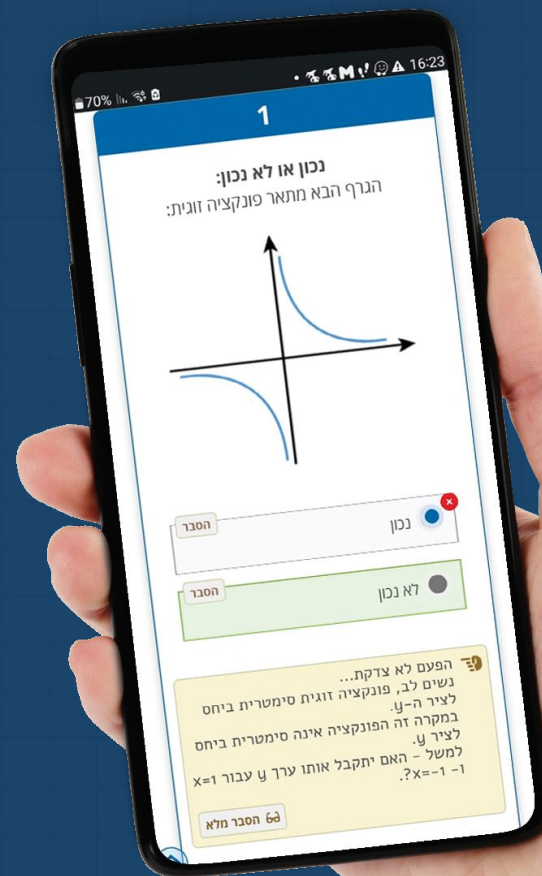


חישוב מקדם המתאם הלינארי שיעורים 2+3

- ← מצגות הוראה
- ← יחידות תרגול
- ← חומרים תיאורטיים



חישוב מקדם המתאם הלינארי

שיעור 2+3 (שיעורים כפולים)



א'

מהי השתנות משותפת?

א'

מהי השתנות משותפת?

נניח, שמישהו שאנחנו לא מכירים הולך להיכנס לכיתה...

בהיעדר מידע ננבא כמובן את הממוצע.

איזה גובה ננבא לו?

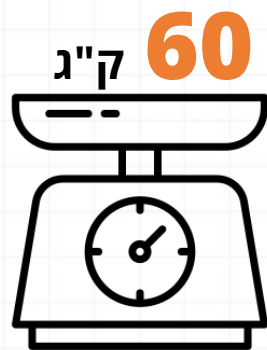
נמוך?

גבוה?



עכשיו נניח, ששתלנו משקל בכניסה לכיתה...

משקל גבוה מהמוצע
"הולך עם"
גובה גבוה מהמוצע...
וכמובן גם להיפך!



איזה גובה ננבא לו?

נמוך?

גבוה?





ב'

