שיטות מחקר תרגיל 2

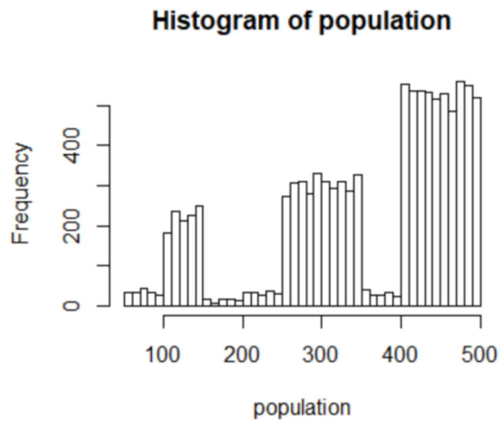
מגישות :

מיכל דגן 315657064

נועה בן דרור 316163260

שאלה 1: ה1 בקובץ המצורף של התרגיל כיתה.

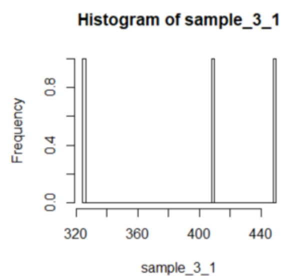
שאלה 1:

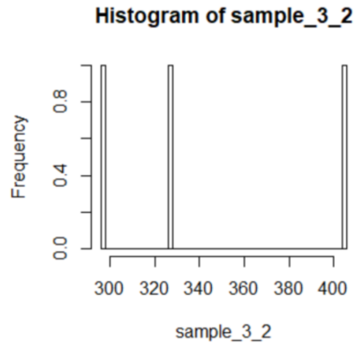
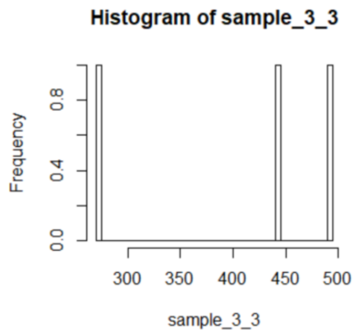
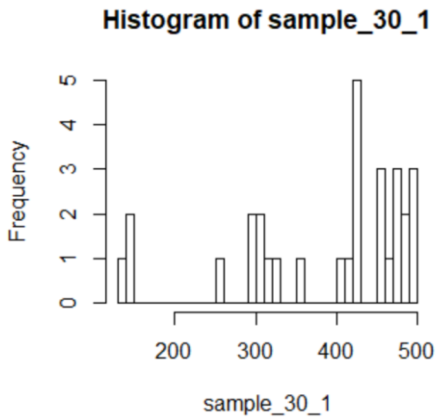
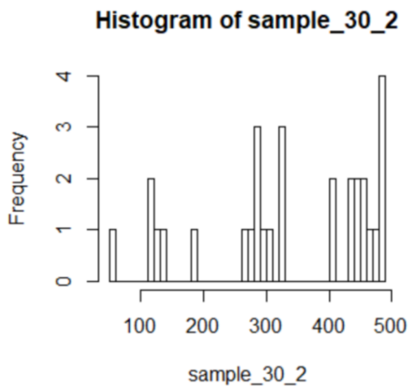


