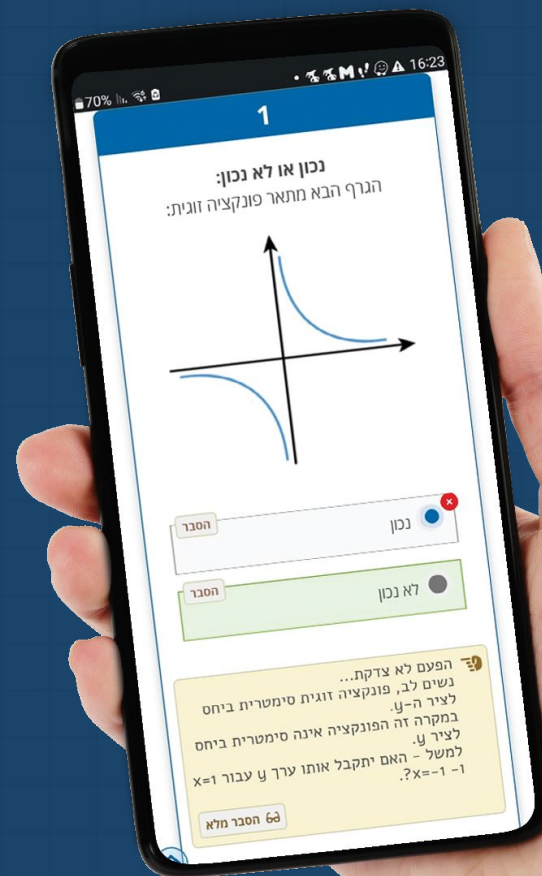




Classsit

שיעור 6

טרנספורמציות על קו הרגרסיה



מבנה ההוראה - רגרסיה



- | | |
|--------------|------------------------|
| (שעות 1-2) | 1. קשר בדיאגרמות פיזור |
| (שעות 3-4) | 2. חישוב מקדם המתאם א' |
| (שעות 5-6) | 3. חישוב מקדם המתאם ב' |
| (שעות 7-8) | 4. קשר בטבלאות נתונים |
| (שעות 9-10) | 5. קו הרגרסיה א' |
| (שעות 11-12) | 6. קו הרגרסיה ב' |
| (שעות 13-14) | 7. טרנספורמציות |
| (שעות 15-16) | 8. חזרות |

תזכורת לקראת טרנספורמציות : שונות וסטית תקן



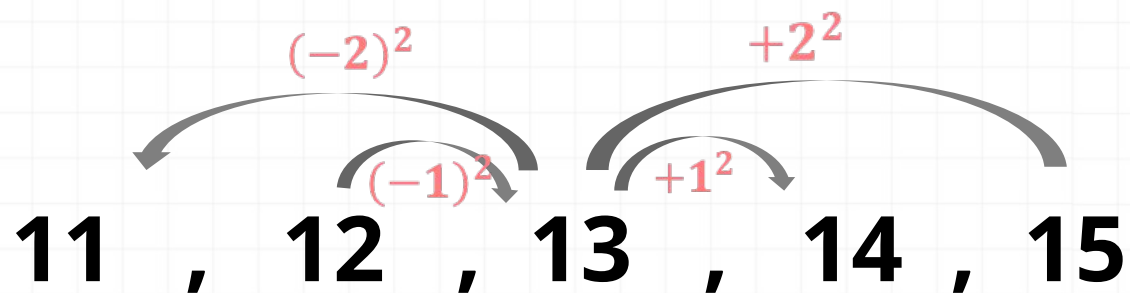
שונות וסטית תקן

שונות היא סכום הסטיות הריבועיות מהמוצע, חלקי מספר האיברים.

$$\frac{\sum(x_i - \bar{x})^2}{n}$$

חשוב לשים לב כי בחישוב השונות יש סכום במונה, חלקי מספר האיברים במכנה
זה מזכיר חישוב של ממוצע והשונות היא אכן מדד ממוצע

השונות היא הסטיה הריבועית הממוצעת מהמוצע.



שונות וסטית תקן

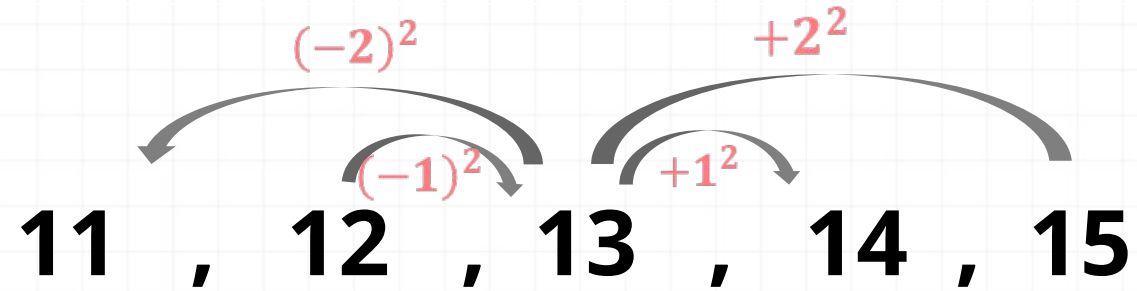
השונות היא מדד מנופח, כי היא מעלה את המרחקים מהממוצע בריבוע, ונותנת משקל לא-פרופורציונלי לערכים קיצוניים.