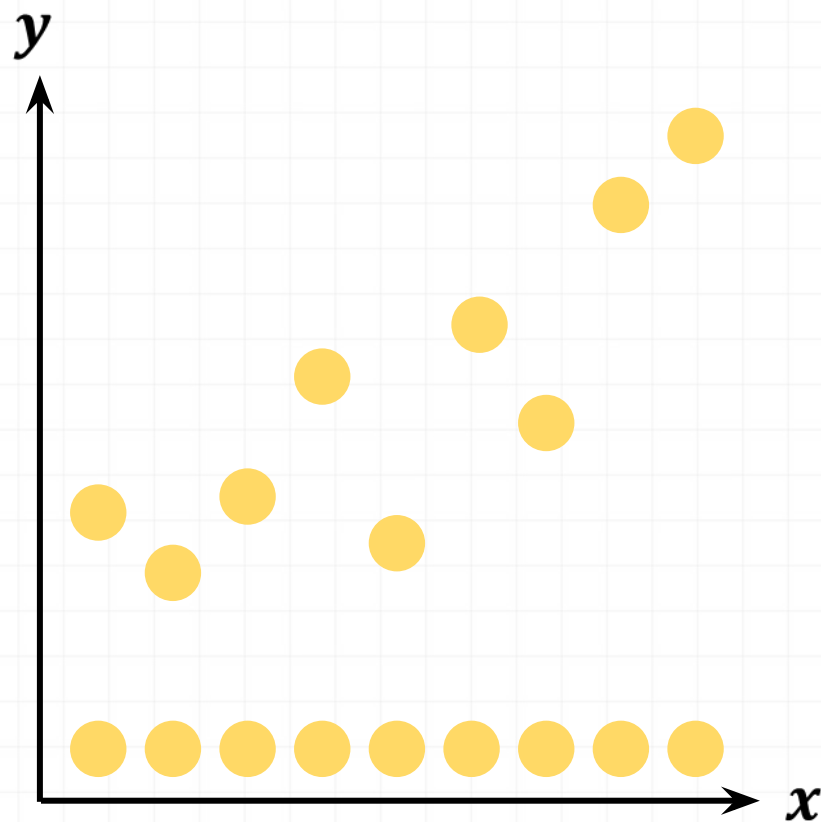
השתנות משותפת בגרף



ב'

השתנות משותפת בגרף

השתנות משותפת בגרף



עכשיו נניח שאותן תצפיות
נמדדו גם במשתנה y

אם נבחן אותן מלמטה למעלה,
למעשה אנחנו מתייחסים
להשתנות של המשתנה y

נתבונן על מספר תצפיות
שנמדדו במשתנה x

אם נבחן אותן משמאל לימין,
למעשה אנחנו מתייחסים
להשתנות של המשתנה x

זו בדיוק ההשתנות המשותפת
של x עם y



שלב ג'

השתנות משותפת - מהגרף לנוסחה



שלב ג'

השתנות משותפת - מהגרף לנוסחה

איך נוכל לחשב מתמטית את מידת התיאום בין שני משתנים?

תצפיות גדולות מהממוצע ב- x נוטות "ללכת עם" תצפיות גדולות מהממוצע ב- y (ולהיפך)

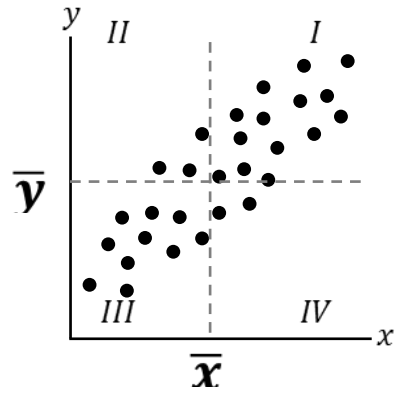
$$\sum (x_i - \bar{x})(y_i - \bar{y})$$

ניקח כל תצפית ונמדוד את מרחק ה- x שלה מהממוצע

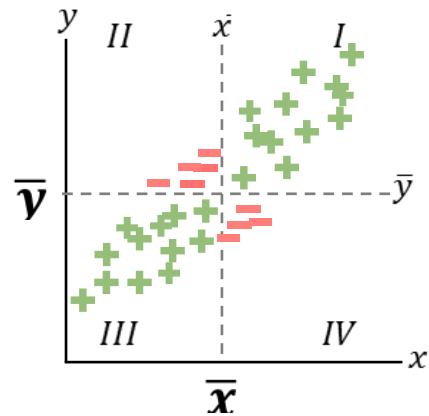
ועבור אותה התצפית, נמדוד את מרחק ה- y שלה מהממוצע

נכפיל את שני הביטויים ונסכום מעבר לכל התצפיות

מה קורה כשהקשר חיובי?



על-פי רוב, ערכים הגבוהים מ- \bar{x} יהיו מתואמים עם ערכים הגבוהים מ- \bar{y}
 על-פי רוב, ערכים הנמוכים מ- \bar{x} יהיו מתואמים עם ערכים הנמוכים מ- \bar{y}

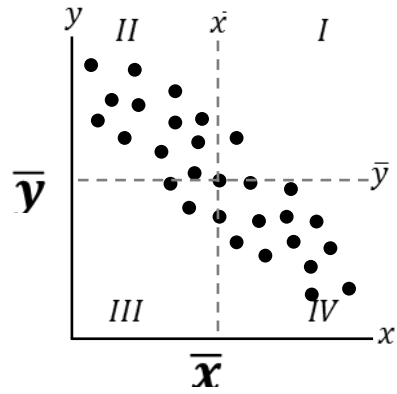


$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) \rightarrow +$$

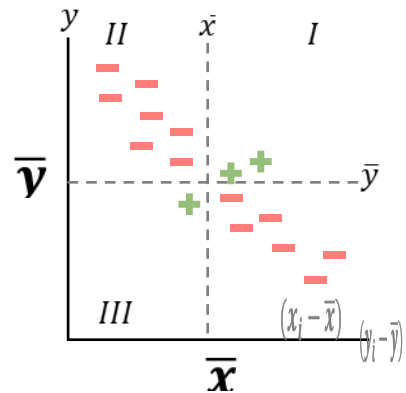
+	+	→	+
-	-	→	+
-	+	→	+

אם התצפיות מתואמות חיובית המכפלות שנחשב תסכמנה לערך חיובי כלשהו

מה קורה כשהקשר שלילי?