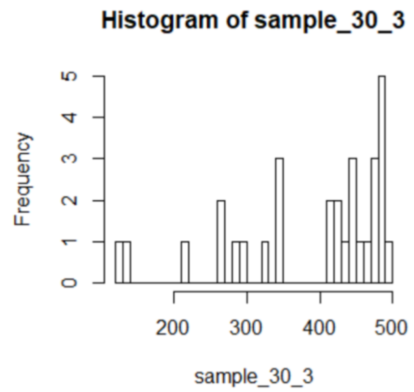
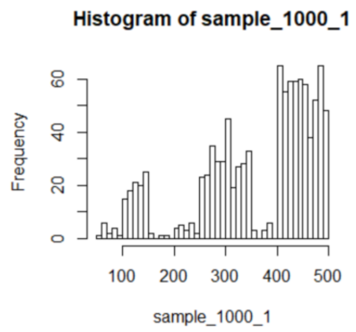
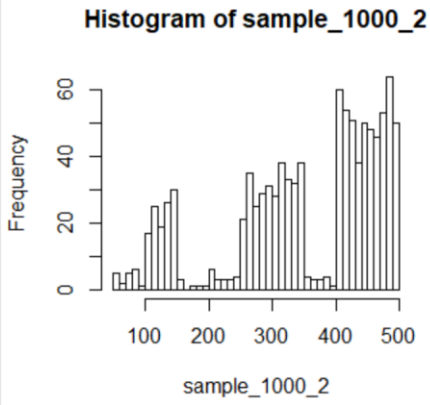
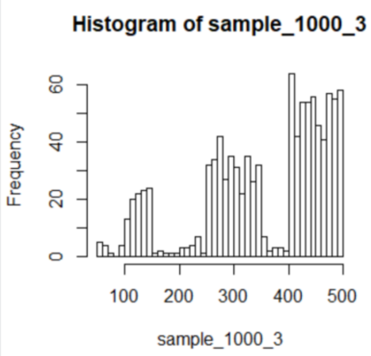
השתמשנו בממוצע – שכן התוחלת במקרה זה היא הממוצע של ציוני הנבדקים, מכיוון שההסתברות בניסוי לקבלת כל תוצאה היא שווה.

```
> mean(population)
[1] 355.8492
> sd(population)
[1] 118.3552
```

סעיף 2:

גודל המדגם	גרף התפלגות המדגם	אומדים במדגם (ממוצע וסטיית תקן)
n=3		> mean(sample_3_1) [1] 394.3608 > sd(sample_3_1) [1] 63.26862

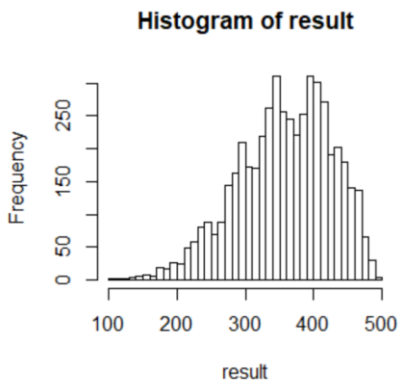
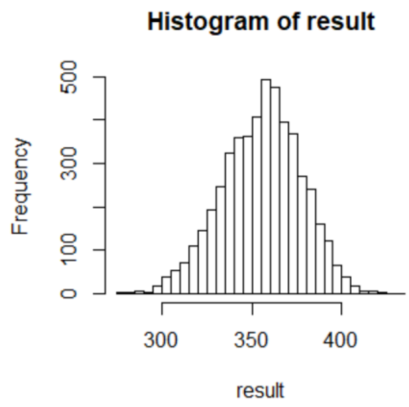
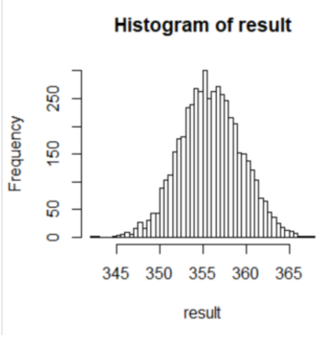
<table>
<tr><td>> mean(sample_3_2)
[1] 342.7559
> sd(sample_3_2)
[1] 56.62176</td><td>
</td><td></td></tr>
<tr><td>> mean(sample_3_3)
[1] 400.5981
> sd(sample_3_3)
[1] 115.214</td><td>
</td><td></td></tr>
<tr><td>> mean(sample_30_1)
[1] 384.13
> sd(sample_30_1)
[1] 108.2927</td><td>
</td><td>n=30</td></tr>
<tr><td>> mean(sample_30_2)
[1] 334.7422
> sd(sample_30_2)
[1] 131.4927</td><td>
</td><td></td></tr>
</table>

