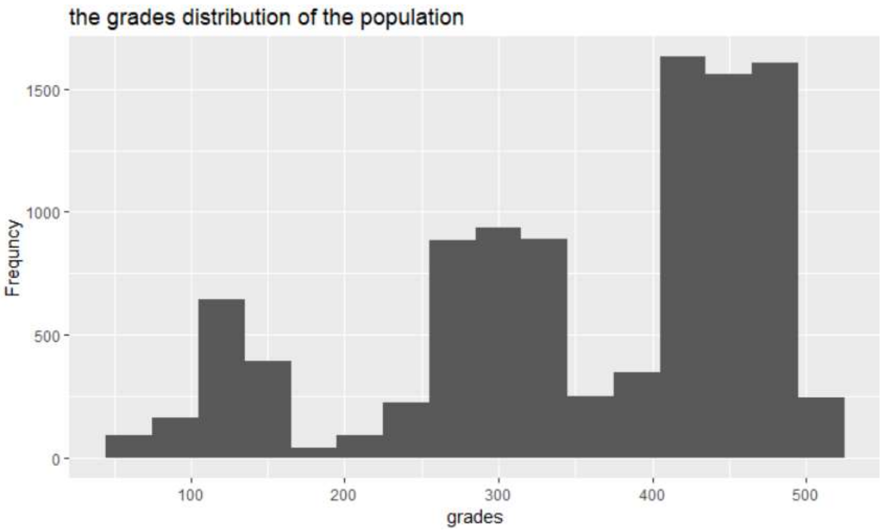
שיטות מחקר- תרגיל מספר 2

מגישות: זוהר בוחניק 311142293, אלה דובדבן 305564866

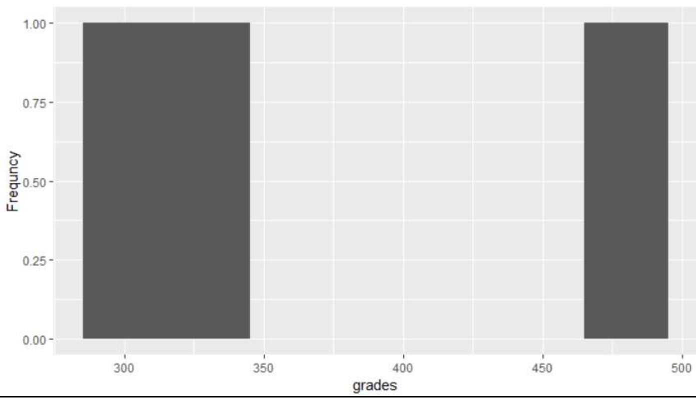
שאלה 1:

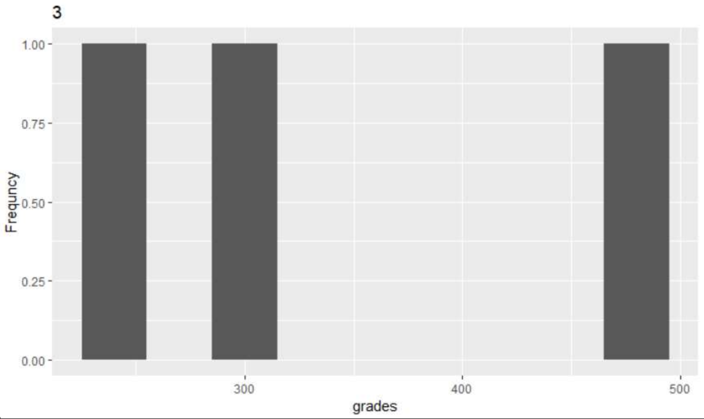
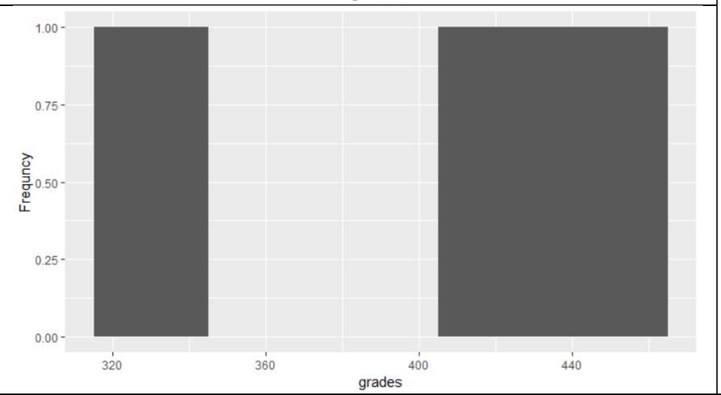
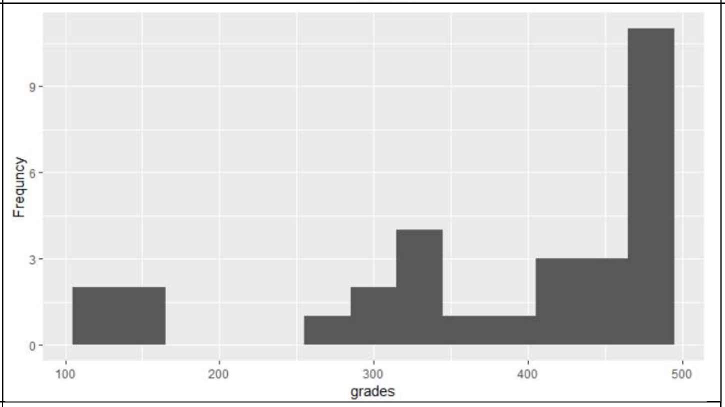
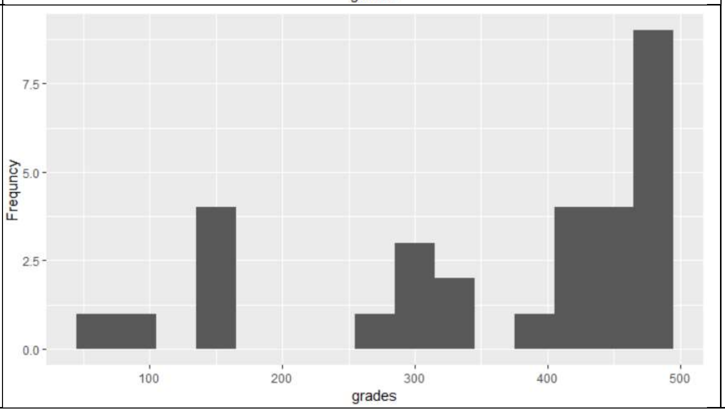
א. ציירו את התפלגות האוכלוסיה, וחשבו את התוחלת וסטיית התקן שלה.

