליגת האלופות?



ט 10
דגיה
מלחמה

קורס "מבוא לסטטיסטיקה"

פרופ' רות הלר, מר צביאל פרוסטיג

יוני 2021

מגישים:

מאי חמרי | 207107111 | Mayhamri8@gmail.com

יובל נוכריאן | 205550528 | Yn8052710@gmail.com

סופי פרחיה | 314710567 | Sofi1103@gmail.com

אמיר חובש | 308369131 | Amirhub1001@gmail.com

תקציר

פרויקט זה עוסק בשאלת הקשר בין רמת העושר של קבוצות כדורגל לבין הצלחתן בליגת האלופות בין השנים 2009-2020. השערת המחקר היא שככל שקבוצה עשירה יותר כך היא מצליחה יותר בליגת האלופות.

אספנו נתונים אודות 11 העונות שצוינו לעיל, נתונים כגון: שווי הקבוצה, רווחי הקבוצה, שכר שחקנים ממוצע, מיקום הקבוצה בעונה כלשהי, דירוג ELO של הקבוצה¹.

הגדרנו שני מדדים מרכזיים שליוו את המחקר: מדד עושר ומדד הצלחה. מדד העושר הינו מדד כמותי משוקלל הנותן משקל שווה לשווי הקבוצה ולרווחי הקבוצה. בנוסף, הגדרנו "רמת עושר", מדד המסווג לפי שלישונים את הקבוצות לרמות עושר בכל שנה. מדד ההצלחה הינו מדד כמותי הנותן דירוג בעבור הגעה לכל שלב בעונת ליגת האלופות מסוימת.

במבחן השוואת פרופורציות בחנו האם השתייכות לרמת עושר גבוהה יותר מעידה על הצלחה גבוהה יותר ואכן דחינו את השערת האפס לפיה הפרופורציה בין רמות העושר השונות שווה.

במבחן הרגרסיה הלינארית שיערנו כי קיים קשר לינארי בין העוצמה הכלכלית של קבוצה להצלחתה בליגת האלופות, ואכן דחינו את השערת האפס ברמת מובהקות גבוהה.

משני המבחנים נסיק כי השערת המחקר נכונה ואכן ככל שקבוצה עשירה יותר (מבחינת רווחים ושווי) היא מצליחה יותר.

מבוא

מאז ומתמיד היה הכדורגל ענף הספורט הפופולרי ביותר בעולם. כדורגל מהווה גאווה לאומית, ומאחד בין קבוצות שונות באוכלוסיית העולם. פופולריות זו הופכת את הכדורגל למשפיע מרכזי בכלכלת העולם ואת קבוצות הכדורגל לעסק לכל דבר. די בלהביט על כניסת המשקיעים ממדינות המפרץ בשנים האחרונות לכדורגל האנגלי והצרפתי (מנצ'סטר סיטי ופריס סן ז'רמן, ממועדוני הכדורגל העשירים בעולם כיום בבעלות שיחיים ממדינות המפרץ) כדי להבין שהכדורגל והכסף כרוכים זה בזה. דוגמאות נוספות שממחישות עד כמה הפופולאריות של ענף זה מתבטאת בכסף – כל שנה נמכרות מיליוני חולצות של שחקנים דומיננטיים. לדוגמה – בשנת 2016 מכרה מנצ'סטר יונייטד 2.85 מיליון חולצות של הקבוצה (עלות ממוצעת של חולצה (100 דולר). כרטיס בשורה האחרונה ביציע (יש שיעידו שממטוס רואים טוב יותר) לגמר ליגת האלופות נמכר בממוצע בכ-3000 יורו.

כמו בכל חברה, ישנן קבוצות עשירות יותר וקבוצות עשירות פחות. העשירות יותר במרבית המקרים סוחפות אחריהן יותר אוהדים, יש להן אצטדיונים מפוארים יותר וכמובן שחקנים דומיננטיים יותר.

אך האם בהכרח הקבוצות העשירות מצליחות יותר? האם הצלחה של קבוצה עם תקציב נמוך היא בגדר 'סיפור סינדרלה' או שמא מדובר בתופעה שכיחה? התשובה הרווחת לשאלה זו ברחבי העולם היא שקיים קשר ברור בין רמת העושר של הקבוצה למידת ההצלחה שלה. זאת, היות שככל שמשאביה של קבוצה מאפשרים יותר, קיימות לה יותר אפשרויות לרכישת שחקנים טובים יותר, להשקיע בקבוצות הנוער של המועדון ואלמנטים נוספים שמשפיעים על הצלחה של מועדון בטווח הקצר ובטווח בארוך. די בלהסתכל על האלופות בכל שנה בליגות המקומיות באירופה ולהיווכח בכך.

עם זאת, שאלה מעניינת שהתשובה עליה לא בהכרח טריוויאלית היא האם גם בליגה שמרכזת את כל הקבוצות הטובות (ומטבע הדברים גם העשירות) באירופה, הלא היא ליגת האלופות סטטוס קוו זה נשמר?

על שאלה זו ננסה לענות במחקר זה, ביתר דיוק – האם עושרן של קבוצות כדורגל משפיע על הצלחתן בליגת האלופות?

¹ שיטה לחישוב רמות המיומנות היחסיות של קבוצות כדורגל אירופיות בהתבסס על הצלחתן במפעלים השונים.

שיטות

תיאור הנתונים ושיטת האיסוף

הנתונים שברשותנו נאגדו על ידנו ממספר מקורות. להלן פירוט הנתונים ומקורותיהם²:

שם העמודה	הסבר	מקור	סוג משתנה	יחידת מידה
TEAM COUNTRY	שם הקבוצה		קטגוריאל	
RANKING	מדינת המקור של הקבוצה	UEFA.com	קטגוריאל	
RATE	מיקום הקבוצה בסיום עונת ליגת האלופות		אורדינלי	1 – מקום ראשון (זכייה) 2 – מקום שני (פיינליסט) S – הגעה לחצי הגמר Q – הגעה לרבע הגמר E – הגעה לשמינית הגמר Y – אי מעבר של שלב הבתים N – אי השתתפות בעונה
SEASON	דירוג שיצרנו		כמותי בדיד	פירוט בהמשך
REV	עונת ליגת האלופות		כמותי בדיד	11 עונות סה"כ, עונה ראשונה-2009-2010 עונה אחרונה-2019-2020 מיליוני יורו
PLAYERS SALARY	רווחי הקבוצה בעונה	Deloitte-football-money-league	כמותי רציף	מיליוני דולר.
VALUE	משכורות ממוצע שחקני הקבוצה בעונה	Globalsportssalaries.com	כמותי בדיד	קיים רק לחלק מהקבוצות; עונה ראשונה-2014-2015. עונה אחרונה-2019-2020 מיליוני דולרים
ELO RATING POWER	שווי הקבוצה – דירוג Forbes	Elo.com	כמותי בדיד	יחידות ELO
CLASS POWER	דירוג ELO באותה עונה		כמותי רציף	פירוט בהמשך
	מדד העוצמה הכלכלית של קבוצה (משתנה שיצרנו)		אורדינלי	A – רמת העושר הגבוהה (שלישון עליון) B – רמת העושר האמצעית (שלישון אמצעי) C – רמת העושר הנמוכה (שלישון תחתון)

מדד ההצלחה

כאמור, בנינו מדד המכמת את הצלחת הקבוצות בליגת האלופות, באופן הבא:

מיקום	מספר קבוצות	ניקוד
זכייה (מקום ראשון)	1	10 נק'
מקום שני	1	9 נק'
הגעה לחצי הגמר	2	8 נק'
הגעה לרבע הגמר	4	6 נק'
הגעה לשמינית הגמר	8	4 נק'
הגעה לשלב הבתים	16	2 נק'
אי השתתפות בליגת האלופות		0 נק'

² הנתונים מאוגדים בטבלת האקסל המצורפת למסמך – ראה נספח א'.

הרציונאל מאחורי שיטת הניקוד הינו שבראייתנו ההבדל בין שלב לשלב בשלבים המכריעים (חצי גמר וגמר) הוא פחות משמעותי מאשר ההבדל בין השלבים עד שלב רבע הגמר. ומכאן נובע הפער בניקוד (2 נק' אל מול נק' אחת בלבד).

