

תרגיל: מגלשת הרים

מה נלמד בתרגיל הזה?

- איך להוסיף קווי שגיאה לתרשים באקסל (סימון של סטיית/טעות תקן)
- איך להוסיף קו מגמה על תרשים פיזור
- איך לחשב מתאם בעזרת אקסל

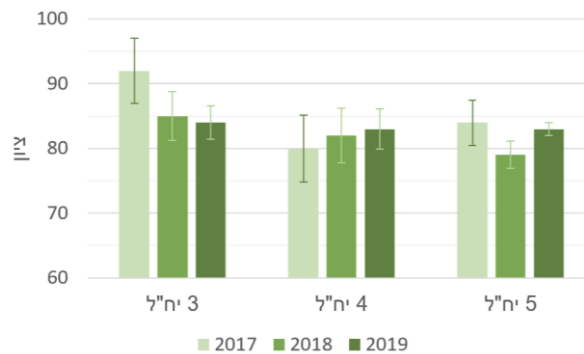
לפני התרגיל

צפו [בסרטון הזה](#).

קווי שגיאה

קווי שגיאה הם הקווים שמסמנים, על תרשים קיים, את סטיית התקן של הנתונים.

לדוגמה, בקובץ charts.xlsx בגיליון English-chart, העמודות מציינות את ממוצע הציונים בכל מקצוע בכל שנה, והקווים האנכיים בראש כל עמודה מייצגים את סטיית התקן של כל קבוצת נתונים. אורך הקו מייצג את סטיית התקן; מקובל שהקו נמשך סטיית תקן אחת מעל ומתחת ראש העמודה.

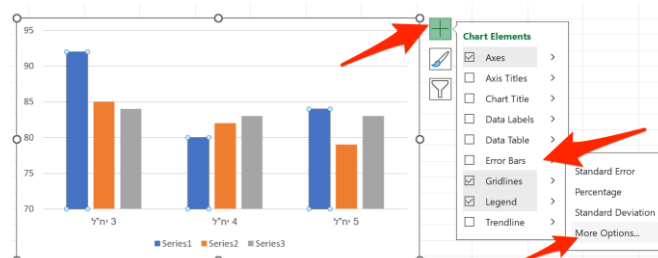


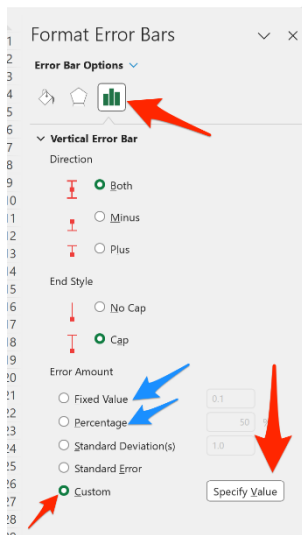
בדוגמה זו, קווי השגיאה מציגים את סטיית התקן (standard deviation). בדוגמה שריקרו הציגו בסרטון, מוצגת טעות התקן (standard error). הקשר בין השניים נקבע ע"י הנוסחה הבאה, בה n הוא כמות הדגימות בכל קבוצת נתונים (במקרה זה – כמות התלמידים):

$$standard\ error = \frac{standard\ deviation}{\sqrt{n}}$$

היתרון של טעות תקן הוא שברוב המקרים היא משקפת באופן ישיר יותר את ה-p-value. למשל, אם אנחנו רוצים לבדוק האם ההבדל בין שתי קבוצות מובהק, הגודל של טעות התקן משקף ישירות את המובהקות הזאת.

1. צרו גרף עמודות רגיל שמציג את הנתונים (פתרון השאלות בסוף הקובץ).
2. את קווי השגיאה לא מוסיפים בבת אחת, אלא לכל סדרה בנפרד. נתחיל בסדרת הנתונים של 2017. סמנו את היסוד, ובעזרת כפתור "+" בחרו באפשרות "קווי שגיאה" ובתוכה ב"אפשרויות נוספות":

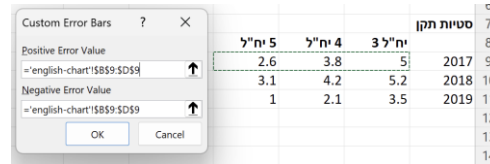




שימו לב: כמו רוב האלמנטים בתרשימים, גם את קווי השגיאה אפשר להוסיף בכמה דרכים. למשל דרך כפתור "Add chart elements" בסרגל התרשים. כרגיל, כל שיטות ההוספה יובילו אותנו בסופו של דבר לאותו מקום – במקרה זה, לחלונית עיצוב התרשים.

3. בחלונית עיצוב התרשים, נבחר לפי מה ייקבע אורכם של קווי השגיאה. יש פה כמה אפשרויות – למשל לכתוב בפירוש מה אורך הקו ביחידות ציר Y או באחוזים מתוך גובה העמודה (חצים כחולים), או – השיטה בה נבחר עכשיו – לקחת מתוך קובץ האקסל עצמו את אורך הקו עבור כל קטגוריה בנפרד. לשם כך, נבחר באפשרות custom, ונלחץ על כפתור specify values כדי לבחור את הערכים הספציפיים (חצים אדומים):

במסך שייפתח, עלינו לציין מה יהיה אורך קווי השגיאה. יש לציין בנפרד מה ארכם לכל כיוון – מעל ראש העמודה (כיוון חיובי) ומתחת ראש העמודה (שלילי).