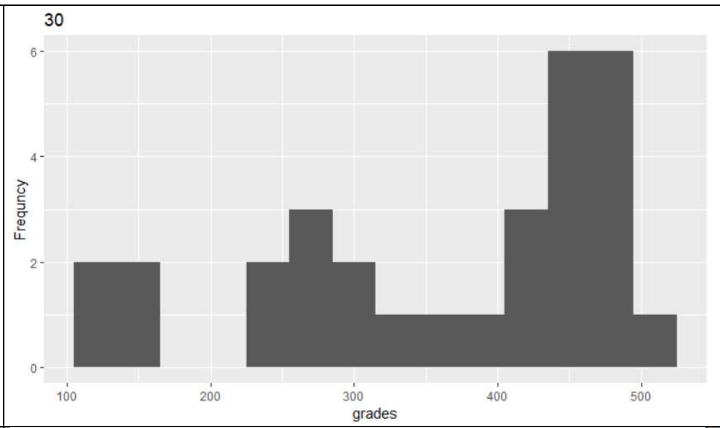
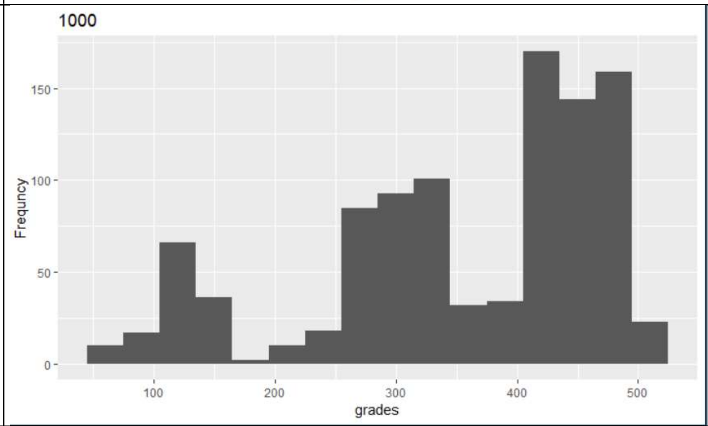
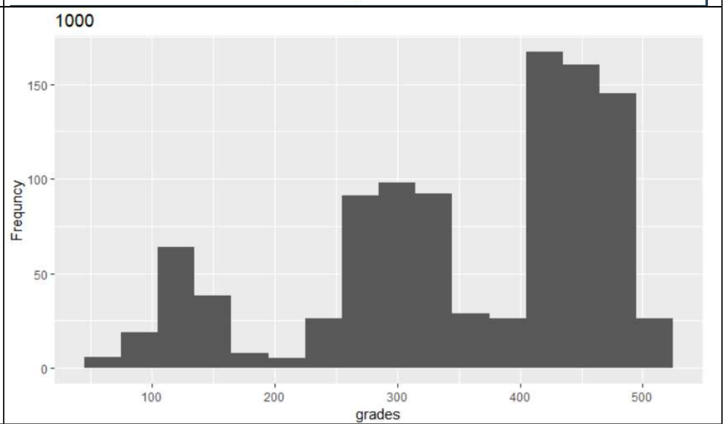
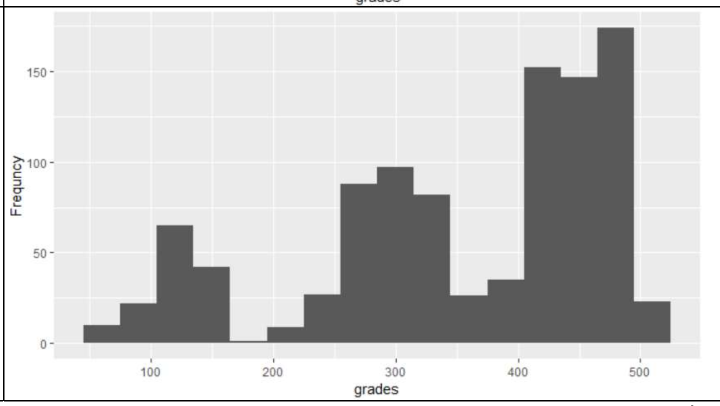


ממוצע: 355.8492 **סטיית תקן**: 118.3552

ב. דגמו מתוך האוכלוסייה מדגמים בגדלים שונים: n=3,30, 1000., 3 פעמים עבור כל גודל מדגם.
עבור כל מדגם, ציירו את התפלגות המדגם, וחשבו את הממוצע וסטיית התקן שקיבלתם במדגם.
העתיקו את הגרפים וסכמו את הנתונים בטבלה למטה. מה ניתן להסיק בעקבות הסימולציה הזו?
דונו בעזרת המושגים שנלמדו בכיתה.

גודל מדגם	**גרף התפלגות המדגם**	**אומדים במדגם (ממוצע וסטיית תקן)**
N=3		mean = 370.6061 sd = 96.06628

mean = 344.4303 sd = 112.4588		
mean = 398.7294 sd = 70.19157		
mean = 378.8241 sd = 119.0951		
mean = 358.5816 sd = 137.6373		N=30