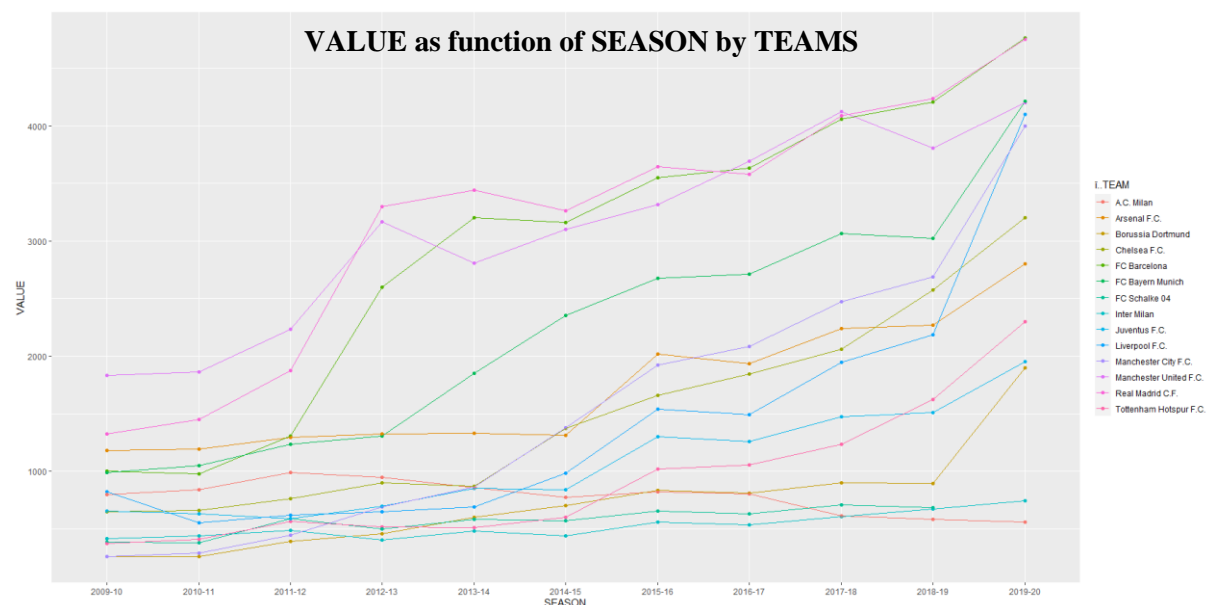
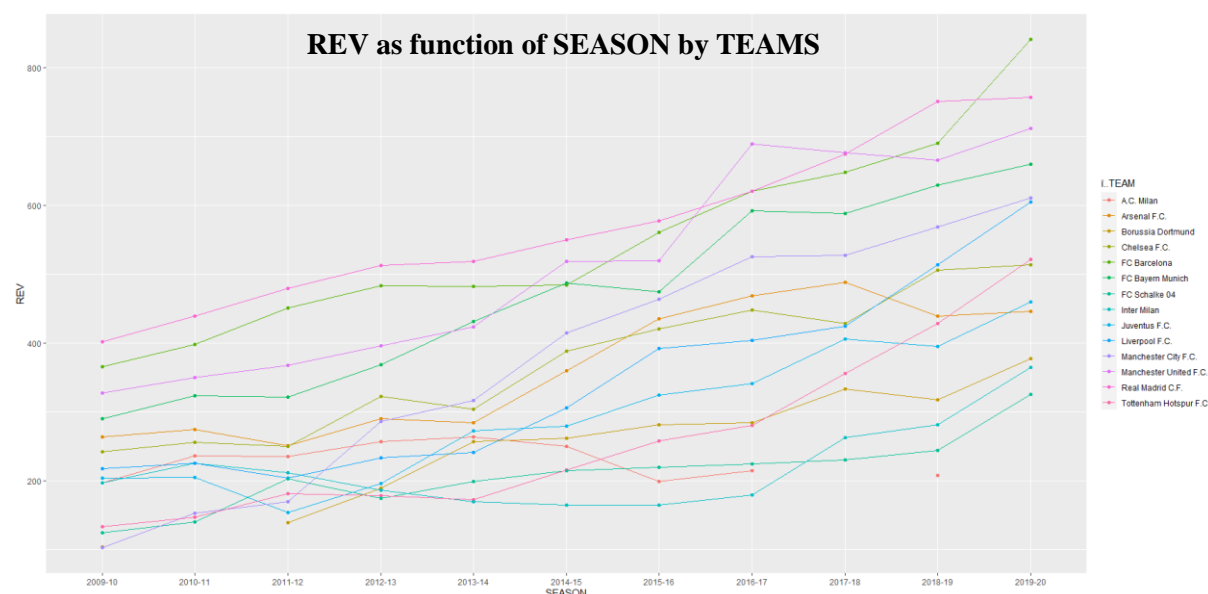
על מנת לבחון האם המדד שלנו משקף דיו את מידת ההצלחה של הקבוצות, צירפנו לקובץ הנתונים, נתונים אודות מדד ה-ELO של הקבוצות ב-11 העונות שנבחנו. בנספח ה' ניתן לראות מבחן רגרסיה לינארית הבוחן את הקשר בין ממוצע מדד ה-ELO של 14 הקבוצות המרכזיות שאותן נבחן לבין מדד העושר של הקבוצות הללו.

מדד העושר

במטרה לייצר מדד משוקקל של שני פרמטרי העושר המרכזיים שלנו: REV, VALUE, הגדרנו משתנה חדש – POWER, ע"י המשוואה הבאה:

$$POWER = \frac{1}{2} * REV + \frac{1}{2} * VALUE$$

נדגיש בשלב זה כי הרגשנו בנוח להגדיר מדד זה, כיוון שניתן לזהות בשני המשתנים הללו עלייה הדרגתית מתמדת ברווחים ובשווי של כל הקבוצות עם השנים, כפי שניתן לראות בשני הגרפים הבאים:



כמו כן, נציין כי הסיבה שלא כללנו את המשתנה PLAYERS SALARY (משתנה נוסף המעיד על עוצמה כלכלית) במדד זה הינה מכיוון שאין ברשותנו נתונים על משתנה זה בכלל השנים הרלוונטיות.

מבחן השוואת פרופורציות

נבחן את פרופורציית הנקודות שכל רמת עושר צברה במהלך 11 העונות (סך כל הנקודות של כל רמה חלקי סך הנקודות הכללי). יצרנו שלושה "רמות עושר" (C,B,A) בהתאם לשלישונים של מדד העושר בכל עונה.

נחלק לשלושה מבחנים: $P_A VS P_B$, $P_A VS P_C$, $P_B VS P_C$. בעבור כל אחד מהמבחנים נבדוק האם פרופורציית רמת העושר הגדולה יותר גדולה מפרופורציית רמת העושר הנמוכה יותר ברמת מובהקות $\alpha = 0.05$.

תיאור הנתונים:

המדגם שלנו מורכב מ-208 משתנים מקריים שטווח הערכים שלהם נע בין 0-10 (בהתאם למדד ההצלחה אותו הגדרנו לעיל). המרנו את טווח הערכים 0-10 לטווח של משתנה מקרי ברנולי (0,1) באופן הבא – "1" במידה והמשתנה שלנו קיבל ערכים גדולים או שווים ל-6 (הגעה לחצי גמר); "0" במידה והמשתנה קיבל ערכים קטנים שווים ל-4. התצפיות במדגם מתחלקות באופן שווה בין רמות העושר השונות:

למה? דא לא כלכלי
אם לא כלכלי
rate = plc
wilcoxon
אם לא כלכלי

1. 70 תצפיות ברמת עושר A $\frac{70}{208} = 0.3365$
2. 73 תצפיות ברמת עושר B $\frac{73}{208} = 0.3509$
3. 65 תצפיות ברמת עושר C $\frac{65}{208} = 0.3125$

השערות:

- השערת האפס – לא קיים הבדל בהפרש הפרופורציות בין זוג רמות עושר.
כלומר $H_0: p_1 = p_0 \rightarrow p_1 - p_0 = 0$
- השערת המחקר – פרופורציית רמת העושר הגבוהה גדולה מפרופורציית רמת העושר הנמוכה.
כלומר $H_1: p_1 > p_0 \rightarrow p_1 - p_0 > 0$
- ההשערות זהות עבור שלושת המבחנים.

להלן סכמה גנרית של המבחן:

$$\begin{array}{ll} H_0: p_1 = p_0 \rightarrow p_1 - p_0 = 0 & H_1: p_1 > p_0 \rightarrow p_1 - p_0 > 0 \\ n_1 = X & n_0 = Y \end{array}$$

נאמוד את פרופורציית ההצלחות לכל רמת עושר באמצעות ממוצעי הדגימות:

$$\widehat{p}_1 = \frac{\sum_{i=1}^X x_i}{X}, \quad \widehat{p}_0 = \frac{\sum_{i=1}^Y y_i}{Y}$$

מכיוון שהמדגם שלנו מכיל יותר מ-30 תצפיות, נחיל על הממוצעים את משפט הגבול המרכזי:

$$\widehat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right), \quad \widehat{p}_0 \sim N\left(p_0, \frac{p_0(1-p_0)}{n_0}\right)$$

כיוון ששני האומדים מתפלגים נורמלית, גם ההפרש שלהם (צירוף לינארי) מתפלג נורמלית, לכן נקבל:

$$\widehat{p}_1 - \widehat{p}_0 \sim N\left(p_1 - p_0, \frac{p_1(1-p_1)}{n_1} + \frac{p_0(1-p_0)}{n_0}\right)$$

תחת השערת האפס, נקבל כי $p_1 = p_0 = p$:

$$\widehat{p}_1 - \widehat{p}_0 \sim N\left(p_1 - p_0, \frac{p_1(1-p_1)}{n_1} + \frac{p_0(1-p_0)}{n_0}\right) = N\left(0, \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_0}\right)$$

נתקן על מנת לקבל התפלגות נורמלית סטנדרטית:

$$\frac{\widehat{p}_1 - \widehat{p}_0}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_0}}} \sim N(0,1)$$

