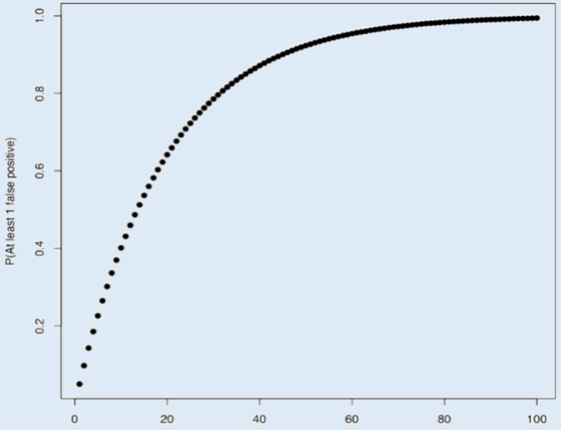
בשיעור שעבר...

ראינו שככל שעושים יותר מבחנים סטטיסטים, כך עולה הסיכוי לטעות מסוג 1 (false positive).

- אם למשל עשינו 40 מבחנים ברמת מובהקות של $\alpha = 0.05$, נצפה ל- $40 \cdot 0.05 = 2$ מבחנים שיצאו מובהקים <u>רק במקרה</u>
- בפועל, הסיכוי לפחות לטעות <u>אחת</u> מסוג I הוא $FWER = 1 - (1 - 0.05)^{40} = 0.87$



<u>דרכים שלמדנו עד כה ע"מ להתמודד עם FWER (או multiple hypothesis testing)</u>

1. **תיקון בונפרוני** – נקבע שהבדל יצא מובהק רק אם עבר סף חדש: $p \leq \frac{\alpha}{m}$ - כאשר m מספר המבחנים – כך ש: $FWER_{new} < \alpha$

   יתרון: כמות נמוכה יותר של טעויות מסוג I. חסרון: מעלה את הסיכוי לטעות מסוג II.

2. **פרמוטציות** – שימוש בנתונים עצמם על מנת לחשב מובהקות – ערבול ה"תוויות" של כל דוגמא וחישוב כמה פעמים קיבלנו תוצאה

   שהיא מובהקת כמו או יותר מהתוצאה האורגנית שלנו.

False Discovery Rate (FDR)

FWER הוא ההסתברות לעשות לפחות טעות אחת מסוג 1 כאשר עושים מספר מבחנים:

$$FWER = 1 - (1-\alpha)^{\#tests}$$

FDR הוא כמות הטעויות מסוג 1 ביחס לסך כל הגילויים (דחיות של H0):

$$\boxed{FDR = \frac{\#false\ discoveries}{\#discoveries} = \frac{False\ Positives}{True\ Positives + False\ Positives}}$$

נגדיר כל מקרה שבו נדחה את H0 בתור discovery ("תגלית") – ואז ה-FDR הוא יחס המקרים בהם עשינו תגלית שגויה מתוך כל מקרי ה-"תגליות".

FDR של 5% אומר ש-5% מכל המדדים <u>שיצאו מובהקים</u> (חיוביים) הם בעצם לא.

מציאות / החלטה	המצב האמיתי: שלילי	המצב האמיתי: חיובי
חיובי	טעות מסוג 1 (false positive) α	החלטה נכונה (true positive) $1-\beta$
שלילי	החלטה נכונה (true negative) $1-\alpha$	טעות מסוג 2 (false negative) β

איך מעריכים את ה-FDR?

נשתמש בפרמוטציות

נעריך את כמות השגיאות באמצעות הנתונים הגולמיים על ידי ערבול הקבוצות.

בשיעור הקודם השתמשנו בפרמוטציות כדי להעריך מובהקות של מבחן בודד – והפעם נשתמש בפרמוטציות עבור כל ההשוואות

<u>איך עושים את זה בפועל?</u>

1. מוצאים p-value לכל אחת מההשוואות (בחינות, בדוגמא של הריטלין)
2. מערבבים את הקבוצות ושוב מוצאים p-value לכל אחת מההשוואות
3. חוזרים על סעיף 2 הרבה פעמים
4. עבור רמת המובהקות שאנחנו רוצים, נבדוק כמה השוואות יצאו מובהקות הן בקבוצות המקוריות (בסעיף 1) והן בקבוצות המעורבבות (בסעיף 2).
5. מחשבים FDR:

$$FDR_{permutation} = \frac{mean\ (\#\ of\ significant\ comparisons\ in\ shuffled\ groups)}{\#\ significant\ comparisons\ in\ original\ group}$$

איך מעריכים את ה-FDR?

נשתמש בפרמוטציות

כאן יהיו False positives

$$FDR_{permutation} = \frac{mean\ (\#\ of\ significant\ comparisons\ in\ shuffled\ groups)}{\#\ significant\ comparisons\ in\ original\ group}$$

כאן יהיו False positives וגם True positives

ככל שמספר הפרמוטציות יהיה גדול יותר, נקבל דיוק גבוה יותר בחישוב ה-FDR.

שימו לב: יש לזכור לבצע פרמוטציות על המשתנה הידוע.

איך עושים את זה ב-R?

<u>חלק א'</u>

מטרה: מציאת כמה מבחנים הם שונים בין סטודנטים הנוטלים ריטלין לאלו שלא

```r
1  load('ritalin_data.Rdata')
2  alpha = 0.05
3
4  # Original Data & statistics
5  # -------------------------------------------
6  yesRitalin = Data[,1:39]
7  noRitalin = Data[,40:78]
8
9  pValuesOriginal <- numeric(dim(Data)[1])
10
11 for (i in 1:40) {
12   pValuesOriginal[i] <- t.test(yesRitalin[i,],noRitalin[i,])$p.value
13 }
14
15 length(which(pValuesOriginal < alpha))
```

שורה	מה עושים בשורה?
1	טוענים את מטריצת הנתונים בעזרת פונקציית load לתוך אובייקט שקוראים לו Data
2	קביעת ה-alpha הרצוי לנו
6-7	הגדרת קבוצות הסטודנטים – קבוצת yesRitalin (מטופלים בריטלין) וקבוצת noRitalin (לא מטופלים בריטלין)
9	יצירת וקטור מספרי (בפונקצייה numeric) באורך של מספר השורות של Data (כלומר – מספר הבחינות)
11	יצירת לולאת for שנותנת למשתנה i ערך עולה בין 1 ל-40 בכל איטרציה (כלומר עבור כל בחינה)
12	הכנסה לתוך וקטור במיקום i את ה-p value המתקבל מה-t test בין שתי הקבוצות. הוקטור הסופי נותן לנו את ה-p value המתקבל לכל בחינה בנפרד.
15	ספירה של כמות הבחינות שעומדות בתנאי הסף של alpha

איך עושים את זה ב-R?

```
17 # Permutations
18 # -------------------------------------------
19 nPermutations = 100
20 
21 pValuesShuffled <- matrix(data = 0, nrow = dim(Data)[1], ncol = nPermutations)
22 
23 for (j in 1:nPermutations) {
24   permute <- sample(78)
25   yesRitalinShuffled <- permute[1:39]
26   noRitalinShuffled <- permute[40:78]
27 
28   for (i in 1:40) {
29     pValuesShuffled[i,j] <- t.test(x = Data[i,yesRitalinShuffled],
30                                    y = Data[i,noRitalinShuffled])$p.value
31   }
32 }
```

<u>חלק ב'</u>

מטרה: עריכת פרמוטציות (ערבולים) על המשתנה "ריטלין" (כלומר על האם הסטודנט נוטל ריטלין או לא) וחישוב מספר הבחינות בהם יש הבדלים בין הקבוצות בנתונים המעורבלים.

שורה	מה עושים בשורה?
19	קביעת מספר הערבולים
21	יצירת מטריצה המכילה אפסים (ערך התחלתי) בגודל מספר הבחינות X מספר הפרמוטציות. המטריצה תכיל את ה-p values של הערכים המעורבלים עבור כל בחינה.
23	יצירת לולאת for שנותנת למשתנה i ערך עולה בין 1 למספר הפרמוטציות בכל איטרציה (כלומר עבור כל פרמוטציה)
24	יצירת הפרמוטציה על המספרים 1:78 כלומר – האינדקס של הסטודנט – הסטודנטים כעת מסודרים בסדר אקראי
25-26	שיבוץ כל סטודנט בקבוצה חדשה לפי מיקומו בוקטור permute. שימו לב שגודל הקבוצות נשאר זהה.
28	יצירת לולאת for שנותנת למשתנה i ערך עולה בין 1 ל-40 בכל איטרציה (כלומר עבור כל בחינה)
29-30	הכנסה לתוך המטריצה במיקום i,j (שורות = בחינה, עמודה = פרמוטציה) את ה-p value המתקבל מה-t test בין שתי הקבוצות המעורבלות.