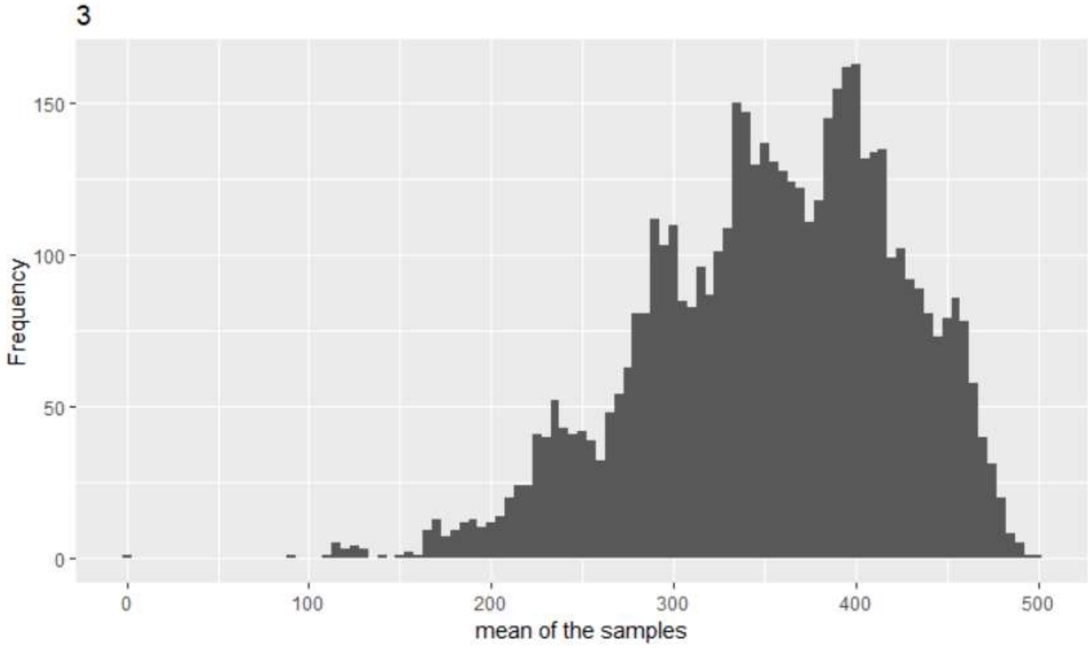
<table>
<tr><td>mean = 361.4182
sd = 122.3804</td><td>
</td><td></td></tr>
<tr><td>mean = 354.724
sd = 117.7839</td><td>
</td><td rowspan="3">N=10
00</td></tr>
<tr><td>mean = 353.9865
sd = 117.1619</td><td>
</td></tr>
<tr><td>mean = 353.2237
sd = 120.8778</td><td>
</td></tr>
</table>

(המשך שאלה מספר 1)