נשים לב כי קיבלנו ביטוי שתלוי ב- p , לכן נאמוד את p עם הממוצע של שני האומדים: $\tilde{p} = \frac{\sum_{i=1}^{70} x_i + \sum_{i=1}^{73} x_i}{143}$ ונקבל את סטטיסטי המבחן $(\theta \sim N(0,1))$:

$$\theta = \frac{\widehat{p}_1 - \widehat{p}_0}{\sqrt{\tilde{p}(1-\tilde{p})\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}} = \frac{\widehat{p}_1 - \widehat{p}_0}{\sqrt{\frac{\sum_{i=1}^{70} x_i + \sum_{i=1}^{73} x_i}{143} \left(1 - \frac{\sum_{i=1}^{70} x_i + \sum_{i=1}^{73} x_i}{143}\right) \left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}$$

לכן, נחפש k כך ש: $p(\theta > k) = 0.05$. מכיוון ש- θ מתפלג נורמלית סטנדרטית, $k = Z_{0.95} = 1.644$

מבחן רגרסיה לינארית

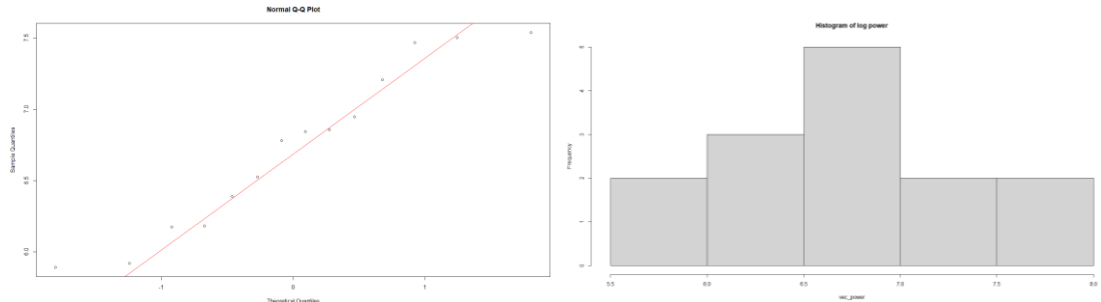
נבחן האם קיים קשר לינארי בין ממוצע משתנה ה-POWER לבין ממוצע משתנה ה-RATE לאורך 11 העונות שנבחנו ברמת מובהקות $\alpha = 0.05$. המדגם שלנו מורכב מ-14 קבוצות להן קיימים הנתונים הנ"ל. נבחן זאת באמצעות רגרסיה לינארית עפ"י שיטת "ריבועים פחותים".

14 הקבוצות במדגם מתפלגות בין הליגות המובילות באירופה באופן הבא:

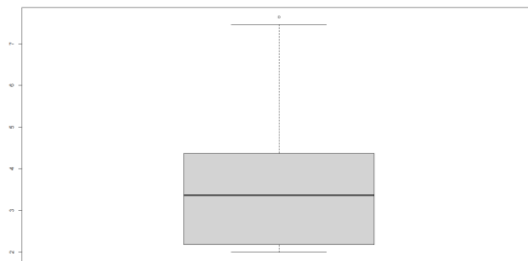
גרמניה		איטליה		אנגליה		ספרד	
	FC Bayern Munich		A.C. Milan		Manchester United F.C.		Real Madrid CF
	Borussia Dortmund		Inter Milan		Manchester City F.C.		FC Barcelona
	FC Schalke 04		Juventus F.C.		Tottenham Hotspur F.C.		
					Chelsea F.C.		
					Arsenal F.C.		
					Liverpool F.C.		

תיאור הנתונים :

- המשתנה הבלתי תלוי (POWER) – נראה כי מתפלג בקירוב נורמלית. כמו כן, כיוון שקיים הבדל משמעותי בסדרי הגודל בין המשתנים הנבחרים, נבצע טרנספורמציה לוג למשתנה POWER בעל הערכים הגבוהים.



- המשתנה התלוי (RATE) – לא ניתן לזהות התפלגות מוכרת כלשהי. להלן תרשים קופסא שמתאר את התפלגות המשתנה :



השערות :

- השערת האפס – לא קיים קשר לינארי בין המשתנים, כלומר $H_0: b = 0$
- השערת המחקר – קיים קשר לינארי חיובי בין המשתנים, כלומר $H_1: b \neq 0$

להלן סכמה של המבחן :

הנחות :

- קיום קשר לינארי $y_i = a + bx_i + \varepsilon_i$. כאשר y_i הוא ממוצע ה-RATE עבור קבוצה i , x_i הוא ממוצע ה-POWER עבור קבוצה i , ε_i הוא הסטייה מקו הרגרסיה עבור קבוצה i .
- הסטיות מתפלגות נורמלית עם תוחלת אפס ושונות לא ידועה, נאמת הנחה זו באמצעות $qqnorm$ בפרק 'תוצאות'.
- קיום אי תלות בין כל זוג קבוצות שונות. נדגיש כי הנחה זו בעייתית, שכן בהכרח קיימת השפעה של מיקום קבוצה אחת בעונה כלשהי על קבוצה אחרת. כמו כן, גם מדד ה-POWER מושפע סביבתית משינויים כלכליים וכיו"ב.

$$T = \frac{\hat{b}_{MLE} - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{14} (x_i - \bar{x})^2}}} \sim t_{12}$$

- נבחין כי אנו במצב של בדיקת השערות על השיפוע כאשר השונות אינה ידועה ולכן נאמוד את השונות.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{14} (y_i - \hat{y}_i)^2}{12}$$

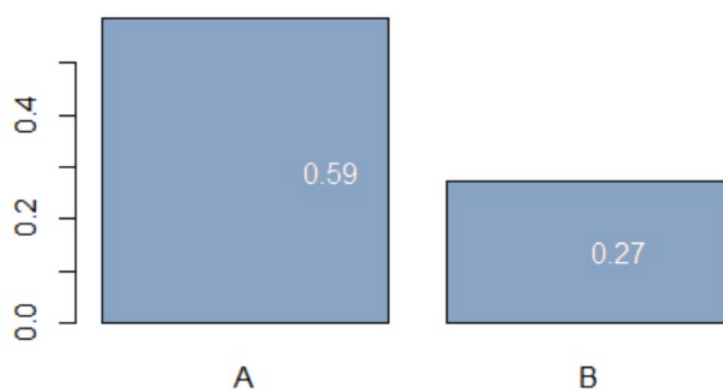
- נדחה את השערת האפס ברמה מובהקות $\alpha = 0.05$ אם $|T| > t_{12, 1-\alpha/2}$ או אם $P_v < \alpha$.
- נבצע את המבחן באמצעות הפקודה ב-R `lm(rate~power)`.

תוצאות

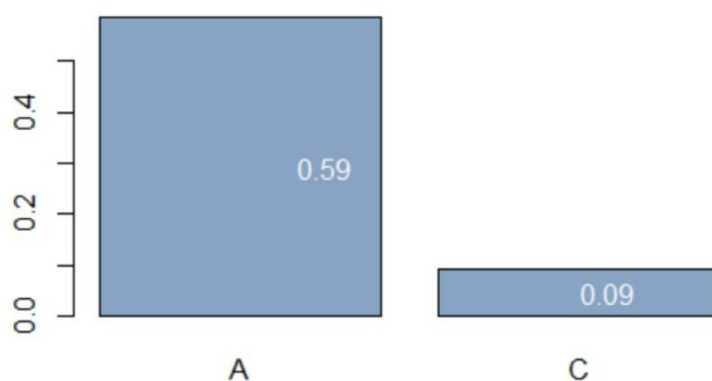
מבחן השוואת פרופורציות

resault	p value	θ	R_α	α	\tilde{p}	n_0	n_1	\hat{p}_0	\hat{p}_1	test
Reject H_0	$8.26 \cdot 10^{-5}$	3.767	$\theta > 1.644$	0.05	0.42657	73	70	0.27	0.59	$\hat{p}_A = \hat{p}_1$ $\hat{p}_B = \hat{p}_0$
Reject H_0	$9.16 \cdot 10^{-10}$	6.012	$\theta > 1.644$	0.05	0.34814	65	70	0.09	0.59	$\hat{p}_A = \hat{p}_1$ $\hat{p}_C = \hat{p}_0$
Reject H_0	0.003	2.724	$\theta > 1.644$	0.05	0.18840	65	73	0.09	0.27	$\hat{p}_B = \hat{p}_1$ $\hat{p}_C = \hat{p}_0$

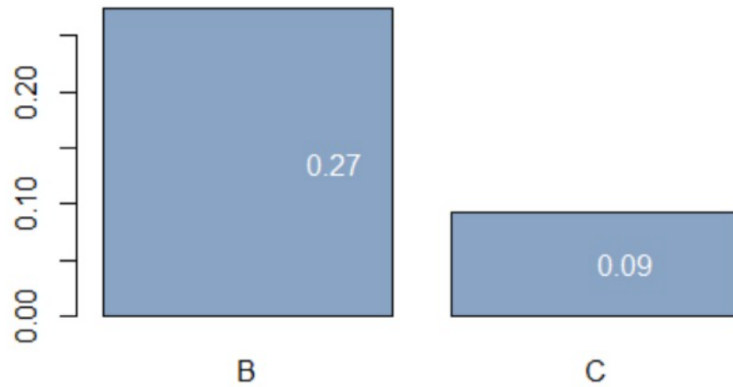
1. p_A VS p_B – נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$. ולכן נקבל את השערת המחקר לפיה $p_A > p_B$. כלומר, פרופרציית ההצלחה ברמת עושר "A" גבוהה מפרופרציית ההצלחה ברמת עושר "B".