מדד השונות "הפחות-מנופח" הוא סטית התקן:

$$\sqrt{\text{שונות}} = \text{סטיית התקן}$$

סטית התקן היא האומדן הנפוץ ביותר להערכת הפיזור היא שורש של הסטיה הריבועית הממוצעת.

איך מחושבת סטיית התקן?



השונות ($SS = 2$) היא מדד
מנופח ביחס לסטיות מוחלטות.

$$s^2 = \frac{1^2 + (-1)^2 + 2^2 + (-2)^2}{5} = \frac{10}{5} = 2$$

סטית התקן ($S = 1.4$) פחות
מנופחת מהשונות, אך עדיין
גדולה מהסטיות המוחלטות.

$$s = \sqrt{s^2} = \sqrt{2} = 1.4$$

תזכורת נוספת - טרנספורמציות

הוספת/הפחתת קבוע לכל הערכים

תוכנית אימונים שנתית בניו זילנד לריצת מרתון אולימפית בדקה את משקל המשתתפים בתחילת התוכנית ובסופה. נמצא שבתחילת התוכנית ממוצע משקלם של המשתתפים היה 68 קילוגרם, ובמהלך התוכנית כל אחד מהמשתתפים הפחית בדיוק 4 ק"ג ממשקלו.

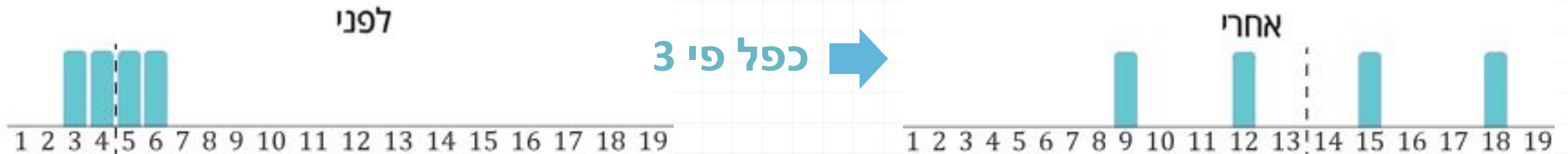


כל אחד מערכי ההתפלגות ירד ב-4 ק"ג, וכך גם כל מדדי המרכז!
אבל: הוספת או הפחתת קבוע לכל הערכים לא תשפיע על הפיזור!

תזכורת נוספת - טרנספורמציות

הכפלה (או חלוקה) בקבוע על כל הערכים

אחד המשתתפים בתוכנית האימונים בניו זילנד רצה להמיר את מחירי הירקות שקנה לשקלים. מחירי הירקות היו 3,4,5,6 דולר ניו זילנדי, וידוע שדולר ניו זילנדי שווה ל-3 שקלים חדשים. מה יהיה ממוצע מחירי הירקות בש"ח?

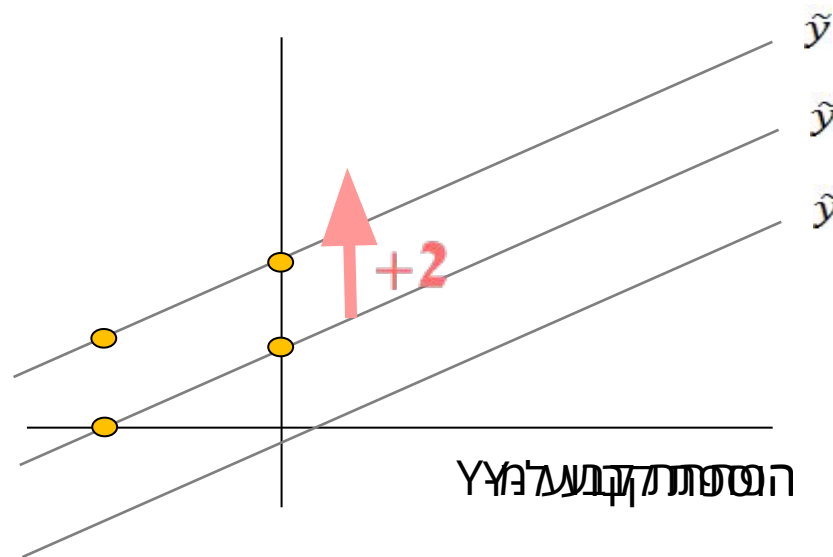
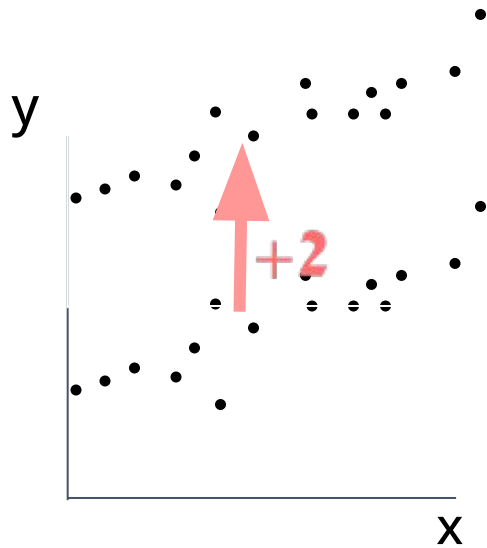