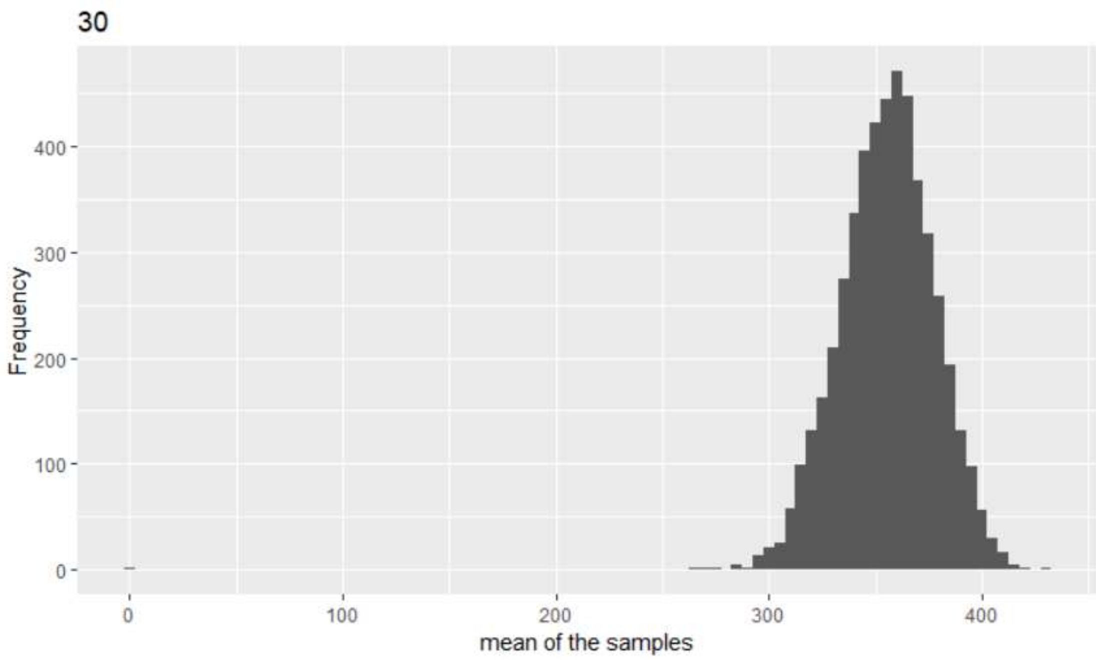
ג. כעת, דגמו מדגמים בגודל 3, 30, ו-1000 מספר רב של פעמים (למשל, 5000 פעמים). עבור כל הרצה, חשבו את ממוצע המדגם. לאחר מכן, עבור כל גודל מדגם, ציירו את **התפלגות הדגימה של הממוצע,** ובדקו מהי התוחלת וסטיית התקן שלה. האם התשובות שקיבלתם מתאימות לנלמד בכיתה? דונו בעזרת המושגים שנלמדו בכיתה.

נציג את הנתונים שקיבלנו עבור כל N :

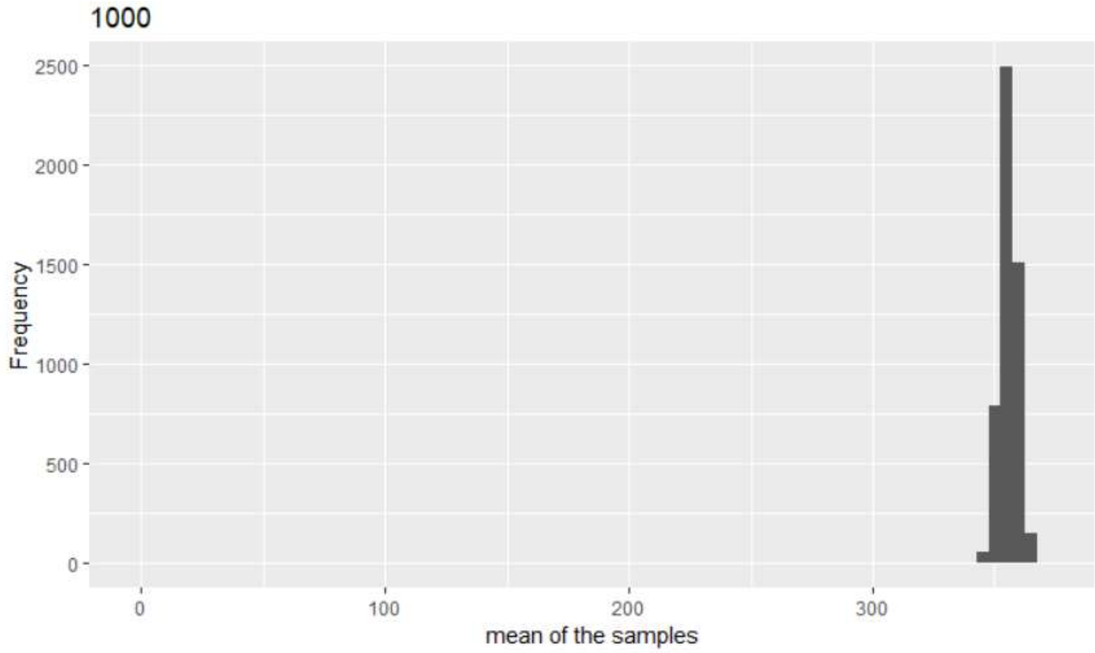
N= 3, **mean**= 354.4215, **sd**= 354.4215