2. p_A VS p_C – נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$. ולכן נקבל את השערת המחקר לפיה $p_A > p_C$. כלומר, פרופרציית ההצלחה ברמת עושר "A" גבוהה מפרופרציית ההצלחה ברמת עושר "C".



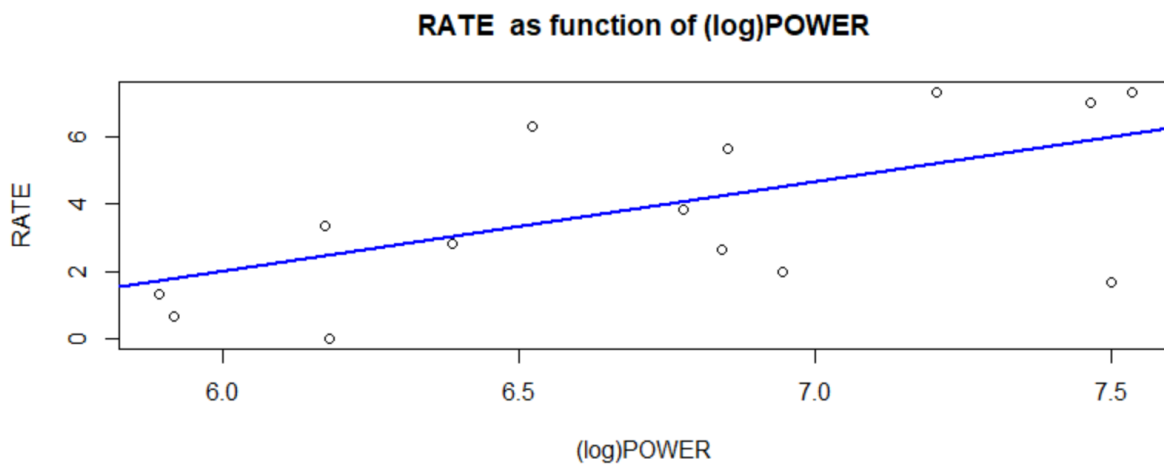
3. p_B VS p_C – נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$. ולכן נקבל את השערת המחקר לפיה $p_B > p_C$. כלומר, פרופרציית ההצלחה ברמת עושר "B" גבוהה מפרופרציית ההצלחה ברמת עושר "C".



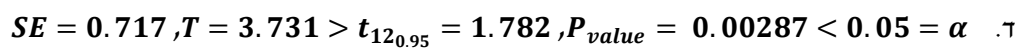
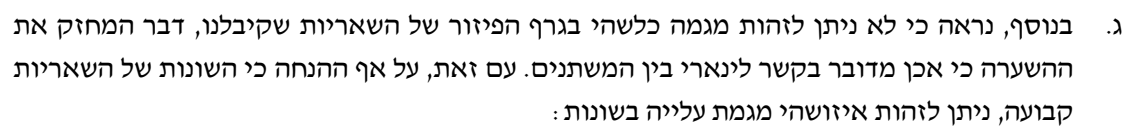
4. נשים לב כי בשלושת המבחנים, בפרט עבור השניים הראשונים קיבלנו P Value קטן מאוד, נתון המחזק את מסקנתנו.
5. סה"כ, מניתוח שלושת המבחנים נקבל כי השתייכות לרמת עושר גבוהה יותר מעידה על הצלחה גדולה יותר.
6. כמו כן, בנספח י' ניתן למצוא גרף נוסף המחזק את השערתנו לפיה השתייכות לרמת עושר גבוהה יותר פירושה הצלחה גבוהה יותר.

מבחן רגרסיה לינארית

1. להלן גרף המתאר את הרגרסיה הלינארית RATE כפונקציה של log POWER:



2. ממצאי הרצת הפקודה `summary(lm(RATE~log(POWER)))`:
 - א. משוואת הרגרסיה הלינארית – $RATE = 2.675 * \log(POWER) - 14.053$.
 - א. $\hat{b}_{MLE} = 2.675 > 0$.
 - ב. נראה כי השאריות בין הערכים שקיבלנו לערכים הצפויים מתפלגות נורמלית באמצעות גרף qqplot, דבר המאמת את הנחתנו:



H_1

דיון ומסקנות

1. בטרם ניגשנו למבחנים השונים, שיערנו כי קיים קשר בין עוצמה כלכלית של קבוצות כדורגל לבין מידת ההצלחה שלהן בליגת האלופות. בפרט, ככל שקבוצה עשירה יותר כך היא מצליחה יותר.
2. תוצאות המבחנים לא הפתיעו אותנו, שכן בשניהם דחינו את השערת האפס ברמת מובהקות 0.05 בצורה משכנעת.
3. מבחן השוואת הפרופורציות הצביע על כך, שהשתייכות לרמת עושר A (שלישון עליון ברמת העושר בכל שנה) מעלה את הסיכוי להצלחת הקבוצה בליגת האלופות.
4. ממבחן הקשר הלינארי הסקנו כי קיים קשר לינארי חיובי בין מדד העושר הממוצע של קבוצה לבין מדד הצלחתה הממוצע.
5. תת שאלה נוספת שהתעוררה במהלך המחקר היא – מי מבין האלמנטים של העושר (REV, VALUE, PLAYERS SALARY) משפיע יותר על ההצלחה.
 - a. ביצענו שלושה מבחני רגרסיה נוספים עבור כל אחד מהמשתנים הללו על מנת לבחון האם קיים קשר לינארי בינם לבין ממוצע מדד RATE (ראה נספחים ו', ז', ח').
 - b. המשתנה PLAYERS SALARY (שכר שחקנים ממוצע) הצביע על הקשר הלינארי החזק ביותר אל מול ממוצע מדד ה-RATE. נסתייג ונאמר כי מבחן זה בוצע על 6 עונות בלבד, בהיעדר נתונים.
 - c. עם זאת, ממצא זה לא בהכרח מפתיע, שכן בל נשכח שהכדור בידיים (או יותר נכון, ברגליים) של השחקנים, שסביר שהטובים שביניהם ישחקו בקבוצות העשירות יותר.
6. לסיכום, ניתן להגיד שעל בסיס הנתונים והמדדים שברשותנו ועל בסיס תוצאות המבחנים שביצענו כי **עושרן של קבוצות כדורגל אכן משפיע על הצלחתן בליגת האלופות**.
7. ממצאי המחקר שביצענו מתיישבים עם מרבית ממצאי מחקרי עבר. בחלק ממחקרי העבר נמצא כי מספיקה השתייכות לרמת עושר מסוימת על מנת להצליח. כלומר, אכן כל הקבוצות ברמת העושר הגבוהה מתווגות כמצליחות, אך ההבדל בין מידת ההצלחה הוא זניח. ניתן להגיד בזהירות (מעט תצפיות) שממצא זה מתיישב עם ממצאי מבחן הרגרסיה הלינארית.
8. אם כן, המסקנה היא שבמידה ומחר אתם מתעוררים כבעלים של קבוצת כדורגל ואתם רוצים להצליח, הדבר החשוב ביותר הוא התקציב... 😊
9. הסתייגויות:
 - א. הגדלת המדגם (יותר קבוצות) במבחן הרגרסיה הלינארית עשויה לשפר / להרע את ממצאי המחקר. כאמור, נלקחו 14 קבוצות בלבד, כיוון שרק להן היו נתונים אודות כל 11 השנים.
 - ב. במבחן הרגרסיה הלינארית עשינו ממוצע של מדד העושר של כל הקבוצות, דבר שעלול להשפיע על ממצאי המבחן הלינארי, כיוון שלא נטרלנו את השפעת השנים. ייתכן כי ביצוע מבחן רגרסיה לינארית עם שני משתנים היה מבחן טוב יותר, שכן היו בו גם יותר תצפיות.
 - ג. מכיוון שהאינטרוול שבחנו הוא של 11 עונות, וכן כי מיצענו את הנתונים, ייתכן כי פספסנו קבוצות שלאור כניסת משקיעים והגדלת הכנסות הצליחו יותר באינטרוול קטן יותר של שנים, אך הדבר לא בא לידי ביטוי בחתך של 11 עונות.
 - ד. לא הסתכלנו באיסוף הנתונים ובביצוע המבחנים על השפעות כלכליות (משבר כלכלי, אינפלציה, חוקי ה-Fair Play וכו') שבבירור השפיעו על נתוני ה"עושר" שברשותנו ועלולים להטות את תוצאות המבחנים.
 - ה. ענינו על שאלת ליגת האלופות, אין הדבר אומר כי בהכרח ניתן להשליך את ממצאי המחקר על ליגות אחרות בענף הכדורגל או בענפי ספורט אחרים.
10. הצעות לכיווני מחקר נוספים:
 - א. העמקה בשאלה "היכן עובר הקו שממנו אין השפעה של התקציב על הצלחת הקבוצה?", במילים אחרות, לבחון מתי שיפוע קו הרגרסיה הופך לקבוע ולא חיובי עולה.