איך עושים את זה ב-R?

```r
38 # Computing FDR for different alphas
39 # ------------------------------------------
40 alphas <- seq(0.005,0.3,0.005)
41 FDR <- numeric(length(alphas))
42 sigOriginal<-numeric(length(alphas))
43
44 for (k in 1:length(alphas)) {
45   sigOriginal[k]<- sum(pValuesOriginal < alphas[k])
46   sigShuffled <- colSums(pValuesShuffled < alphas[k])
47   FDR[k]<-mean(sigShuffled)/sigOriginal[k]
48   FDR[k]<-max(FDR[1:k])  # avoid non-monotonic portions
49 }
50
51 plot(x = sigOriginal, y = FDR, type='b', xlab='# significant comparisons', ylab='FDR')
```

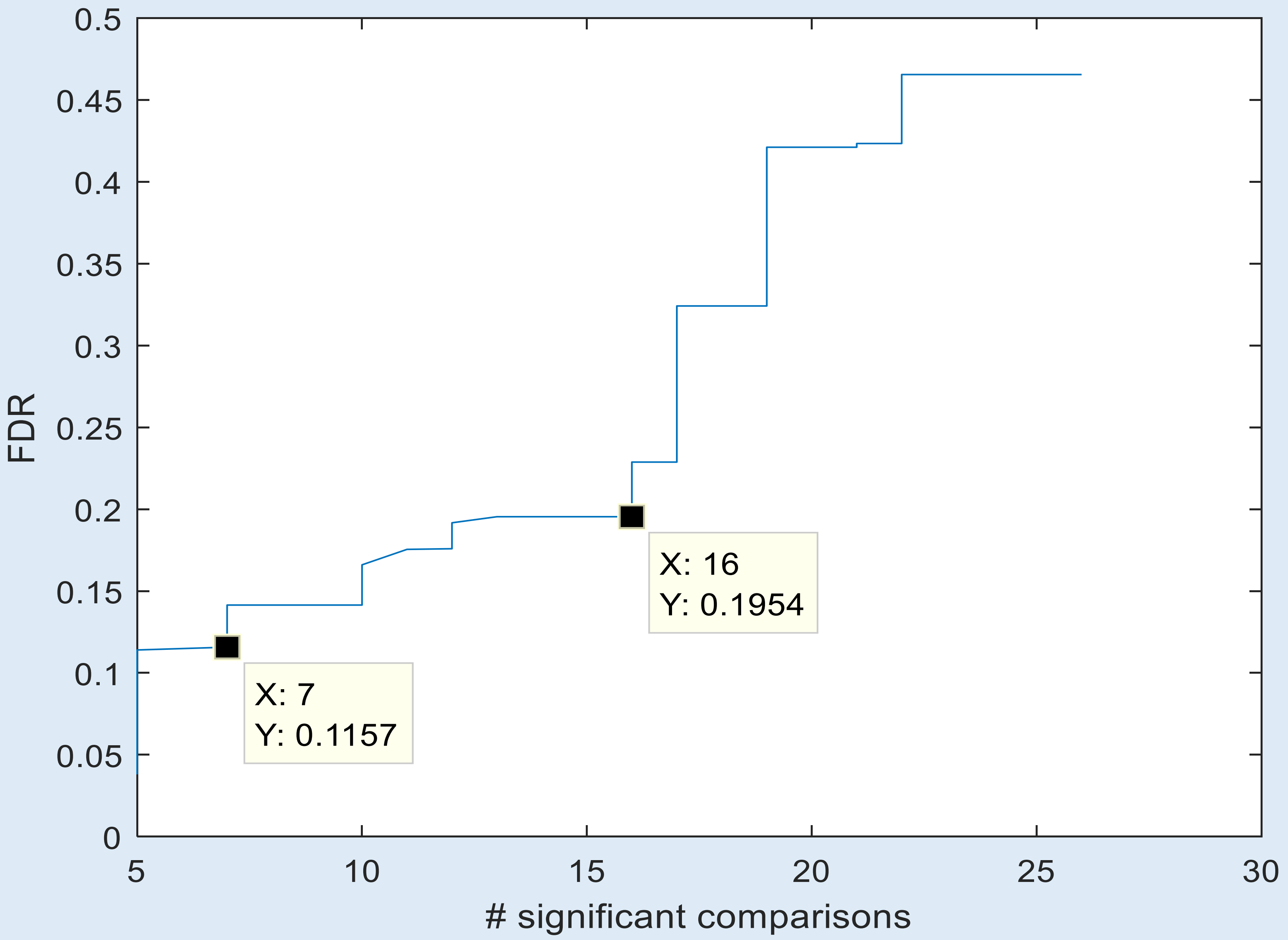
<u>חלק ג'</u>

מטרה: חישוב FDR ומציאת alpha המתאים לצרכינו

$$FDR_{perm} = \frac{mean\,(\#\ of\ significant\ comparisons\ in\ shuffled\ groups)}{\#\ significant\ comparisons\ in\ original\ group}$$

שורה	מה עושים בשורה?
40	יצירת משתנה alphas המכיל ערכים שונים מ-0.005 ועד 0.3 בקפיצות של 0.005. נבדוק את ה-FDR עבור כל אחת מה-alphas
41	יצירת משתנה FDR ריק באורך מספר ה-alphas שיכיל את ערך ה-FDR לכל alpha
42	משתנה ריק באורך מספר ה-alphas שיכיל את מספר הבחינות המקורי (ללא פרמוטציה) שעובר את סף ה-alpha שקבענו בכל פעם
44	יצירת לולאת for שנותנת למשתנה k ערך עולה בין 1 למספר ה-alphas בכל איטרציה (כלומר עבור כל alpha)
45	הכנסה למשתנה sigOriginal (ראה שורה 42) את מספר הבחינות המקורי שעובר את סף ה-alpha באיטרציה הנוכחית בלולאת ה-for
46	ראה שורות 21,29-30 – בדיקה של מספר הבחינות בכל פרמוטציה שעוברות את סף ה- alpha באיטרציה הנוכחית בלולאת ה-for. colSums עוברת עמודה עמודה וסוכמת לפי התנאי שהצבנו (במקרה הזה שהערכים יהיו קטנים מה- alpha הנבדקת).
47	הכנסה למשתנה FDR את הממוצע של מספר הבחינות שעברו את סף ה-alpha בכל הפרמוטציות חלקי המספר המקורי (ראה שורה 45)
48	על מנת ליצר גרף מונוטוני מוודאים שהערך שהכנסנו בשורה 47 גדול או שווה לערך שהיה לפניו (פונקציית ההסתברות האמתית היא בהכרח מונוטונית).
51	ציור גרף ה-FDR כתלות במספר הבחינות שעברו את הסף. type='b' מציין שצריך לצייר גם (<u>b</u>oth) נקודות וגם קווים. ylab ו-xlab נותנים שמות לצירים.