כל אחד מערכי ההתפלגות הוכפל פי 3, וכך גם כל מדדי המרכז
שימו לב: הכפלה (או חלוקה) של כל הערכים תשפיע גם על מדדי הפיזור!

טרנספורמציות על קו הרגרסיה

$$m = r_{xy} \cdot \frac{S_y}{S_x}$$

הוספה/הפחתה
על משתנה y בלבד



העיקר: r נשאר!

השיפוע אינו משתנה

השונות אינה משתנה

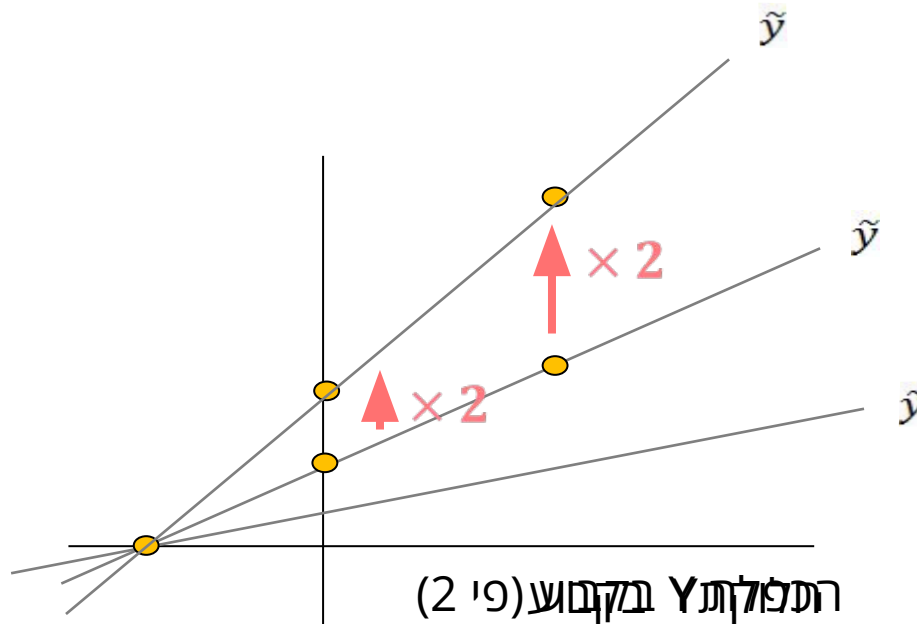
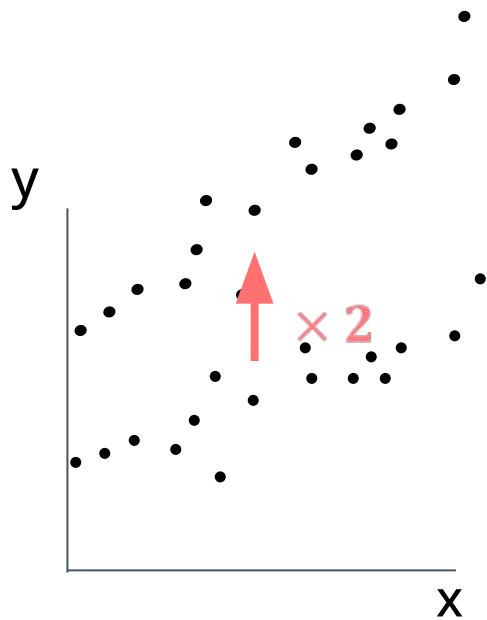
נקודות החיתוך עם הצירים משתנות

$$\tilde{y} = mx + b + 2$$

טרנספורמציות על קו הרגרסיה

הכפלה / חלוקה
על משתנה y בלבד

$$m = r_{xy} \cdot \frac{S_y}{S_x}$$



העיקר: r נשאר!

אבל: השיפוע משתנה

השוונות משתנה

נקודת החיתוך עם ציר y משתנה

נקודת חיתוך עם ציר x נשארת כ"עוגן"

$$m = r_{xy} \cdot \frac{2 \cdot S_y}{S_x}$$

סוף

סוף

סוף החומר ברגרסיה.....



Classsit