על-פי רוב, ערכים הנמוכים מ- \bar{x} יהיו מתואמים עם ערכים הגבוהים מ- \bar{y}
 על-פי רוב, ערכים הגבוהים מ- \bar{x} יהיו מתואמים עם ערכים הנמוכים מ- \bar{y}

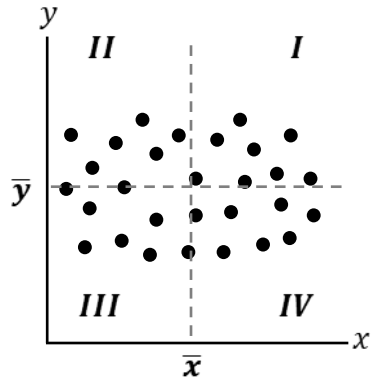


$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) \rightarrow -$$

$$\begin{array}{ccc} - & + & \rightarrow - \\ + & - & \rightarrow - \end{array}$$

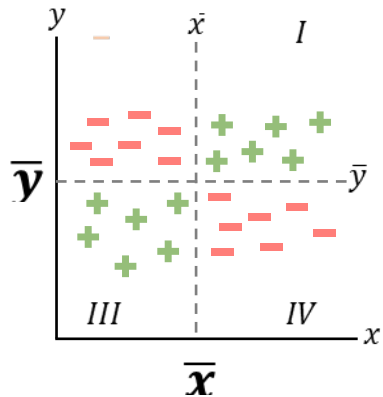
אם התצפיות מתואמות שלילית המכפלות שנחשב תסכמה לערך שלילי כלשהו

מה קורה כשאין קשר?



ערכים הגבוהים מ- \bar{x} לעתים יהיו מתואמים עם ערכים הגבוהים מ- \bar{y}
 אבל לעיתים יהיו מתואמים עם ערכים הנמוכים מ- \bar{y}

ערכים הנמוכים מ- \bar{x} לעתים יהיו מתואמים עם ערכים הנמוכים מ- \bar{y}
 ולעתים יהיו מתואמים עם ערכים הגבוהים מ- \bar{y}



$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) \rightarrow 0$$

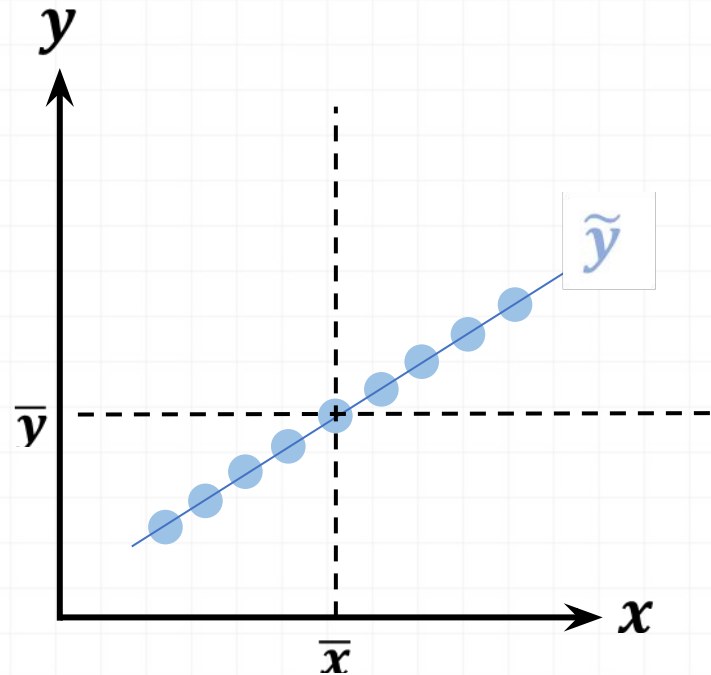
+	+	→	+
+	-	→	-
-	-	→	+
-	+	→	-

אם התצפיות לא מתואמות המכפלות יתקזזו ויתקבל ערך שואף לאפס

אז יש לנו מדד מתמטי לגבי כיוון הקשר!!

$$\Sigma(x_i - \bar{x})(y_i - \bar{y})$$

אבל מה לגבי עוצמת הקשר?