> mean(sample_30_3) [1] 388.6952 > sd(sample_30_3) [1] 107.5582		
> mean(sample_1000_1) [1] 361.6202 > sd(sample_1000_1) [1] 114.6515		n=1000
> mean(sample_1000_2) [1] 352.3896 > sd(sample_1000_2) [1] 119.87		
> mean(sample_1000_3) [1] 356.6969 > sd(sample_1000_3) [1] 116.0811		

מסקנות:

נוכל להסיק ע״י התבוננות בנתונים והשוואתם לסעיף א׳ כי ככל שהמדגם שלנו גדול יותר הממוצע וסטיית התקן הולכים ומתקרבים (שואפים) אל הממוצע וסטיית התקן שנמדדו מהתבוננות בכלל האוכלוסייה.

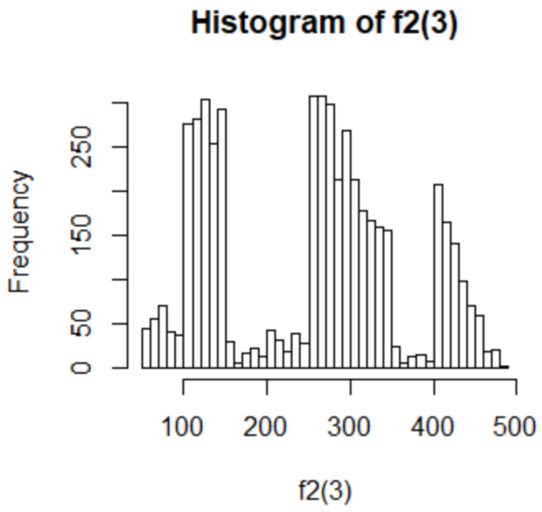
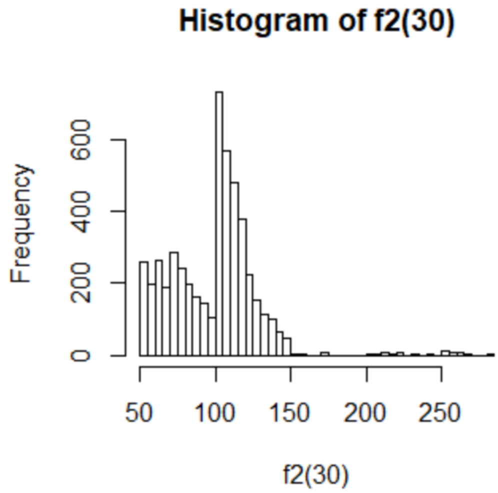
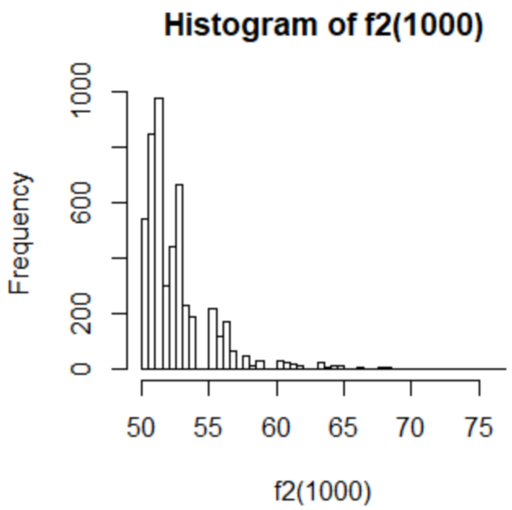
סעיף 3

גודל המדגם	התפלגות הדגימה של הממוצע	תוחלת וסטיית התקן
$n=3$		> mean(result) [1] 357.2776 > sd(result) [1] 68.32333
$n=30$		> mean(result) [1] 355.3891 > sd(result) [1] 21.89853
$n=1000$		> mean(result) [1] 355.919 > sd(result) [1] 3.542965