N= 30, **mean**= 355.5271, **sd**= 355.5271



N= 1000, **mean**= 355.8385, **sd**= 355.8385



ניתן לראות כי ככל שמספר הדגימות שלקחנו ומיצענו היה גדול יותר (3,30,1000) כך גם השונות בין הממוצעים שקיבלנו ירדה. במילים אחרות, ככל שהיו לנו יותר דגימות כך התכנס בהתאם הממוצע של הדגימות לתוצאה דומה. נתון זה עולה בקנה אחד עם חוק המספרים הגדולים שעליו דיברנו בכיתה האומר כי כאשר מספר הדגימות שואף לאינסוף, כך בתוחלת הממוצע של הדגימות מתכנס לתוחלת של הדגימות. ניתן לראות גם שלא משנה עבור איזה N חישבנו, קיבלנו תוחלת דומה בכולם מה שגם מחזק את הנאמר.

<u>שאלה 2:</u>

א+ב+ג.

manual calculation:

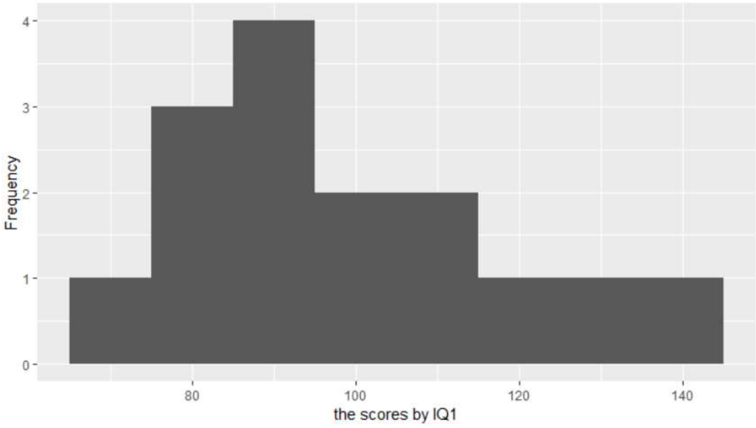
estimated var = 375.6952   **sd** = 19.38286

r check:

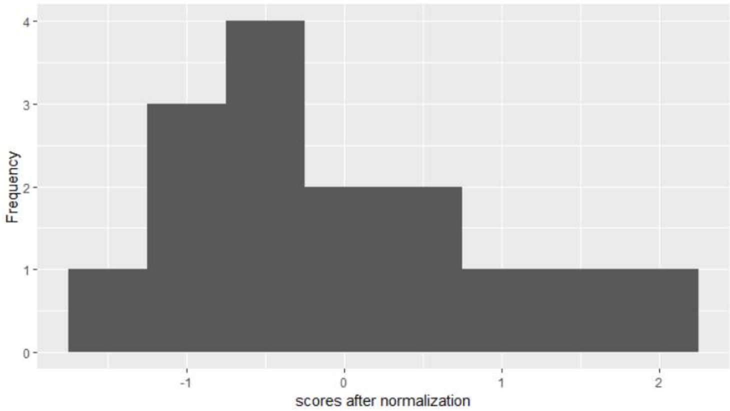
var = 375.6952   **sd** = 19.38286

ד. קיבלנו כי התוחלת קרובה מאוד ל-0 ($8.883591*10^{-17}$), וקיבלנו כי השונות היא 1. התוצאות מתאימות לנירמול ההתפלגות לכדי התפלגות נורמלית סטנדרטית – כלומר חישוב ערכי ה-z המייצגים את ציון התקן כפי שלמדנו בכיתה.

ה. גרף של התפלגות הציונים המקוריים:



גרף של התפלגות הציונים אחרי נרמול:



נבחין כי לאחר הנרמול נשמרה ההתפלגות של הנתונים, השוני היחידי שקרה הוא בטווח שבו הנתונים נמצאים. בציונים ללא נרמול הטווח הוא בין 0 ל-140 ואילו לאחר הנרמול הטווח הוא בין 1- ל2.
נשים לב שיש חשיבות רבה לנירמול תוך שמירה על ההתפלגות המקורית, שכן עבור משתנה המתפלג נורמלי סטנדרטי יותר קל לבצע את החישובים השונים וזאת בלי לפגוע בהתפלגות המקורית.