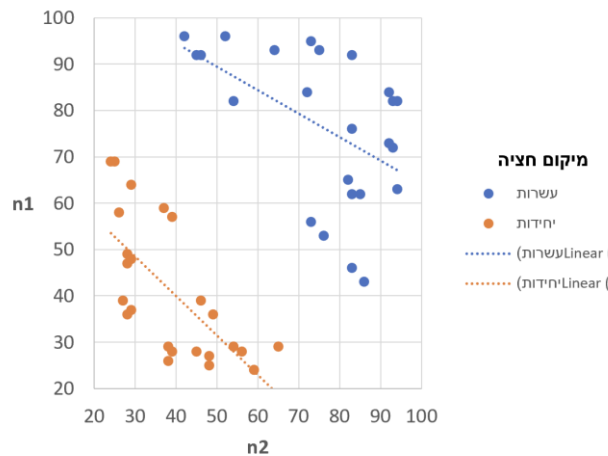
במקרה זה, ציירנו קווים באורך זהה לשני הכיוונים. זה מה שעושים בדרך כלל. אבל זה לא הכרחי. למשל, לפעמים מצירים בתרשים רק את הקו העליון, ללא הקו התחתון (אורך 0):

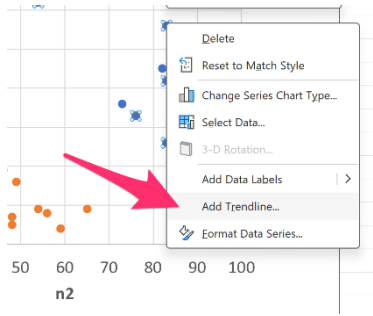
4. כמובן שאפשר לעצב מבחינה גרפית את קו השגיאה עצמו כמו שמעצבים כל קו אחר בתרשימים – צבע, עובי וכו'. נסו זאת!

קווי מגמה

קווי מגמה מיועדים לתרשים פיזור, והם מראים מה הקשר בין המשתנה בציר X למשתנה בציר Y. סטטיסטית, קו המגמה הוא בעצם הקו שמציג את נוסחת הרגרסיה הליניארית בין שני המשתנים. מידת ההתאמה בין הקו לבין הנתונים ("עד כמה הנקודות מסתדרות על הקו") משקפת את המתאם בין שני המשתנים. השיפוע של קו המגמה משקף את המקדם של X בנוסחת הרגרסיה (נהוג לסמן אותו באות b).

לדוגמה, כך נראים קווי המגמה בגיליון calc-stimuli-chart:





איך נוסיף קווי מגמה?

הקליקו עם הכפתור הימני של העכבר על אחת הנקודות בגרף, ובחרו באפשרות של "הוספת קו מגמה":

יקרו שני דברים: ראשית, קו המגמה יתווסף מייד. שנית, בצד המסך תיפתח חלונית עיצוב תרשים עם ההגדרות של קו המגמה, שם תוכלו לקבוע כל מיני מאפיינים שלו:

- האם הקו יתאר את הקשר הליניארי (ישר) בין שני המשתנים, או קשר אחר (אקספוננציאלי, לוגריתמי וכו' – נסו זאת!)
- אפשר להציג את הנוסחה של קו המגמה (בפועל זו נוסחת הרגרסיה), ואת מידת ההתאמה שלו לנתונים (R^2).
- וכמובן כרגיל, אפשר לקבוע את המאפיינים היוזואליים של הקו.

באופן דומה, נוסיף קו מגמה לסדרה השנייה.

חישוב מתאם

אם בחרנו קו מגמה ליניארי, הקו משקף את תוצאת הרגרסיה הליניארית של משתנה Y לפי משתנה X. אנחנו יודעים שיש קשר הדוק בין רגרסיה למתאם – למעשה, ערך R^2 של הרגרסיה (אקסל מאפשר להציג אותו ליד קו המגמה) הוא הריבוע של מתאם פירסון בין שני המשתנים.

אקסל מאפשר לנו לחשב את המתאם בין שני המשתנים גם קשר לתרשים. עושים זאת בעזרת הפונקציה CORREL, ויש לה שני ארגומנטים: סדרת הנתונים של המשתתף הראשון, וסדרת הנתונים של המשתתף השני.

לדוגמה, כדי לחשב את מתאם פירסון בין שני המשתנים בקובץ שלנו (n_1, n_2) במקרה של חציית עשרת בעשרות, כתבו את הנוסחה:

$$=CORREL(A2:A25, B2:B25)$$

5. חשבו גם את המתאם בין שני המשתנים במקרה של חציית עשרת ביחידות.

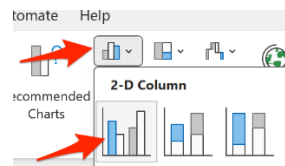
שימו לב: הפונקציה CORREL מחזירה את המתאם (כלומר ערך r), לא את המובהקות שלו.

למרבה הצער, אין לאקסל פונקציה לחישוב מובהקות המתאם. אבל אפשר לכתוב בקלות נוסחה שמחשבת את זה. אם בגיליון מסוים רשמנו בתא A1 את המתאם (כלומר ערך r, כפי שהוחזר ע"י הפונקציה CORREL) ובתא A2 את גודל המדגם (כמה דגימות היו לכל אחד מהמשתנים – לא לשניהם ביחד!), הנוסחה הבאה תחזיר את ערך p-value דו-זנבי של המתאם:

$$=T.DIST.2T(A1 * SQRT(A2-2) / SQRT(1-A1^2), A2-2)$$

פתרונות השאלות

1. נסמן את הנתונים הרלוונטיים, ובסרגל "הוספה" נלחץ על הכפתור של הוספת תרשים עמודות:



E	D	C	B	A	
	5 יח"ל	4 יח"ל	3 יח"ל		1
אחוז נעדר	84	80	92	2017	2
5%	79	82	85	2018	3
15%	83	83	84	2019	4
7%					5

5. $CORREL(A26:A49,B26:B49)$