- ב. ביצוע מבחן רגרסיה לינארית תוך נטרול השפעת השנים, כלומר מבחן רגרסיה לינארית עם יותר משני משתנים.
- ג. הרחבת המחקר לליגות נוספות בענף הכדורגל ובענפי ספורט נוספים.

רפרנסים

1. מחקר עבר (2018) של Tiago Manuel Valente Pereira הבוחן קשר בין כוח פיננסי של קבוצות כדורגל להצלחה – [קישור](#).
2. דוחות שנתיים של חברת Deloitte, בהם פירוט על 20 הקבוצות הרווחיות באירופה – [קישור](#).
3. דוחות שנתיים של אתר globalsportssalaries, בהם פירוט של שכר ממוצע של שחקנים בחלוקה לקבוצות – [קישור](#).
4. דירוג פורבס של שווי מועדוני הכדורגל – [קישור](#).
5. אתר UEFA, דירוג הקבוצות בעונות השונות בליגת האלופות – [קישור](#).
6. מדד Elo המדרג את קבוצות הכדורגל בעולם לפי דירוג Elo – [קישור](#).

נספחים

נספח א' – טבלת איסוף הנתונים של המחקר



Microsoft Excel
Worksheet

	A	B	C	D	E	F	G	H	I	J	K	L
1	TEAM	COUNTRY	RANKING	RATE	SEASON	REV	PLAYERS	VALUE	ELO_RAT	POWER	CLASS_POWER	
2	FC Barcelona	Spain	q	6	2019-20	840.8	\$12,284,55	4760	2228	2800.4	A	
3	Real Madrid	Spain	e	4	2019-20	757.3	\$11,154,65	4750	2186	2753.65	A	
4	Manchester	England	n	0	2019-20	711.5	\$7,657,000	4200	2168	2455.75	A	
5	FC Bayern	Germany	1	10	2019-20	660.1	\$8,119,896	4215	2353	2437.55	A	

נספח ב' – דוח שנתי לדוגמא Deloitte

Deloitte Football Money League 2020 | Ups and downs

Ups and downs

2018/19 Revenue (€m)

1	↑	1	FC Barcelona	840.8
2	↓	(1)	Real Madrid	757.3
3	↔	0	Manchester United	711.5
4	↔	0	Bayern Munich	660.1
5	↑	1	Paris Saint-Germain	635.9
6	↓	(1)	Manchester City	610.6
7	↔	0	Liverpool	604.7
8	↑	2	Tottenham Hotspur	521.1
9	↓	(1)	Chelsea	513.1
10	↑	1	Juventus	459.7
11	↓	(2)	Arsenal	445.6
12	↔	0	Borussia Dortmund	377.1
13	↔	0	Atlético de Madrid	367.6
14	↔	0	FC Internazionale Milano	364.6
15	↑	1	Schalke 04	324.8
16	↓	(1)	AS Roma	231.0
17	n/a	new	Olympique Lyonnais	220.8
18	↑	2	West Ham United	216.4
19	↓	(2)	Everton	213.0
20	n/a	new	SSC Napoli	207.4











● DFML position ● Change on previous year ● Number of positions changed

2017/18 Revenue (€m)

1	↑	1	Real Madrid	750.9
2	↑	1	FC Barcelona	690.4
3	↓	(2)	Manchester United	665.8
4	↔	0	Bayern Munich	629.2
5	↔	0	Manchester City	568.4
6	↑	1	Paris Saint-Germain	541.7
7	↑	2	Liverpool	513.7
8	↔	0	Chelsea	505.7
9	↓	(3)	Arsenal	439.2
10	↑	1	Tottenham Hotspur	428.3
11	↓	(1)	Juventus	394.5
12	↔	0	Borussia Dortmund	317.2
13	↔	0	Atlético de Madrid	304.4
14	↑	1	FC Internazionale Milano	280.8
15	n/a	new	AS Roma	250.0
16	↔	0	Schalke 04	243.8
17	↑	3	Everton	212.9
18	n/a	new	AC Milan	207.7
19	n/a	new	Newcastle United	201.5
20	↓	(3)	West Ham United	197.9

נספח ג' – דוח שנתי לדוגמא sportingintelligence

GLOBAL SPORTS SALARIES SURVEY 2014 TOP 12

	RANK & TEAM (2013 RANK)	LEAGUE	AVE ANNUAL PAY £ (AVE WEEK PAY)	AVE ANNUAL PAY \$ (AVE WEEK PAY)
	1 (1) Manchester City	EPL	£5,337,944 (£102,653)	\$8,109,912 (\$155,960)
	2 (5) New York Yankees	MLB	£5,286,628 (£101,666)	\$8,031,948 (\$154,461)
	3 (2) Los Angeles Dodgers	MLB	£5,119,701 (£98,456)	\$7,778,336 (\$149,583)
	4 (3) Real Madrid	La Liga	£4,993,393 (£96,027)	\$7,586,438 (\$145,893)
	5 (4) Barcelona	La Liga	£4,901,327 (£94,256)	\$7,446,562 (\$143,203)
	6 (16) Brooklyn Nets	NBA	£4,485,019 (£86,250)	\$6,814,067 (\$131,040)
	7 (9) Bayern Munich	Bundesliga	£4,402,905 (£84,671)	\$6,689,311 (\$128,641)
	8 (12) Manchester United	EPL	£4,322,251 (£83,120)	\$6,566,775 (\$126,284)
	9 (19) Chicago Bulls	NBA	£3,985,706 (£76,648)	\$6,055,463 (\$116,451)
	10 (8) Chelsea	EPL	£3,984,536 (£76,626)	\$6,053,686 (\$116,417)
	11 (15) Arsenal	EPL	£3,901,923 (£75,037)	\$5,928,172 (\$114,003)
	12 (20) NY Knicks	NBA	£3,862,191 (£74,273)	\$5,867,808 (\$112,842)

נספח ד' – דירוג ה-Elo נכון למאי 2021 :

Euro Top 25	
Club	Elo
1 Man City	2009
2 Bayern	1987
3 Liverpool	1936
4 Real Madrid	1935
5 Man United	1922
6 Chelsea	1919
7 Barcelona	1910
8 Inter	1894
9 Atlético	1892
10 Paris SG	1872
11 Arsenal	1863
12 Juventus	1846
13 Sevilla	1844
14 Dortmund	1839
15 Atalanta	1833
16 Tottenham	1827
17 Villarreal	1821
18 RB Leipzig	1812
19 West Ham	1807
20 Ajax	1806
21 Napoli	1800
22 Porto	1799
23 Milan	1798
24 Leicester	1798
25 Leeds	1780

נספח ה' – מבחן רגרסיה לינארית בין המשתנים ELO RATING ל-POWER :

נבחן האם קיים קשר לינארי בין ממוצע משתנה ה-POWER (המשתנה הבלתי תלוי) לבין ממוצע משתנה ה-ELO RATING (המשתנה התלוי) לאורך 11 העונות שנבחנו ברמת מובהקות $\alpha = 0.05$. המדגם שלנו מורכב מ-14 קבוצות להן קיימים הנתונים הנ"ל. נבחן זאת באמצעות רגרסיה לינארית עפ"י שיטת "ריבועים פחותים".

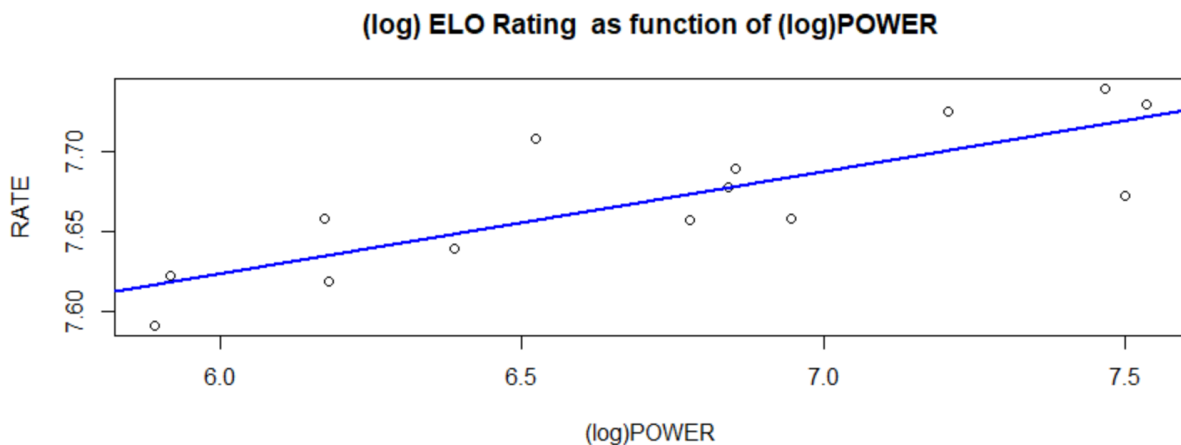
השערות :

- השערת האפס – לא קיים קשר לינארי בין המשתנים, כלומר $H_0: b = 0$
- השערת המחקר – קיים קשר לינארי בין המשתנים, כלומר $H_1: b \neq 0$

מהלך המבחן – המבחן יבוצע באופן זהה למבחן שבוצע בגוף העבודה, ועל כן לפירוט נוסף אודות סכמת המבחן הכללית וההנחות הרלוונטיות ראה תת סעיף "מבחן רגרסיה לינארית" תחת פרק "שיטות".