מסקנות

ניתן לראות כי ככל שהגדלנו את n, כך ההתפלגות הדגימה של הממוצע שואפת אל ההתפלגות הנורמלית (משפט הגבול המרכזי). בתרגול ראינו כי בדרך כלל התפלגות הדגימה עבור מדגם הגדול מ-30 הינה מספיק קרובה להתפלגות נורמלית ואכן בגרף התוצאות שלנו ניתן לראות זאת, כלומר אין הבדל משמעותי כש-$n=1000$ וכש-$n=30$.

סעיף 4: בחרנו את הסטטסטי min

גודל המדגם	התפלגות הדגימה של הממוצע
n=3	
n=30	
n=1000	

שאלה 2:

סעיפים 1 +2 : בקובץ R

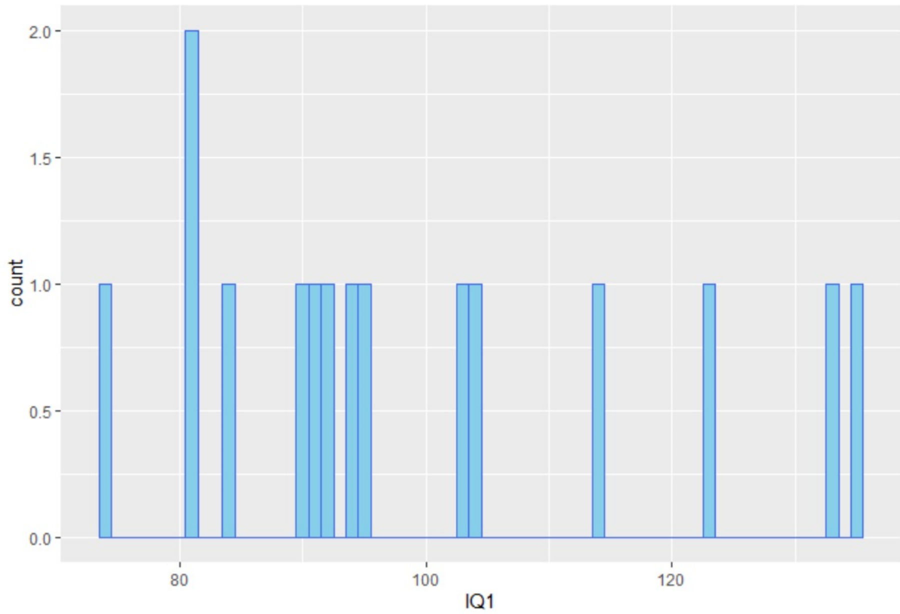
סעיף 3 : הממוצע של המשתנה החדש :