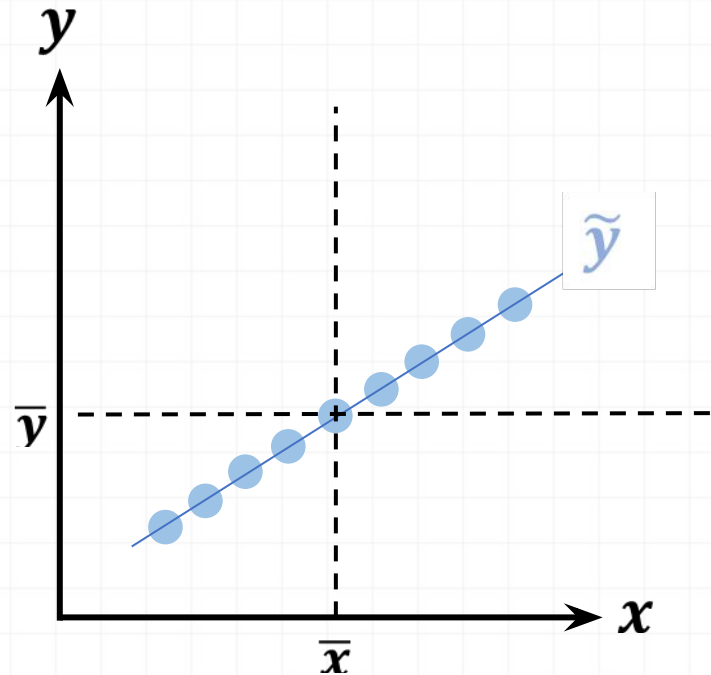


הבעיה היא שהמדד שלנו מושפע מגודל המדגם...

אז יש לנו מדד מתמטי לגבי כיוון הקשר!!

$$\Sigma(x_i - \bar{x})(y_i - \bar{y})$$

אבל מה לגבי עוצמת הקשר?



$$\text{Covariance}_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n}$$



שלב ד'

מהשתנות משותפת - למקדם המתאם



שלב ד'

מהשתנות משותפת - למקדם המתאם

השתנות משותפת

$$COV_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n}$$

COV_{xy} מצביע על **כיוון** הקשר, אבל הוא לא אינדיקטור יעיל לבירור **עוצמת** הקשר

הבעיה היא שהיא "רגיש" לסקאלת הערכים - ליחידות המידה של התצפיות

אם נחשב את COV_{xy} בין
גובה בס"מ לבין משקל בגרמים

אותו אדם ייצר את המכפלה הבאה:

$$(180 - 175)(80,000 - 75,000) = 25,000$$

אם נחשב את COV_{xy} בין
גובה במטרים לבין משקל בק"ג

אדם בגובה 1.8 ובמשקל 80 ייצר את
המכפלה הבאה:

$$(1.8 - 1.75)(80 - 75) = 0.25$$

וכל זאת כאשר מדובר באותה תצפית ובאותו קשר!

מקדם המתאם של פירסון

$$COV_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n}$$

אם המדד רגיש לסטיות התקן של המשתנים, אז כדי לבטל רגישות זו יש לחלק את המדד בסטיות התקן של שני המשתנים

$$r_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x \cdot S_y}$$

המדד המתוקנן הוא מתאם פירסון

$$-1 \leq r_{xy} \leq 1$$

מקדם המתאם של פירסון

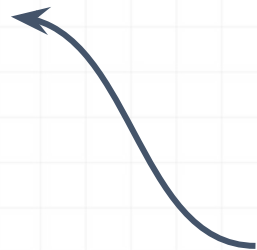
$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x \cdot S_y}$$

נבליט בו שני חלקים:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x \cdot S_y}$$

$$r_{xy} = \frac{\sum Z_{x_i} \cdot Z_{y_i}}{n}$$

ואכן ראינו שבקשר לינארי
נבדק התיאום בין מידת
ההתרחקות מהמוצע של
התצפיות במשתנה x לבין
מידת **ההתרחקות**
מהמוצע של התצפיות
במשתנה y





Classsit

סוף שיעורים 2+3 :-)