<u>שאלה 3:</u>

א. עוזר המחקר הוריד שלוש תצפיות קיצוניות ולכן שונות התפלגות הדגימה תגדל.

לא נכון. הורדת התצפיות הקיצוניות דווקא מקטינה את השונות.

ב. עוזר המחקר הוריד את התצפית הגבוהה ביותר ואת התצפית הנמוכה ביותר ולכן ממוצע המדגם ישתנה.

נכון חלקית. הורדת התצפית הנמוכה ביותר והגבוהה ביותר משפיעה על הסכום של תוצאות התצפיות ולכן כאשר נחשב ממוצע על הדגימות (בעזרת סכום חלקי מספר הדגימות) התוצאה יכולה להשתנות. חלקית- כי במידה וכל הדגימות בעלות תוצאה זהה הורדה של שתי דגימות לא תשנה את הממוצע, לא משנה איזה דגימות נוריד)

ג. עוזר המחקר הוריד את התצפית הגבוהה ביותר ואת התצפית הנמוכה ביותר ולכן שונות המדגם תישאר זהה.

לא נכון. כאשר מורידים את התצפיות הגבוהה ביותר והנמוכה ביותר אנו מקטינים את השונות.

ד. עוזר המחקר הכפיל כל תצפית של המשתנה ב-2 ולכן תוחלת המשתנה תגדל פי חמש.

לא נכון. היא תגדל בדיוק פי 2 מלינאריות התוחלת.

ה. עוזר המחקר החסיר מכל תצפית 10 נקודות ולכן ממוצע המדגם יישאר אותו דבר.

לא נכון. הממוצע ירד כי הוא מושפע באופן ישיר מסכום הדגימות

ו. עוזר המחקר החסיר מכל תצפית 10 נקודות ולכן שונות המדגם תשתנה.

לא נכון. השונות תישאר אותו דבר כי הוספת מספר קבוע על הדגימות לא משנה את השונות.

ז. עוזר המחקר הכפיל כל תצפית של המשתנה ב-2, הדבר ישפיע על הממוצע ועל שונות המדגם, אך השונות תושפע מכך פחות.

לא נכון. מלינאריות התוחלת, היא תגדל פי 2, אמנם עבור השונות- כאשר מכפילים את תוצאות המדגם בקבוע, אנו מכפילים את השונות של הדגימות בריבוע של הקבוע, לכן תושפע יותר.

__שאלה 4:__

התשובה היא **ג'**

ראינו בכיתה כי מתקיים:

$$V(\bar{X}) = V\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{1}{n^2}V(x_1 + x_2 + \cdots + x_n)$$
$$= \frac{1}{n^2}V(x_1) + V(x_2) + \cdots + V(x_n) = \frac{1}{n^2}[\sigma^2 + \sigma^2 + \cdots + \sigma^2] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

כלומר סטיית התקן היא בריבוע בניגוד מה שכתוב בג'.

__שאלה 5:__

התשובה היא- 3 שתיהן טועות

הפונקציה dnorm מחשבת את פונקציית הצפיפות, ואילו הפונקציה qnorm מחשבת את האחוזון (הערך שx אחוז מהתצפיות קטנות ממנו) שזהו בעצם ההופכי של pnorm

כדי לחשב את מה שדנה מעוניינת למצוא היא צריכה להשתמש בפונקציה pnorm, אשר מחשבת את האינטגרל ממינוס אינסוף עד לערך המבוקש – ובכך מחזירה את ההסתברות שאנשים יקבלו באופן רנדומלי בובה ממשקל 0.46 או פחות.

הפונקציה ב-R מפורטת בקוד המצורף,

התוצאה אשר התקבלה הינה 0.4207403 .