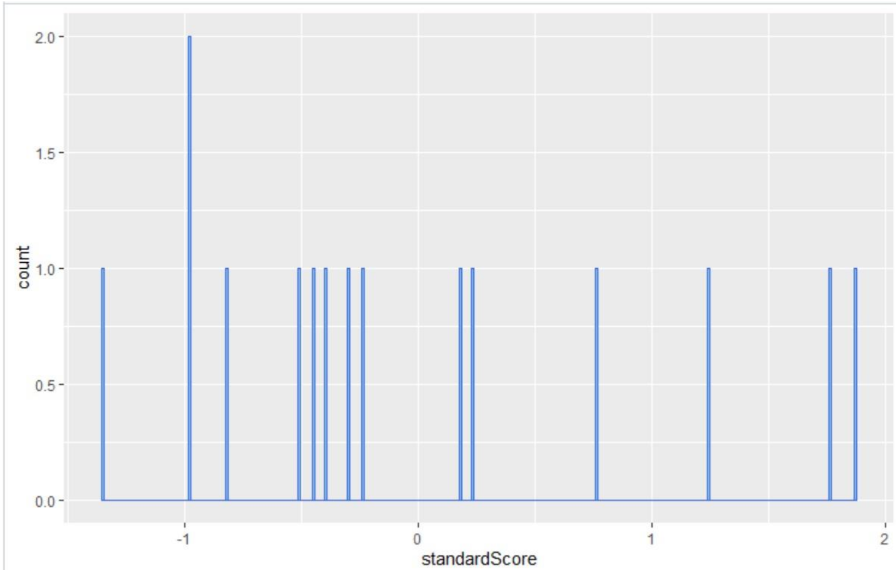
```
> meanz
[1] 3.053113e-16
```

סטיית התקן של המשתנה החדש :

```
> sdz
[1] 1
```

סעיף 4+5:





ניתן לראות כי הגרפים שקיבלנו דומים מאוד, שכן ביצענו שינוי קבוע על כל ציון כדי להפוך אותו לציון התקן, ולכן נקבל התאמה בין שני הגרפים, ודמיון רב בין ההתפלגויות. ההבדל בין הגרפים הוא בערכי ציר

ה-x, בגרף המציג את התפלגות הציונים המקוריים – ישנה התייחסות לציונים עצמם בין 74 ל135, ואילו בגרף המציג את התפלגות ציוני התקן ניתן לראות את התפלגות הציונים המציינים את המרחק של כל ציון ממוצע הציונים ביחידות של סטית התקן.

שאלה 3:

1. עוזר המחקר הוריד שלוש תצפיות קיצוניות ולכן שונות התפלגות הדגימה תגדל.
לא בהכרח נכון - אם חצי מהתצפיות היו זהות לתצפית הגבוהה ביותר והחצי השני היה זהה לתצפית הנמוכה ביותר היינו מקבלים כי השונות תישאר זהה לאחר הורדת שלוש התצפיות הקיצוניות. על כן, לפי הנוסחה לחישוב שונות התפלגות הדגימה, נקבל כי שונות התפלגות הדגימה תגדל ולא תקטן שכן השונות נותרה זהה אך n קטן ב-3 .
מנגד, אם ניקח מקרה שבו 30 התצפיות האמצעיות הן ממוצע כל התצפיות ושלוש התצפיות הקיצוניות במרחק שווה מהממוצע. נקבל כי השונות של 30 הדגימות הנותרות הינה 0 ולכן שונות התפלגות הדגימה קטנה כאשר הורדנו את התצפיות הקיצוניות.
2. עוזר המחקר הוריד את התצפית הגבוהה ביותר ואת התצפית הנמוכה ביותר ולכן ממוצע המדגם ישתנה.
לא בהכרח נכון – אם מרחק התצפיות הגבוהה ביותר והנמוכה ביותר שונה מהממוצע נקבל כי ממוצע המדגם ישתנה, אך אם המרחק מהממוצע שווה אז הממוצע לא ישתנה.
3. עוזר המחקר הוריד את התצפית הגבוהה ביותר ואת התצפית הנמוכה ביותר ולכן שונות המדגם תישאר זהה.
לא בהכרח נכון – השונות הינה סיכום ריבועי המרחקים מהממוצע לחלק בגודל המדגם. נוכל לדוגמא לקחת מקרה שבו 30 התצפיות האמצעיות הן ממוצע כל התצפיות והתצפית הנמוכה ביותר והגבוהה ביותר במרחק שווה מהממוצע. כשנוריד את התצפיות האלו נקבל כי הממוצע לא השתנה, אך השונות הינה 0 כי כל יתר התצפיות שוות לממוצע. מצד שני, הטענה היתה נכונה אם חצי מהתצפיות היו זהות לתצפית הגבוהה ביותר והחצי השני זהות לתצפית הנמוכה ביותר.
4. עוזר המחקר הכפיל כל תצפית של המשתנה ב2 ולכן תוחלת המשתנה תגדל פי חמש.
לא נכון – לפי לינאריות התוחלת התוחלת תגדל פי 2 ולא פי 5.
5. עוזר המחקר החסיר מכל תצפית 10 נקודות ולכן ממוצע המדגם ישאר אותו דבר.
לא נכון – מהגדרת הממוצע, כאשר מפחיתים ערכים הממוצע גם הוא קטן בהתאם :
$$\overline{X'} = \frac{\sum_{i=1}^{n} x_i - 10}{n} = \frac{(\sum_{i=1}^{n} x_i) - (10n)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - 10 = \bar{X} - 10$$
6. עוזר המחקר החסיר מכל תצפית 10 נקודות ולכן שונות המדגם תשתנה.
לא נכון – לפי תכונות השונות : מתקיים : var(x)=var(a+x) עבור a∈R.
7. עוזר המחקר הכפיל כל תצפית של המשתנה ב2, הדבר ישפיע על הממוצע ועל שונות המדגם, אך השונות תושפע מכך פחות.
לא נכון – לפי הנוסחאות שראינו הממוצע יגדל פי 2 בעוד שהשונות תגדל פי 4 שכן :
var(2x)=4var(x).