ממצאי המבחן :

- להלן גרף המתאר את הרגרסיה הלינארית log ELO RATING כפונקציה של log POWER :

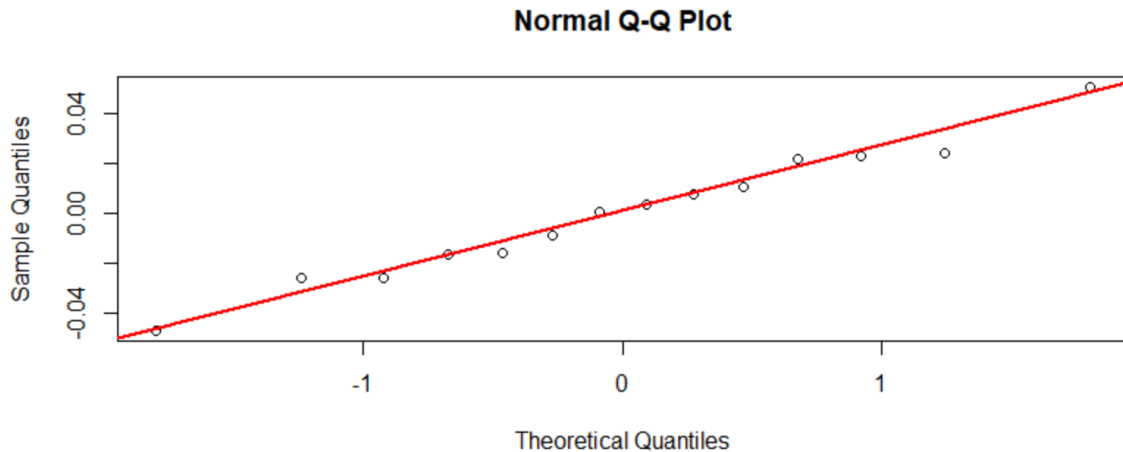


- ממצאי הרצת הפקודה `summary(lm(RATE~log(REV)))` :

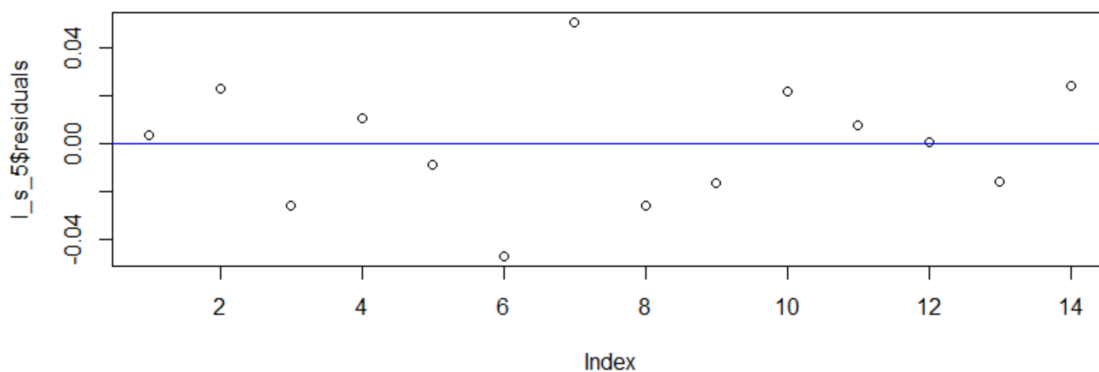
1. משוואת הרגרסיה הלינארית – $(\log)ELORating = 0.0638 * \log(POWER) - 7.240$.

$$\hat{b}_{MLE} = 0.0638 > 0 \quad .a$$

2. נראה כי השאריות בין הערכים שקיבלנו לערכים הצפויים מתפלגות נורמלית באמצעות גרף qqplot :



3. בנוסף, נראה כי לא ניתן לזהות מגמה כלשהי בגרף הפיזור של השאריות שקיבלנו, דבר המחזק את ההשערה כי אכן מדובר בקשר לינארי בין המשתנים:



4. $SE = 0.0638, T = 4.916 > t_{12,0.95} = 1.782, P_{value} = 0.000356 < 0.05 = \alpha$

5. לכן, נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$ ונקבל את ההשערה האלטרנטיבית, לפיה קיים קשר לינארי בין ממוצע מדד ה-POWER של קבוצות כדורגל לממוצע מדד ה-ELO. **בפרט, מכיוון שהסטטיסטי שלנו התברר כחיובי, יכולנו לקיים מבחן חד צדדי ימני ולקבל $\frac{P_{value}}{2}$ ולכן נוכל להסיק כי מדובר בקשר לינארי חיובי.** בנוסף, נבחין כי קיבלנו P_{value} קטן מאוד, מה שמעיד על חוזק הקשר הלינארי, ובכך מחזק את השערתנו.

6. נסתייג ונאמר כי מדד ה-ELO מתבסס על הצלחה גם בליגות המקומיות ועל כן אין הדבר בהכרח מעיד על הצלחה בליגת האלופות בלבד, אלא הצלחה כוללת בכלל המפעלים.

נספח ו' – מבחן רגרסיה לינארית בין המשתנים REV (רווחים במליוני דולרים) ל-RATE:

נבחן האם קיים קשר לינארי בין ממוצע משתנה ה-REV (המשתנה הבלתי תלוי) לבין ממוצע משתנה ה-RATE (המשתנה התלוי) לאורך 11 העונות שנבחנו ברמת מובהקות $\alpha = 0.05$. המדגם שלנו מורכב מ-14 קבוצות להן קיימים הנתונים הנ"ל. נבחן זאת באמצעות רגרסיה לינארית עפ"י שיטת "ריבועים פחותים".

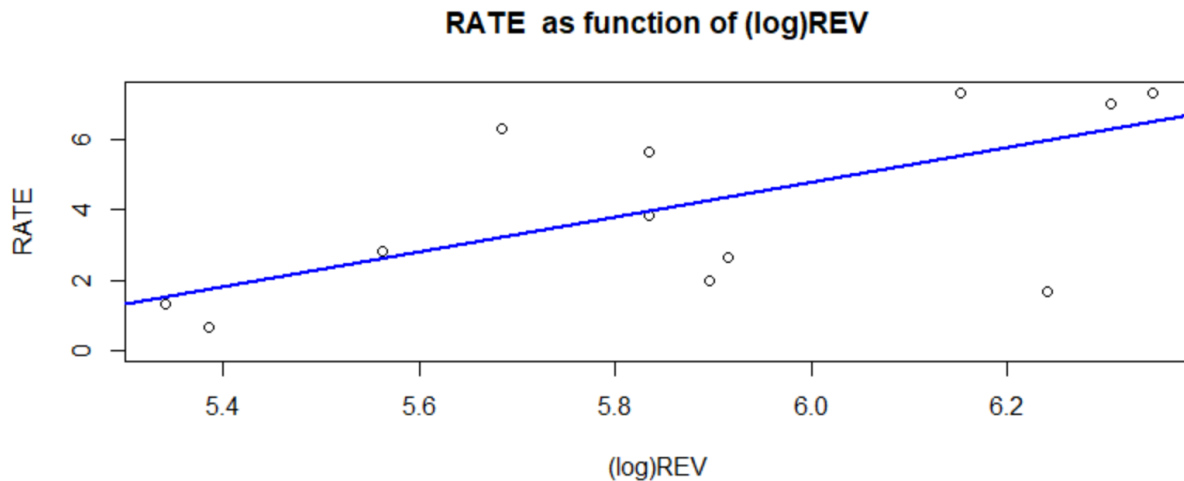
השערות:

- השערת האפס – לא קיים קשר לינארי בין המשתנים, כלומר $H_0: b = 0$
- השערת המחקר – קיים קשר לינארי בין המשתנים, כלומר $H_1: b \neq 0$

מהלך המבחן – המבחן יבוצע באופן זהה למבחן שבוצע בגוף העבודה, ועל כן לפירוט נוסף אודות סכמת המבחן הכללית וההנחות הרלוונטיות ראה תת סעיף "מבחן רגרסיה לינארית" תחת פרק "שיטות".