נחזור לדוגמת הריטלין

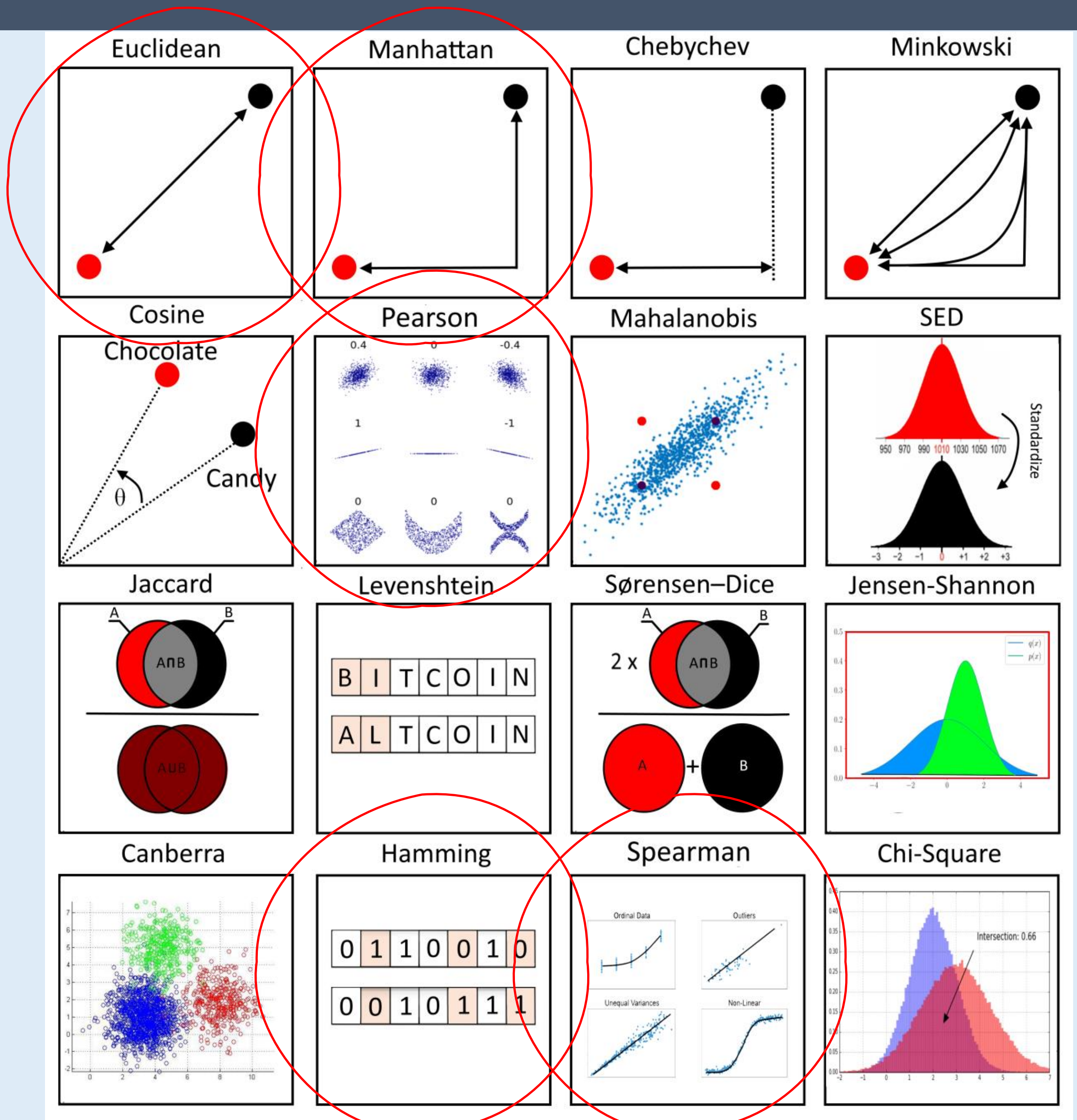


איך נבחר את הסף החדש? (FDR)

ככל ש-α עולה נגלה יותר מובהקויות אמיתיות (שקודם היו $p > \alpha$) אבל מצד שני נקבל גם יותר טעויות מסוג 1.

נוכל עכשיו לבחור α לפי כמות המובהקויות המתקבלות וההסתברות שלהן להיות טעויות.

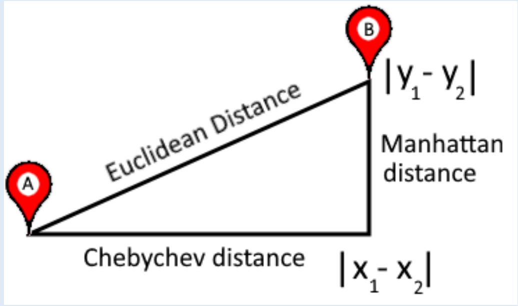
יש המון דרכים למדוד מרחק בין שתי נקודות



דמיון הוא מדד כמותי לכמה אובייקטים מרוחקים אחד מהשני – ככל שהם יותר רחוקים, כך הם פחות דומים.

כאשר יש לנו מידע חד-מימדי, למשל רמות סוכר בדם, קל לנו להשוות בין רמות הסוכר בדם בין שני מטופלים ולהגיד אם הם דומים או שונים זה מזה.

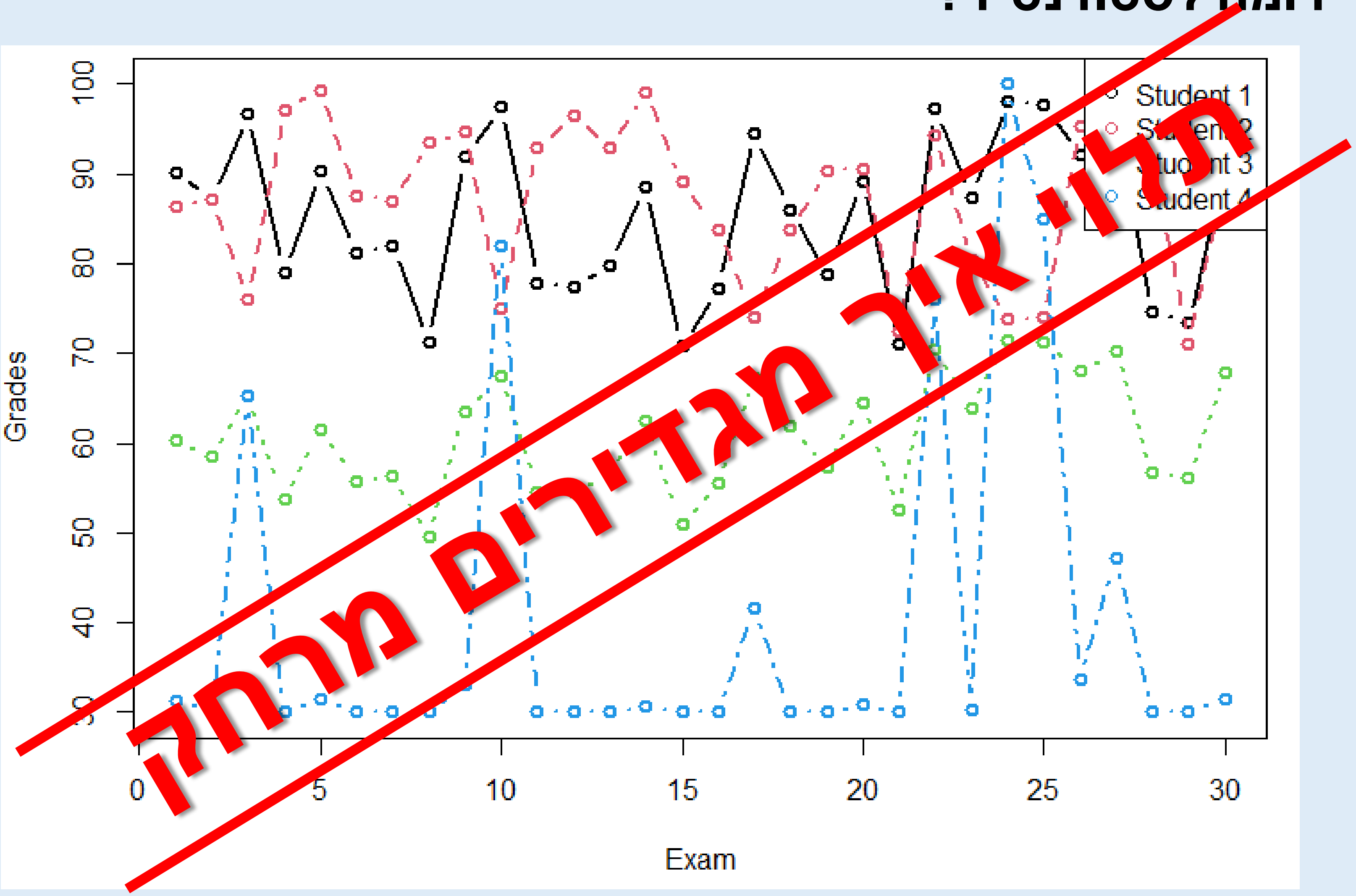
אך ככל שמספר המדדים שלנו גדל, כך קשה לנו יותר (אינטואיטיבית) להגדיר מרחק ודמיון בין נקודות.



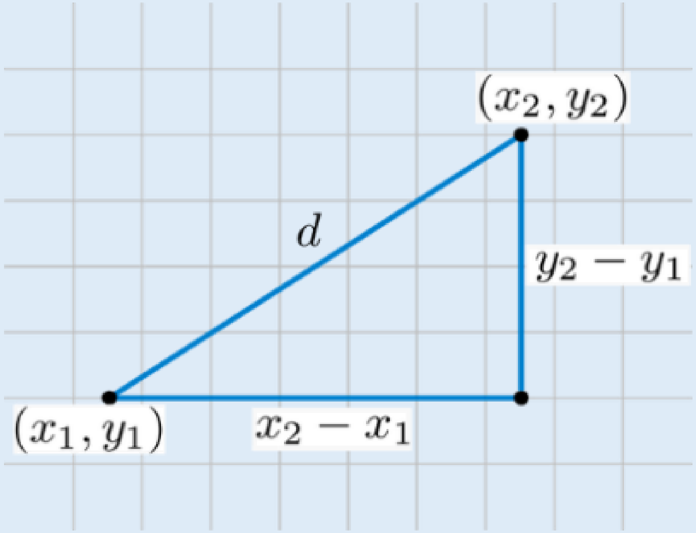
https://towardsdatascience.com/17-types-of-similarity-and-dissimilarity-measures-used-in-data-science-3eb914d2681

דמיון ומרחק

איזה סטודנט הכי דומה לסטודנט 1?



מרחק אוקלידי



בשני מימדים

מרחק אוקלידי זהה למשפט פיתגורס

$$d(1,2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

ברב מימד

$$d(1,2) = \sqrt{\Sigma_i \left(x_2^i - x_1^i\right)^2}$$

כאשר מחשבים מרחקים של מספר
דוגמאות מקבלים מטריצת מרחקים.
למשל בדוגמא עם ארבעת הסטודנטים:

0	73.2775	136.2525	262.5041
73.2775	0	156.1022	289.6258
136.2525	156.1022	0	143.6521
262.5041	289.6258	143.6521	0

מרחק אוקלידי

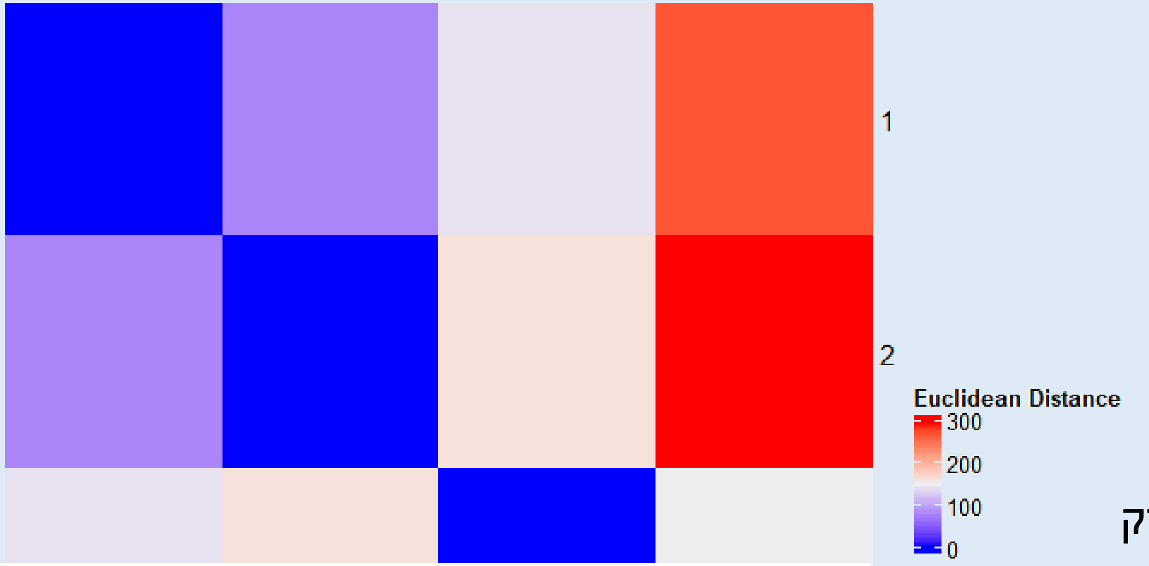
איזה סטודנט הכי דומה לסטודנט 1?

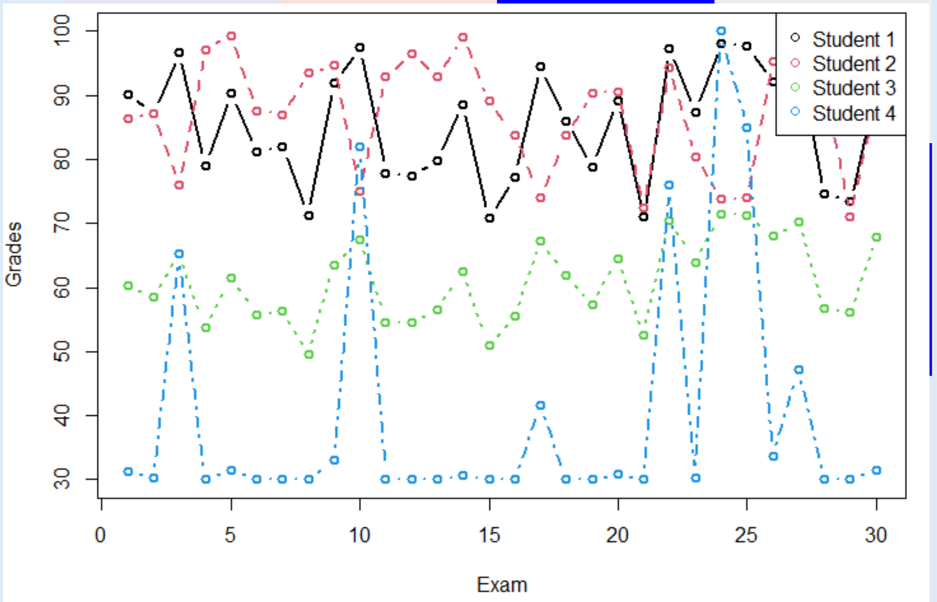
<u>נקודות מעניינות:</u>

1. יש לנו קו אלכסוני בצבע כחול (= מרחק 0 לפי סקלת הצבעים) שנמצאת בדיוק בנקודת המפגש של כל סטודנט עם עצמו. זה הגיוני מכיוון שאין מרחק בין כל סטודנט לעצמו.

2. התמונה משני צידי האלכסון היא זהה – כלומר, מספיק לנו לדעת רק את אחד הצדדים של המטריצה כדי לדעת את המרחקים.

3. על מנת למצוא את המרחק המקסימלי נסתכל בגרף ונראה שהמרחק הגדול ביותר הוא בין סטודנט 2 לסטודנט 4 (הצבע הכי אדום)





0	73.2775	136.2525	262.5041
73.2775	0	156.1022	289.6258
136.2525	156.1022	0	143.6521
262.5041	289.6258	143.6521	0

קורולציית פירסון (Pearson)

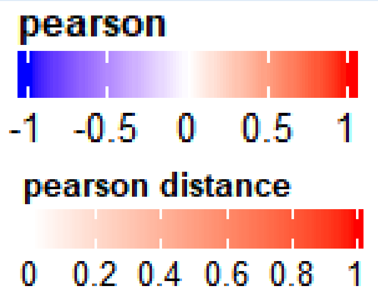
מודדת איך שני משתנים משתנים יחד (co - יחד, variance - שונות ← משתנים יחד)

$$r_{x,y} = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n}\left((x_i - \bar{x}) \cdot (y_i - \bar{y})\right)}{\sigma(x) \cdot \sigma(y)} = \frac{covariance(x,y)}{\sigma(x) \cdot \sigma(y)}$$

$r_{x,y}$ הוא מקדם פירסון, σ היא סטיית התקן ו-n מספר המימדים.

לדוגמא, ייתכן שממוצעי הציונים של מיטל ואריק שונים לחלוטין, אבל אם הם טובים באותם מקצועות וגרועים באותם מקצועות (כל אחד ביחס לממוצע שלו) – אז יש בינהם קורולציה חיובית.

- ערכי r תמיד יהיו בין -1 ל-1
- כדי לתרגם את פירסון למרחק נשתמש בנוסחא: $d = 1 - |r|$ או $d = 1 - r$

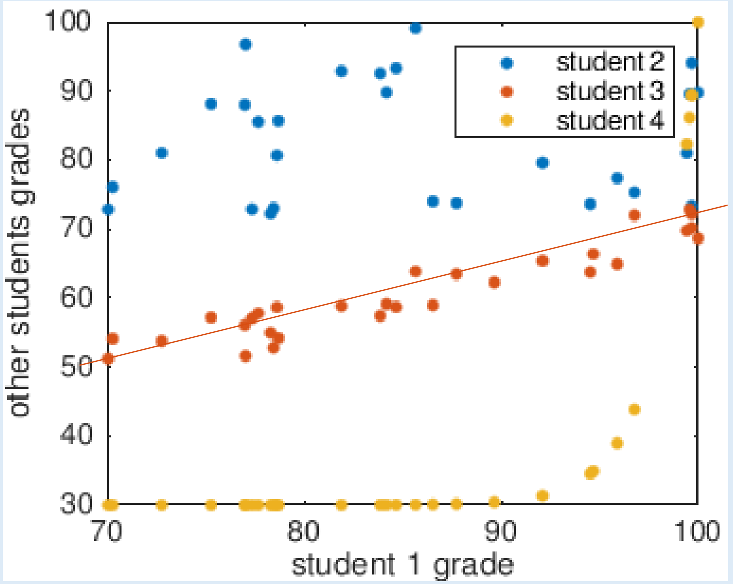


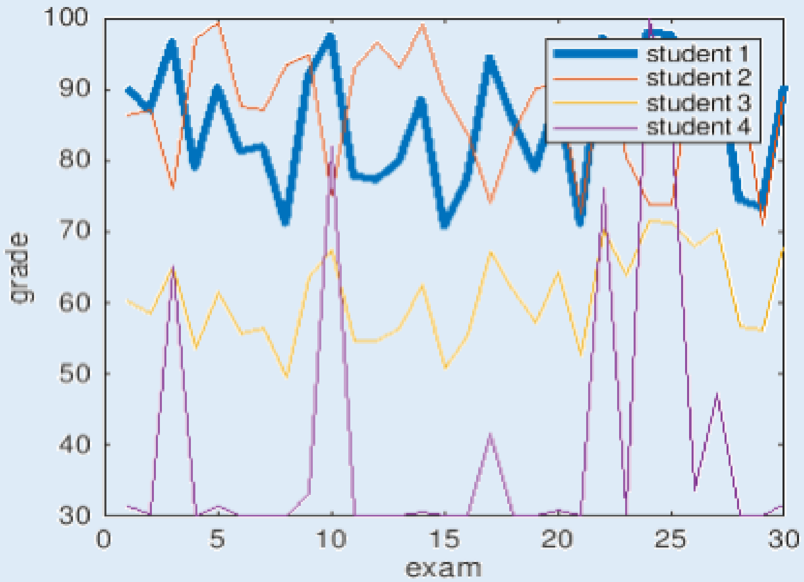
קורלציית פירסון

איזה סטודנט הכי דומה לסטודנט 1?

נחשב את המרחק בין כל שני סטודנטים לפי $1-r$

	1	2	3	4
1	0	1.1914	0.0593	0.3283
2	1.1914	0	1.2720	1.4742
3	0.0593	1.2720	0	0.3205
4	0.3283	1.4742	0.3205	0

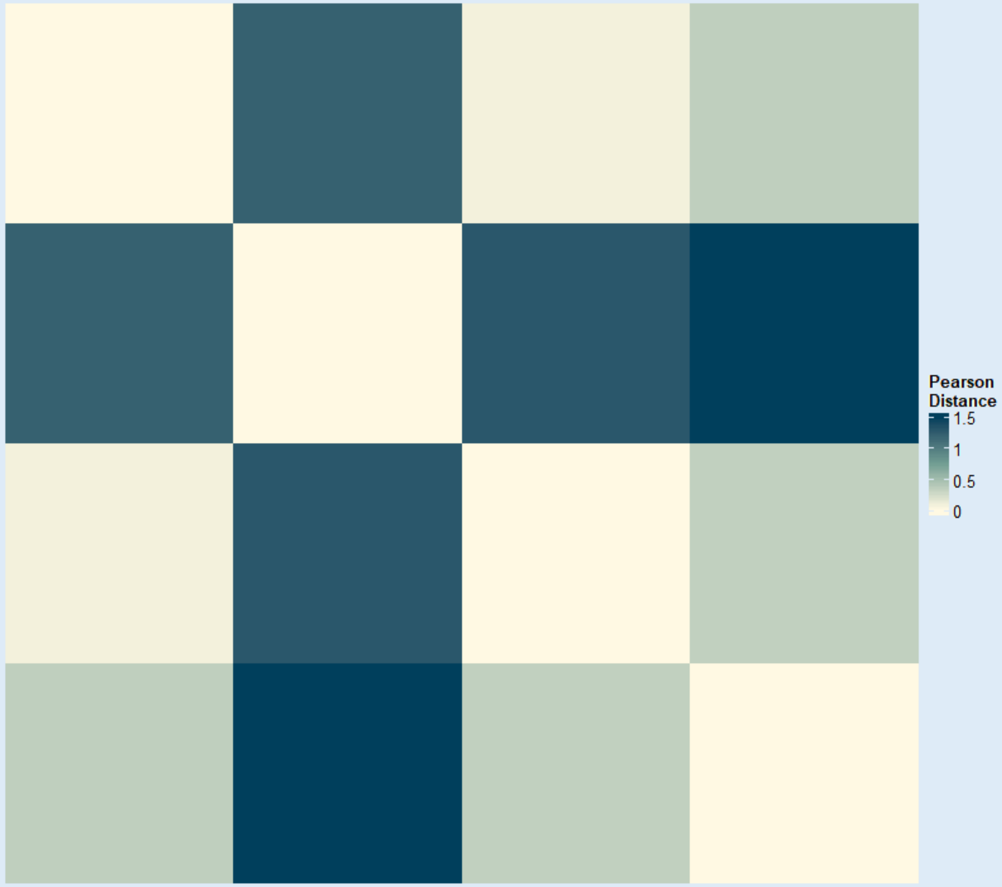




קורלציית פירסון

איזה סטודנט הכי דומה לסטודנט 1?

כאשר מחשבים מרחק אוקלידי שתי דוגמאות קרובות יהיו גם קרובות בגרף – אבל כאשר מחשבים פירסון, שתי דוגמאות קרובות יראו דומות, למרות מרחק אוקלידי שיכול להיות גדול.



קורלציית ספירמן

קורולציית ספירמן דומה לקורולציית פירסון (למעשה יש להן את אותה הנוסחה) – אבל במקום להשתמש בערך עבור כל דוגמא/מדד, היא משתמשת בדירוג שלו.

$$\rho = \frac{1}{n} \cdot \frac{\sum_i^n \left(\left(R(x_i) - \overline{R(x_i)} \right) \cdot \left(R(y_i) - \overline{R(y_i)} \right) \right)}{\sigma(R(x)) \cdot \sigma(R(y))}$$

תלמיד	מקצוע	ציון (x_i)	דירוג $(R(x_i))$
אריק	ביולוגיה של התא	100	1
	אותות ומערכות ב'	50	3
	פיזיקה 2ר'	70	2
אורנית	ביולוגיה של התא	30	3
	אותות ומערכות ב'	90	2
	פיזיקה 2ר'	100	1

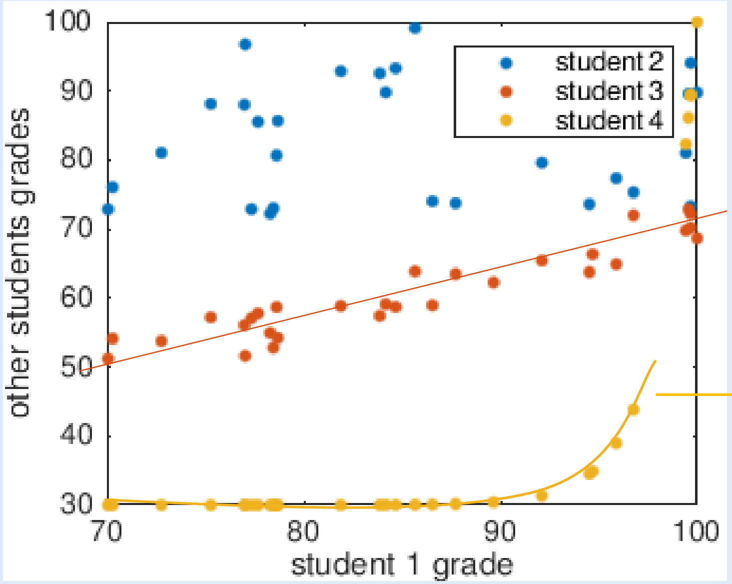
- כדי לתרגם את ספירמן למרחק נשתמש בנוסחא: $d = 1 - |\rho|$ או $d = 1 - \rho$

קורלציית ספירמן

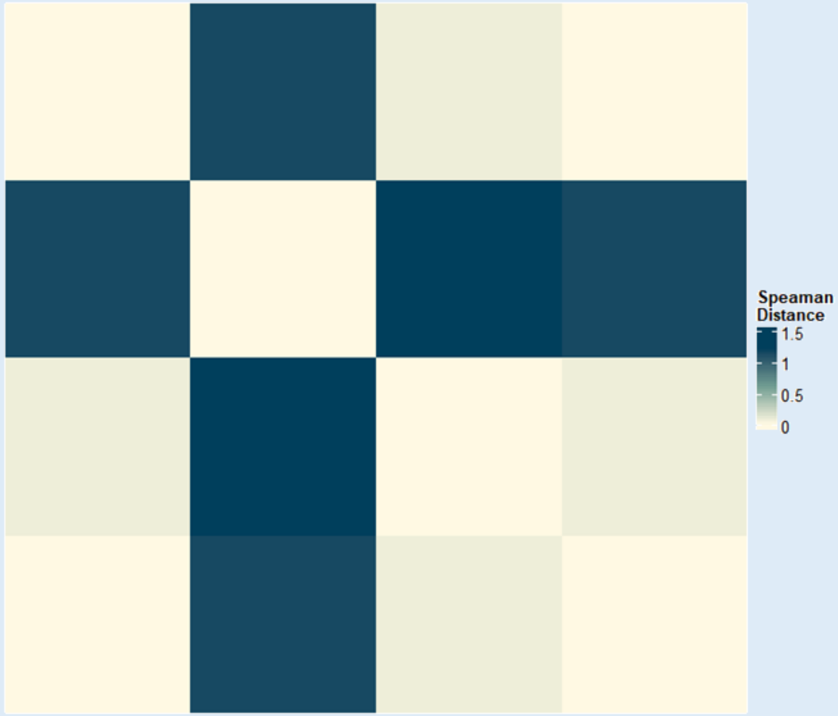
איזה סטודנט הכי דומה לסטודנט 1?

	1	2	3	4
1	0	1.1448	0.0681	0
2	1.1448	0	1.2067	1.1448
3	0.0681	1.2067	0	0.0681
4	$\boxed{0}$	1.1448	0.0681	0

שימו לב הפונקציה מחזירה מרחק ולא קורלציה

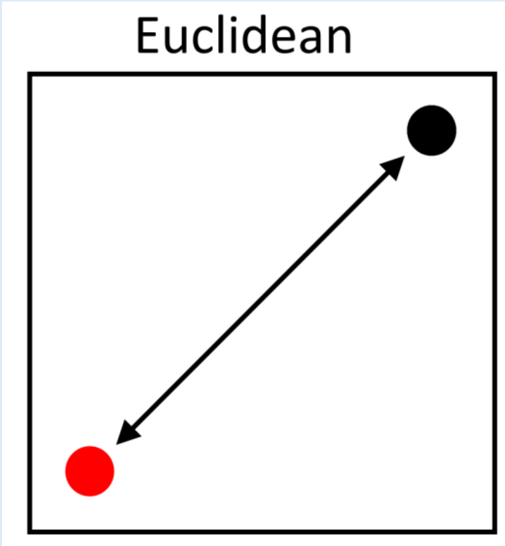


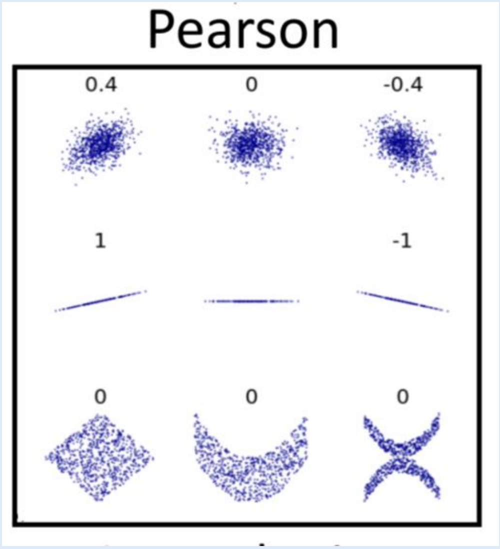
לא לינארי אבל כן מונוטוני

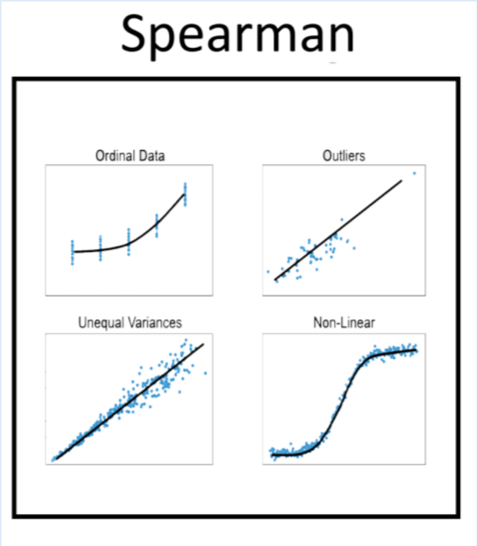


סיכום מרחקים

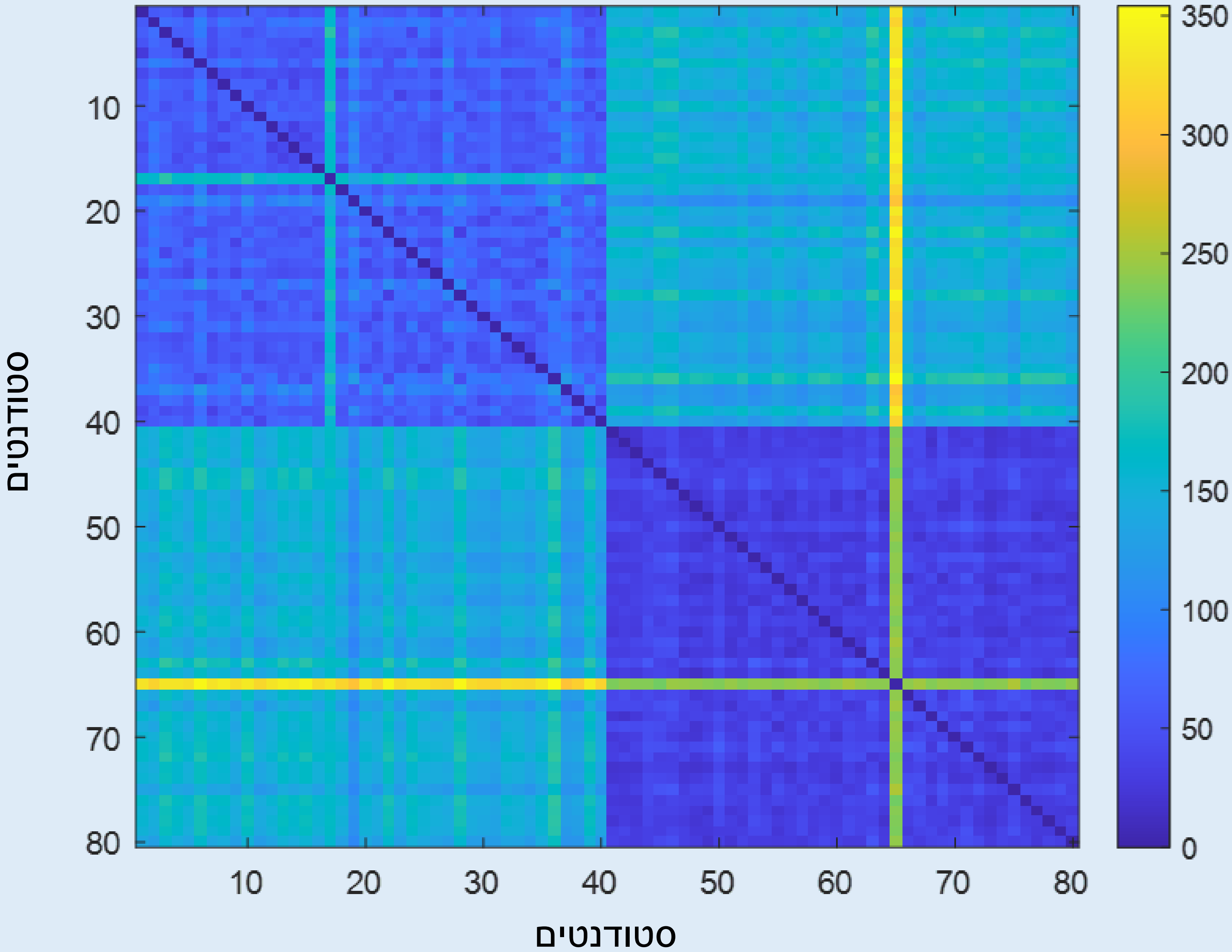
1. מרחק אוקלידי נותן לנו מרחק לינארי בין דוגמאות/מדדים
2. קורולציית פירסון נותנת לנו קורולציה המתארת קשר לינארי בין דוגמאות/מדדים
   - כאשר הקשר הוא לינארי
   - כאשר אנחנו רוצים להשוות בין מדדים עם יחידות מידה שונות
3. קורולציית ספירמן נותנת לנו קורולציה המתארת קשר מונוטוני בין דוגמאות/מדדים
   - כאשר יש לנו דוגמאות חריגות
   - כאשר אנחנו רוצים להשוות בין מדדים עם יחידות מידה שונות







נסגור מעגל...האם יש הבדל בין סטודנטים עם ובלי ריטלין?

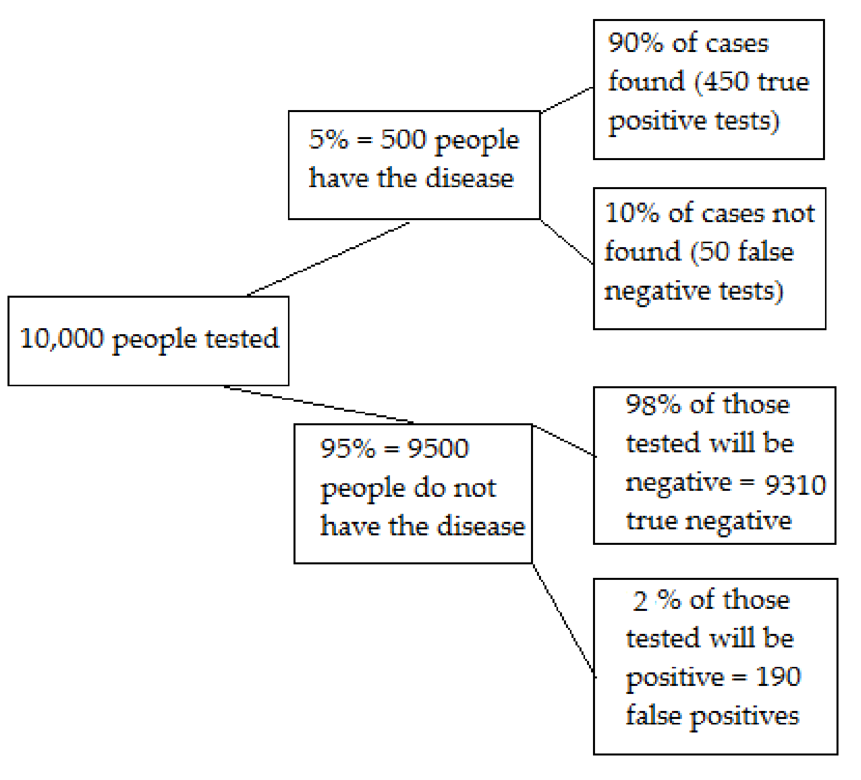


להתעלם מהפסים

שאלה 1

$$FDR = \frac{190}{190 + 450} = 0.296875$$

להלן תיאור של מבחן רפואי אשר מזהה באופן מדוייק 90% מהמקרים בהם יש את המחלה. המבחן נבדק על אוכלוסייה של 10,000 אנשים.



מהו ה-FDR של המבחן?

יש לבחור תשובה אחת:

- 0.3 (selected)
- 0.38
- 0.02
- 0.9

שאלה 2

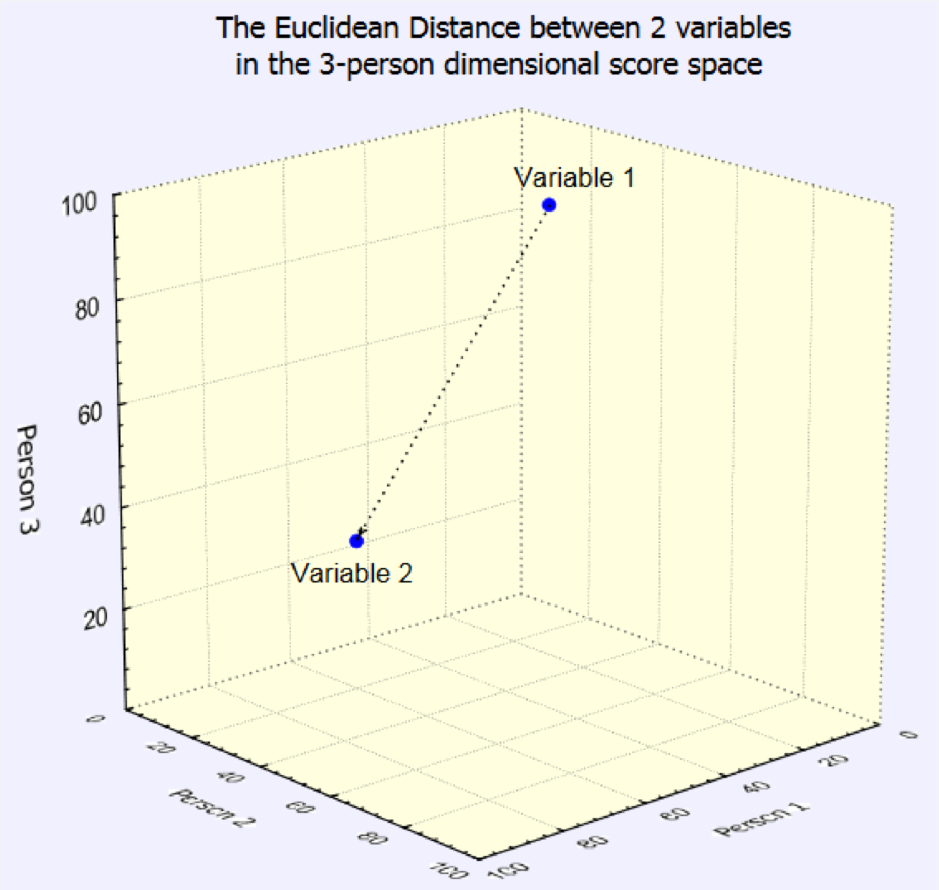
רמת ביטוי של 5,000 גנים נמדדה בדם של כ-40 נבדקים שונים. חצי מהנבדקים טופלו בתרופה וחצי לא. נרצה לדעת אילו גנים מושפעים מהטיפול. בגלל ריבוי ההשוואות ברצוננו לחשב את ה– false discovery rate (FDR) בעזרת פרמוטציות. כיצד עלינו לעשות זאת?

יש לבחור תשובה אחת:

- יש לספור את מספר הגנים שעברו את סף המובהקות, לבחון את ההבדל בין הקבוצות ולאחר מכן לבצע חלוקה של הנבדקים לשתי קבוצות אקראיות ועבור אלפא נתון לספור את מספר הגנים שעברו את סף המובהקות. היחס בין שני אלו הוא ה-FDR
- יש לספור את מספר הגנים שעברו את סף המובהקות לאחר תיקון בנפורוני, ערך אלפא הוא ה-FDR
- יש לספור את מספר הגנים שעברו את סף המובהקות, לבחון את ההבדל בין הקבוצות ולאחר מכן לבצע ערבול של הגנים בכל נבדק ועבור אלפא נתון לספור את מספר הגנים שעברו את סף המובהקות. היחס בין שני אלו הוא ה-FDR
- יש לבצע חלוקה אקראית של הנבדקים לשתי קבוצות, לבחון את ההבדל בין הקבוצות החדשות, ועבור אלפא נתון נספור את מספר הגנים שעברו את סף המובהקות

שאלה 3

נתון המרחק האוקלידי בין 3 אנשים ו-2 משתנים:



עם הנתונים:

	1 Var1	2 Var2
Person 1	20	80
Person 2	30	44
Person 3	90	40

מה המרחק בין שני האנשים הרחוקים ביותר זה מזה בחישוב המרחק האוקלידי?

דיוק של שתי ספרות אחרי הנקודה

תשובה: 37.36

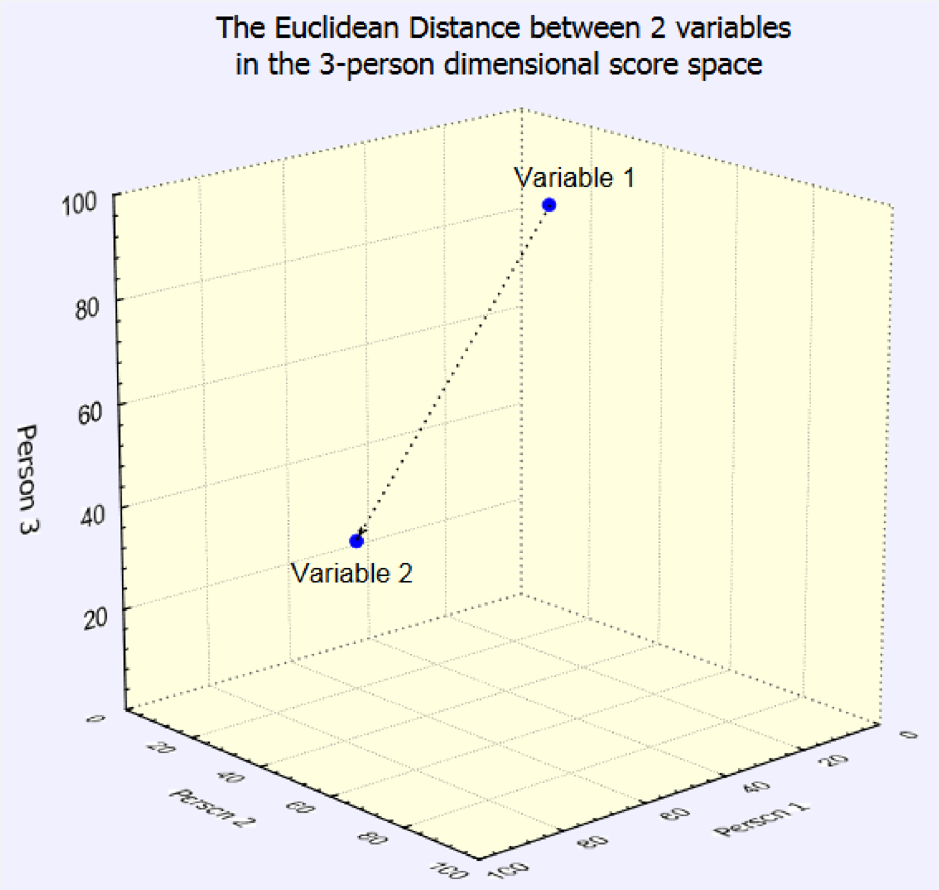
$$d(1,2) = \sqrt{(20-30)^2 + (80-44)^2} = 37.36$$

$$d(1,3) = \sqrt{(20-90)^2 + (80-40)^2} = 80.62$$

$$d(2,3) = \sqrt{(30-90)^2 + (44-40)^2} = 60.13$$

שאלה 4

נתון המרחק האוקלידי בין 3 אנשים ו-2 משתנים:



עם הנתונים:

	1 Var1	2 Var2
Person 1	20	80
Person 2	30	44
Person 3	90	40

מהו המרחק בין שני המשתנים (Var 1 מול Var 2)?

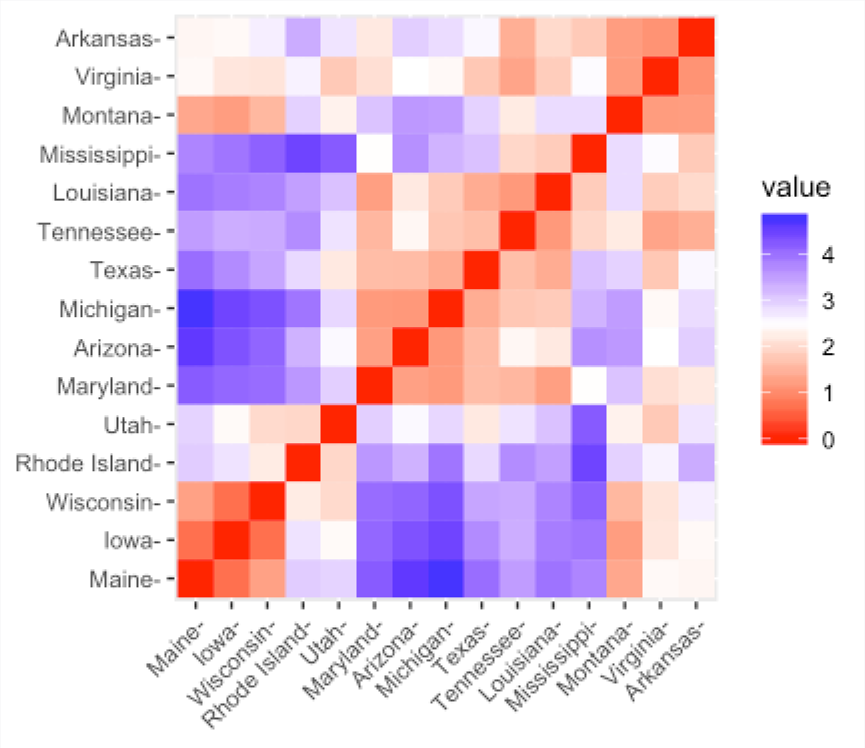
דיוק של שתי ספרות אחרי הנקודה

תשובה: 79.35

$$d(1,2) = \sqrt{(20-80)^2 + (30-44)^2 + (90-40)^2}$$
$$= 79.347$$

שאלה 5

נתונה מטריצת מרחקים:



בעזרת איזה מרחק חישבנו את המטריצה?

יש לבחור תשובה אחת:

- פירסון
- אוקלידי
- ספירמן

טווח הערכים אינו בין 0 ל-1

שאלה 6

שני רופאים מעריכים את מצבם הרפואי של 8 מטופלים הסובלים ממיגרנות.

הם דירגו את המטופלים לפי חומרת הסימפטומים כאשר 1 - מצב מצויין ו-8 מצב נוראי.

מטופל	רופא מספר 1	רופא מספר 2
א	4	5
ב	1	3
ג	3	1
ד	2	2
ה	6	6
ו	5	4
ז	7	8
ח	8	7

מה נצפה שתהיה קורולציית ספירמן?

יש לבחור תשובה אחת:

- קרובה ל-1-
- קרובה ל-0
- קרובה ל-1
- לא ניתן לחשב כאן קורולציית ספירמן

הם נוטים להסכים,
כאשר רופא 1 נותן
ציון גבוה גם רופא 2
נותן ציון גבוה
ולהפך

שאלה 7

לרופא מסויים יש מטופלים עם יתר לחץ דם להם הוא נותן טיפול להורדת לחץ הדם. להלן דוגמא של מטופל כזה לו הוא בודק לחץ דם פעם בחודש:

ינואר	פברואר	מרץ	אפריל	מאי	יוני	יולי	אוגוסט	ספטמבר	אוקטובר	נובמבר	דצמבר
172	155	146	149	154	166	170	157	135	140	140	150

בחודש דצמבר הגיע אל אותו הרופא מטופל חדש עם נתוני לחץ הדם מהשנה החולפת:

ינואר	פברואר	מרץ	אפריל	מאי	יוני	יולי	אוגוסט	ספטמבר	אוקטובר	נובמבר	דצמבר
122	105	96	89	104	116	120	107	85	90	90	100

הרופא רוצה לדעת האם לתת טיפול ליתר לחץ דם למטופל החדש. על מנת לעשות זאת, הרופא החליט למדוד את המרחק בלחץ הדם בין המטופל החדש למטופל שלו עם יתר לחץ דם.

טווח הערכים התקין של לחץ הדם הוא בין 120-140 מ"מ כספית ויתר לחץ דם הוא מעל 140 מ"מ כספית.

האם לתת למטופל החדש טיפול ליתר לחץ דם?

יש לבחור תשובה אחת:

- כן, מכיוון שכאשר משתמשים בקורולציית ספירמן הקורולציה בין המטופלים היא 1
- כן, מכיוון שההבדל בינהם הוא רק 50.
- כן, מכיוון שכאשר משתמשים בקורולציית פירסון הקורולציה בין המטופלים היא 1
- לא, מכיוון שהמרחק האוקלידי בין המטופלים הוא גדול וטווח הערכים של המטופל החדש לא דומה לטווח הערכים של המטופל הותיק.

שאלה 8

איזה מהמשפטים הבאים לא נכון?

יש לבחור תשובה אחת:

- כפל כל קואורדינטה במספר קבוע לא תשפיע על קורולציית ספירמן ופירסון
- הוספת מספר קבוע לכל קואורדינטה לא תשפיע על המרחק האוקלידי בין הנקודות.
- הוספת מספר קבוע לכל קואורדינטה לא תשפיע על קורולציות ספירמן ופירסון
- הוספת מספר קבוע לא תשפיע על קורולציית פירסון אבל כן תשפיע על קורולציית ספירמן
- כפל כל קואורדינטה במספר קבוע משפיע על המרחק האוקלידי בין הנקודות.

שאלה 9

כמו שראינו בשאלות קודמות, הבחירה בשיטת חישוב דמיון יכולה להיות בעלת השפעה מכרעת על התוצאות.

מי מאלו אינו נכון?

יש לבחור תשובה אחת:

- ○ נשתמש בקורולציית פירסון כשנרצה להעריך את הקורולציה של משתנים הנמצאים ביחס לינארי
- ○ נשתמש בקורולציית פירסון כשהכיווניות חשובה לנו אך הגודל לא
- ◉ נשתמש בקורולציית ספירמן במקום בפירסון במצב בו המדדים המושווים נמדדו ביחידות שונות
- ○ נשתמש במרחק אוקלידי כשגודל השינוי חשוב לנו יותר מכיוון ומגמת השינוי