שאלה 4

$\sigma = sd\ of\ population$

$n = sample\ size$

1. כאשר האוכלוסייה שממנה נלקח המדגם מתפלגת נורמלי, התפלגות הדגימה בהכרח נורמלית .
נכון – התפלגות הדגימה היא התפלגות תיאורטית של דגימה מתוך האוכלוסייה אינסוף פעמים, מדגמים בלתי תלויים בגודל n . ולכן אם ההתפלגות המקורית נורמלית, בהכרח גם התפלגות הדגימה תהיה כזאת.
2. התוחלת של התפלגות הדגימה זהה לתוחלת של האוכלוסיה ללא תלות בגודל n.
נכון - התפלגות דגימה היא התפלגות תאורטית של אינסוף המדגמים שניתן לבנות בגודל n. אם נבנה התפלגות דגימה על בסיס אינסוף מדגמים בגודל n ואז נחשב את תוחלת התפלגות הדגימה נקבל כי תוחלת ההתפלגות תהיה שווה לממוצע האוכלוסייה ממנה נלקחו המדגמים.
3. סטיית התקן של התפלגות הדגימה שווה ל n/σ. **לא נכון** – סטית התקן של התפלגות הדגימה שווה ל- $\frac{\sigma}{\sqrt{n}}$ כאשר n שווה לגודל המדגמים שנדגמו.
4. לפי משפט הגבול המרכזי, התפלגות הדגימה לא בהכרח נורמלית אלא כתלות בגודל של n.
נכון- משפט הגבול המרכזי מגדיר כי התפלגות הדגימה תהיה נורמלית, גם במצב בו ההתפלגות המקורית באוכלוסייה אינה נורמלית אם גודל המדגם מספיק גדול, כאשר הקונבנציה היא שמספיק גדול = 30.

שאלה 5

התשובה הנכונה היא 3- שתיהן טועות. רונה טועה מכיוון שהפונקציה qnorm מקבלת הסתברות ומחזירה את הציון המתאים לה- אבל דנה רוצה לגלות מה ההסתברות לכך שתתקבל בובה במשקל 0.46 או פחות.
דנה טועה משום שהפונקציה dnorm מקבלת ערך x ונתונים של התפלגות נורמלית ומחשבת את ההסתברות לקבל ערך ספציפי זה מתוך ההתפלגות הנורמלית. לעומת זאת, דנה מעוניינת לבדוק טווח של ערכי משקל, 0.46 ומטה.

כדי לענות על שאלתה של דנה, עלינו להשתמש בפונקציה pnorm שמקבלת ערך ומחשבת את ההסתברות לקבלת ערך זה, או ערך קטן ממנו.

```
> q5 = pnorm(0.46, 0.5, sqrt(0.04))
> q5
[1] 0.4207403
>
```