ממצאי המבחן :

- להלן גרף המתאר את הרגרסיה הלינארית RATE כפונקציה של $\log REV$:

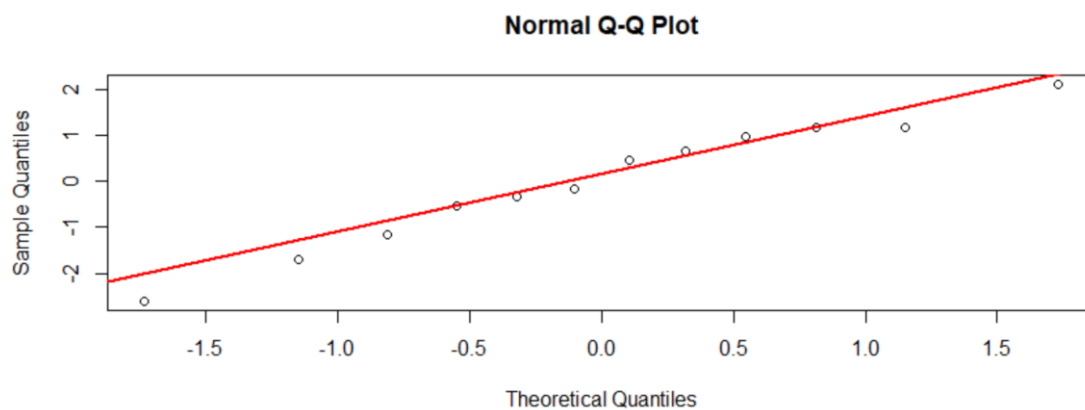


- ממצאי הרצת הפקודה `summary(lm(RATE~log(REV)))` :

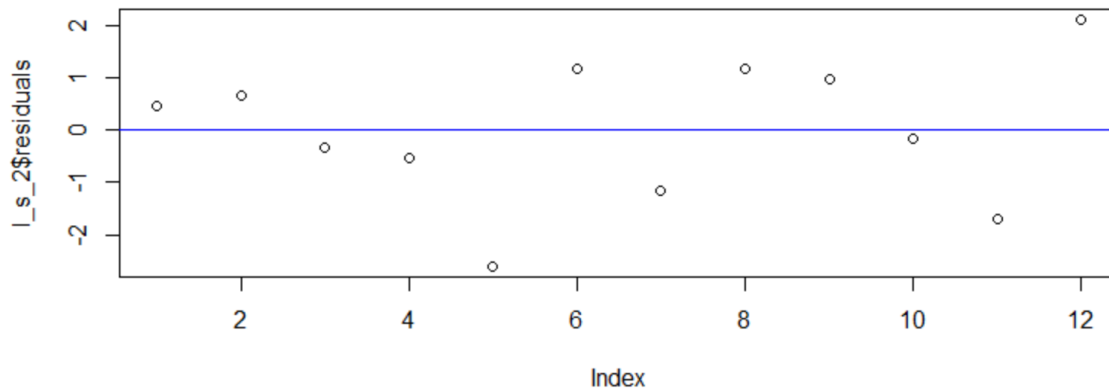
1. משוואת הרגרסיה הלינארית – $RATE = 4.955 * \log(REV) - 24.954$.

b. $\hat{b}_{MLE} = 4.955 > 0$

2. נראה כי השאריות בין הערכים שקיבלנו לערכים הצפויים מתפלגות נורמלית באמצעות גרף qqplot :



3. בנוסף, נראה כי לא ניתן לזהות מגמה כלשהי בגרף הפיזור של השאריות שקיבלנו, דבר המחזק את ההשערה כי אכן מדובר בקשר לינארי בין המשתנים :



$$SE = 1.252, T = 3.957 > t_{12,0.95} = 1.782, P_{value} = 0.00270 < 0.05 = \alpha \quad .4$$

5. לכן, נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$ ונקבל את ההשערה האלטרנטיבית, לפיה קיים קשר לינארי בין רווחים במיליוני דולרים של קבוצות כדורגל להצלחתן בליגת האלופות. **בפרט, מכיוון שהסטטיסטי שלנו התברר כחיובי, יכולנו לקיים מבחן חד צדדי ימני ולקבל $\frac{P_{value}}{2}$ ולכן נוכל להסיק כי מדובר בקשר לינארי חיובי.** בנוסף, נבחין כי קיבלנו P_{value} קטן מאוד, מה שמעיד על חוזק הקשר הלינארי, ובכך מחזק את השערתנו.

נספח ז' – מבחן רגרסיה לינארית בין המשתנים VALUE (שווי קבוצה במיליוני יורו) ל-RATE:

נבחן האם קיים קשר לינארי בין ממוצע משתנה VALUE (המשתנה הבלתי תלוי) לבין ממוצע משתנה ה-RATE (המשתנה התלוי) לאורך 11 העונות שנבחנו ברמת מובהקות $\alpha = 0.05$. המדגם שלנו מורכב מ-14 קבוצות להן קיימים הנתונים הנ"ל. נבחן זאת באמצעות רגרסיה לינארית עפ"י שיטת "ריבועים פחותים".

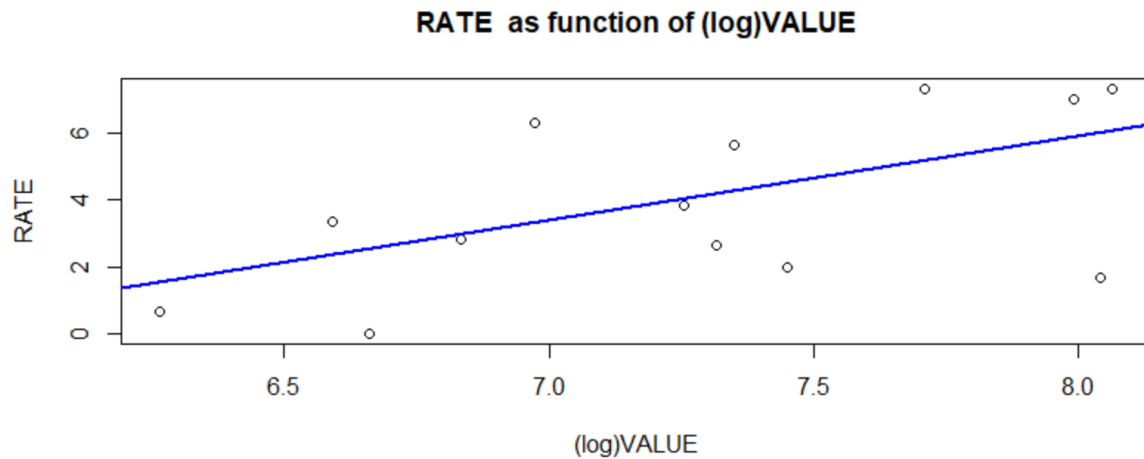
השערות:

- **השערת האפס – לא קיים קשר לינארי בין המשתנים, כלומר $H_0: b = 0$**
- **השערת המחקר – קיים קשר לינארי בין המשתנים, כלומר $H_1: b \neq 0$**

מהלך המבחן – המבחן יבוצע באופן זהה למבחן שבוצע בגוף העבודה, ועל כן לפירוט נוסף אודות סכמת המבחן הכללית וההנחות הרלוונטיות ראה תת סעיף "מבחן רגרסיה לינארית" תחת פרק "שיטות".

ממצאי המבחן:

- להלן גרף המתאר את הרגרסיה הלינארית RATE כפונקציה של $\log \text{VALUE}$:

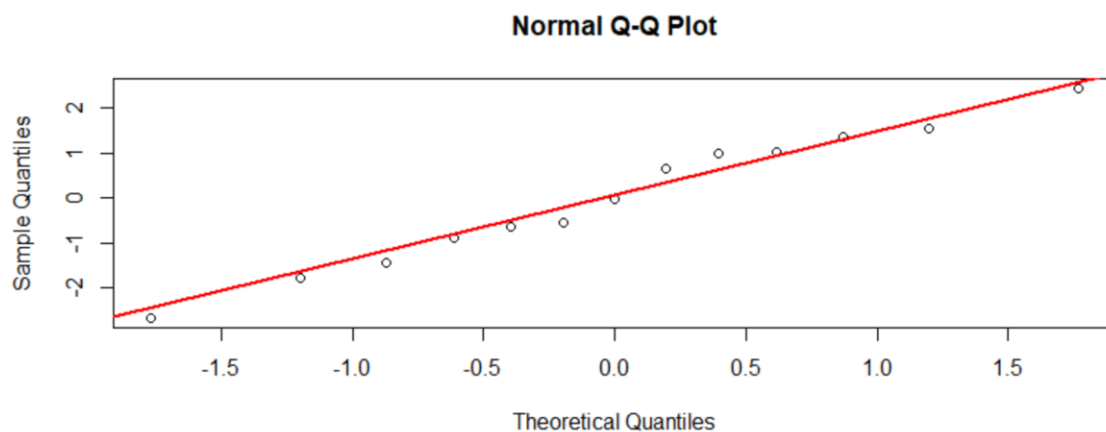


• ממצאי הרצת הפקודה : `summary(lm(RATE~log(VALUE)))`

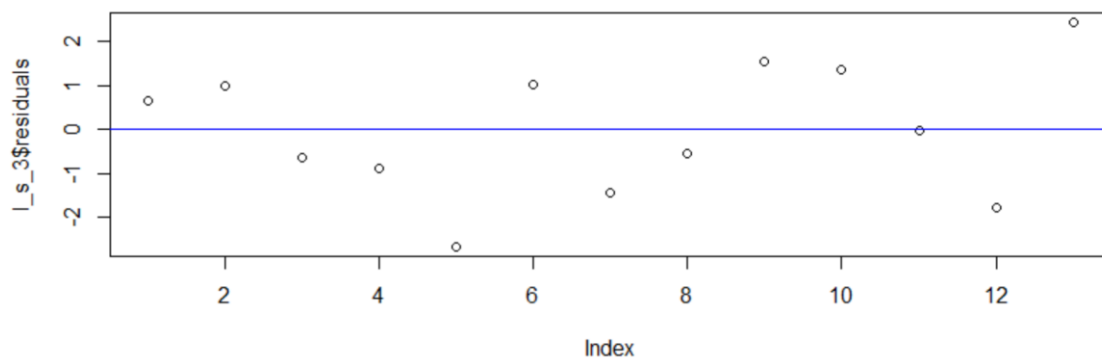
1. משוואת הרגרסיה הלינארית – $RATE = 2.533 * \log(VALUE) - 14.334$

a. $\hat{b}_{MLE} = 2.533 > 0$

2. נראה כי השאריות בין הערכים שקיבלנו לערכים הצפויים מתפלגות נורמלית באמצעות גרף qqplot:



3. בנוסף, נראה כי לא ניתן לזהות מגמה כלשהי בגרף הפיזור של השאריות שקיבלנו, דבר המחזק את ההשערה כי אכן מדובר בקשר לינארי בין המשתנים:



4. $SE = 0.768, T = 3.299 > t_{12,0.95} = 1.782, P_{value} = 0.00709 < 0.05 = \alpha$

5. לכן, נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$ ונקבל את ההשערה האלטרנטיבית, לפיה קיים קשר לינארי בין שווי במיליוני יורו של קבוצות כדורגל להצלחתן בליגת האלופות. **בפרט, מכיוון שהסטטיסטי שלנו התברר כחיובי, יכולנו לקיים מבחן חד צדדי ימני ולקבל $\frac{P_{value}}{2}$ ולכן נוכל להסיק כי מדובר בקשר לינארי חיובי.** בנוסף, נבחין כי קיבלנו P_{value} קטן מאוד, מה שמעיד על חוזק הקשר הלינארי, ובכך מחזק את השערתנו.

נספח ח' – מבחן רגרסיה לינארית בין המשתנים PLAYERS SALARY (שכר שחקנים ממוצע בדולרים) ל-RATE:

נבחן האם קיים קשר לינארי בין ממוצע משתנה PLAYERS SALARY (שכר שחקנים ממוצע בדולרים) לבין ממוצע ה-RATE לאורך 6 העונות שנבחנו (2014/15 – 2019/20) ברמת מובהקות $\alpha = 0.05$. המדגם שלנו מורכב מ-14 קבוצות להן קיימים הנתונים הנ"ל. נבחן זאת באמצעות רגרסיה לינארית עפ"י שיטת "ריבועים פחותים".

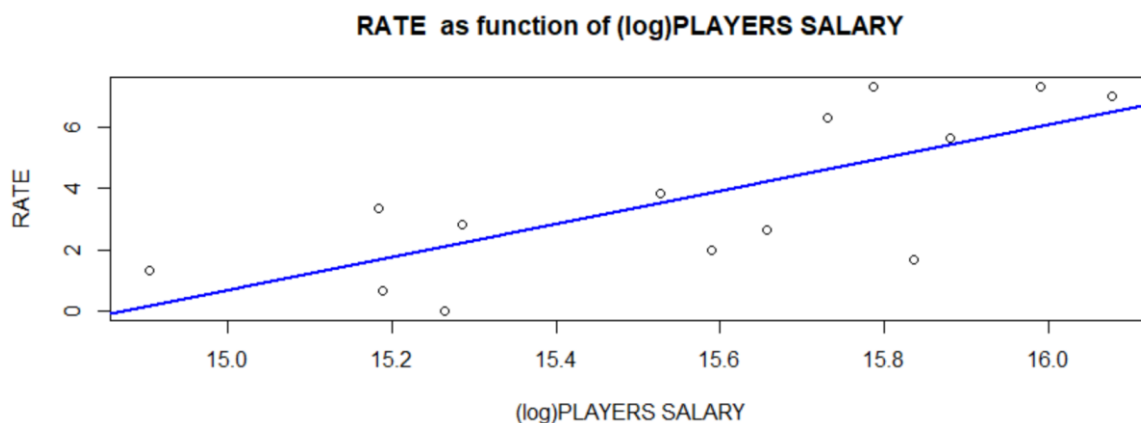
השערות:

- **השערת האפס** – לא קיים קשר לינארי בין המשתנים, כלומר $H_0: b = 0$
- **השערת המחקר** – קיים קשר לינארי בין המשתנים, כלומר $H_1: b \neq 0$

מהלך המבחן – המבחן יבוצע באופן זהה למבחן שבוצע בגוף העבודה, ועל כן לפירוט נוסף אודות סכמת המבחן הכללית וההנחות הרלוונטיות ראה תת סעיף "מבחן רגרסיה לינארית" תחת פרק "שיטות".

ממצאי המבחן:

- להלן גרף המתאר את הרגרסיה הלינארית RATE כפונקציה של $\log \text{PLAYERS SALARY}$:

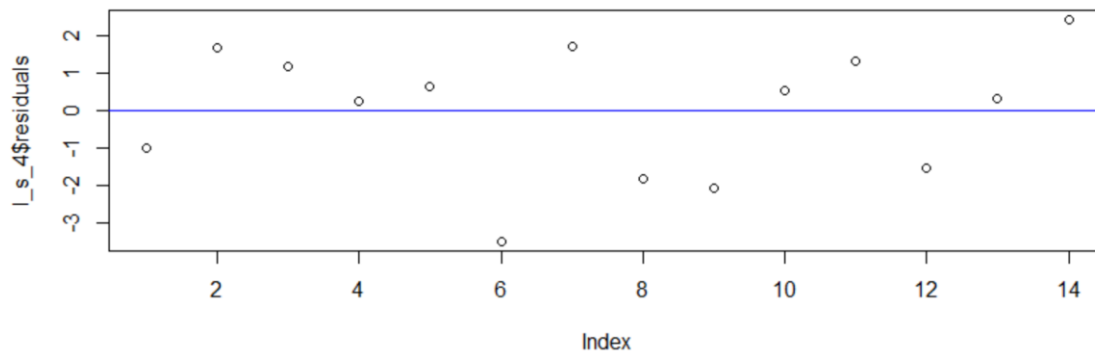


- ממצאי הרצת הפקודה `summary(lm(RATE~logPLAYERS SALARY))`:

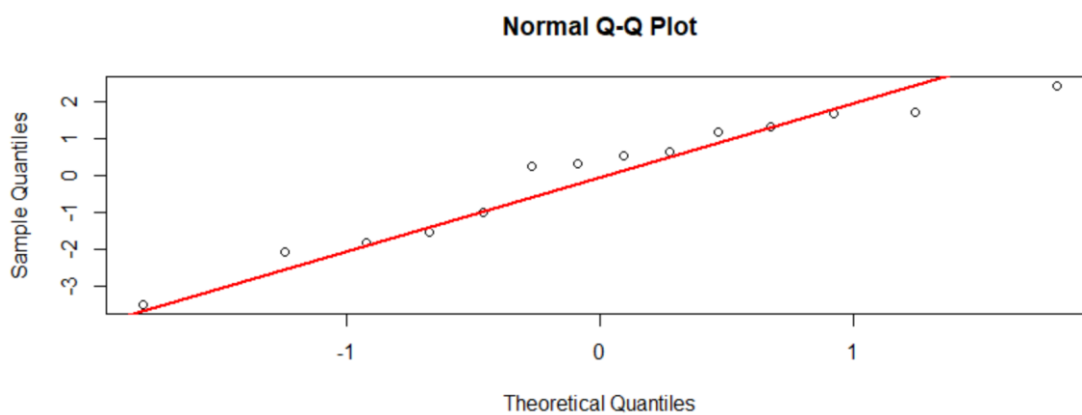
1. משוואת הרגרסיה הלינארית – $RATE = 5.390 * \log(PLAYERS SALARY) - 80.183$

$$a. \hat{b}_{MLE} = 5.390 > 0$$

2. נראה כי השאריות בין הערכים שקיבלנו לערכים הצפויים מתפלגות נורמלית באמצעות גרף qqplot:



3. בנוסף, נראה כי לא ניתן לזהות מגמה כלשהי בגרף הפיזור של השאריות שקיבלנו, דבר המחזק את ההשערה כי אין מדובר בקשר לינארי בין המשתנים:



4. $SE = 1.431, T = 3.768 > t_{12,0.95} = 1.782, P_{value} = 0.00268 < 0.05 = \alpha$

5. לכן, נדחה את השערת האפס ברמת מובהקות $\alpha = 0.05$ ונקבל את ההשערה האלטרנטיבית, לפיה קיים קשר לינארי בין גובה שכר שחקנים של קבוצות כדורגל להצלחתן בליגת האלופות. **בפרט, מכיוון שהסטטיסטי שלנו התברר כחיובי, יכולנו לקיים מבחן חד צדדי ימני ולקבל $\frac{P_{value}}{2}$ ולכן נוכל להסיק כי מדובר בקשר לינארי חיובי.** בנוסף, נבחין כי קיבלנו P_{value} קטן מאוד, מה שמעיד על חוזק הקשר הלינארי, ובכך מחזק את השערתנו.

נספח ט' – גרף המתאר את פרופורציות הנקודות של כל רמת עושר לאורך 11 העונות:

