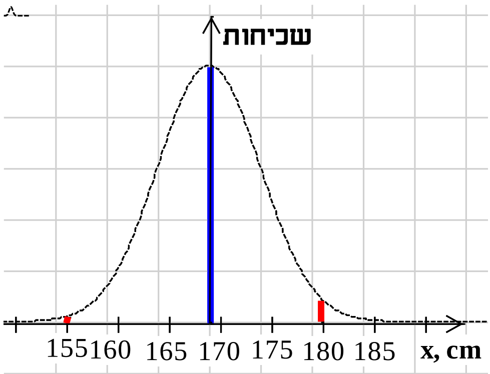
במקרים רבים בחיי היום-יום, גודל אקראי, שמאפיין תהליך מסוים, מקבל את רוב הערכים הקרובים לאיזה ערך מרכזי שהוא, ולעתים נדירות – ערכים הרחוקים ממנו.

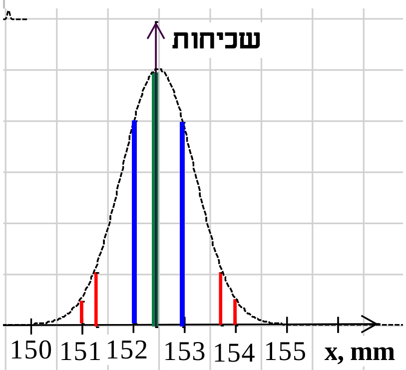


לדוגמה: מדדו את גובה תלמידי שכבת י׳ בבית ספר מסוים. רוב המדידות בקרב 120 תלמידי השכבה היו סביב 168 ס״מ, ורק בודדים היו מעל 180 או מתחת ל- 155 ס״מ.

גם מדידות משקל התלמידים, או מידת הנעליים מראות התפלגות דומה של נתונים.

ההתפלגות של משתנה אקראי מסוג זה מכונה התפלגות **נורמלית**.

הערה: לא כל הגדלים האקראיים מתנהלים כך: למשל, אם נמדוד או נעריך את משקל כל העצמים שסביבנו, מחיידקים עד לכוכבים רחוקים, נמצא שהוא יכול לקבל כל ערך שהוא, ממיליונית גרם עד למיליארדי מיליארדים של טון, ללא העדפה לערך מרכזי מסוים.



☜ **התפלגות נורמלית** מופיעה בכל תהליכי מדידה, שבהם מעורבים גורמי שגיאה רבים.

לדוגמה: עשרה תלמידים קיבלו משימה למדוד אורך של עיפרון בדיוק גבוה עד כמה שאפשר.

הם קיבלו עיפרון ומדדו את אורכו עם הסרגלים הנמצאים ברשותם.

התוצאות שהתלמידים רשמו היו:

153.5, 154.1, 152.6, 152.7, 152.1, 152.5, 152.4, 151.3, 150.9, 152.3 ס"מ.

מדוע היו התוצאות שונות?

מכיוון ש:

א. הסרגלים היו שונים ;   ב. התלמידים לא היו מיומנים באותה מידה ;

ג. תנאֵי התאורה היו שונים (חלק מהתלמידים ישבו ליד חלון וחלק – רחוק ממנו).

בכל התהליכים שבהם מעורבים גורמים אקראיים רבים מופיעה התפלגות נורמלית.

תכונות הגרף של התפלגות נורמלית

מהשוואת שני הגרפים שבעמוד הקודם אפשר להסיק שצורתם דומה : לעקומה המתארת את ההתפלגות הנורמלית צורה של **פעמון.**

לעתים היא מכונה גם **עקומת גאוס** (או גאוסיאן) על שמו של המתמטיקאי הדגול, **פרידריך קָרְל גאוס** (1777-1855-), שתרם רבות לתורת ההסתברות.

תכונות הגרף של התפלגות נורמלית :

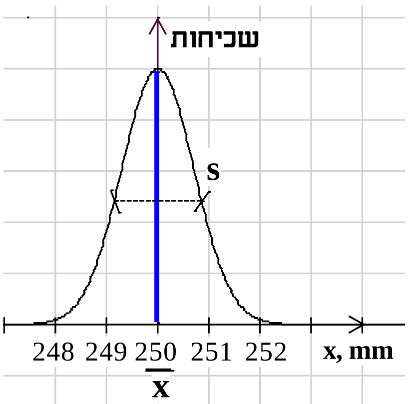
א. הגרף הוא סימטרי יחסית לערך המרכזי.

ב. הערך המרכזי הוא הנתון **השכיח** בין כל הנתונים.

ג. הערך המרכזי הוא **הממוצע** ו**החציון** של כל הנתונים.

כלומר, בהתפלגות נורמלית, **השכיח שווה לממוצע ולחציון**.

מאפייני הגרף - הממוצע וסטיית התקן



בניסוי מדידה שני קיבלו התלמידים משימה למדוד את אורך הספר. התוצאות שקיבלו היו מקובצות סביב מספר אחר - 250 מ״מ, אולם **פיזור התוצאות**, המתבטא **ברוחב** העקומה, היה דומה לזה שהתקבל במדידת אורך העיפרון :

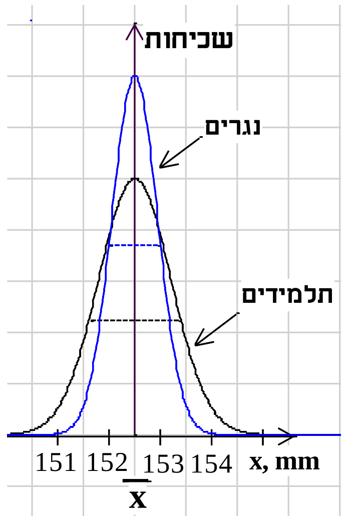
האופי של שגיאות המדידה לא השתנה, לכן רוחב העקומה, המתבטא בגודל של סטיית התקן s לא השתנה; השתנה רק ה**מיקום** של מרכז העקומה המתבטא בממוצע $\overline{x}$.

בניסוי שלישי, עשרה נגרים ותיקים התבקשו למדוד את אורך העיפרון .

התוצאות שלהם היו כדלקמן :

152.1, 152.7, 152.5, 152.9, 151.8, 153.2,

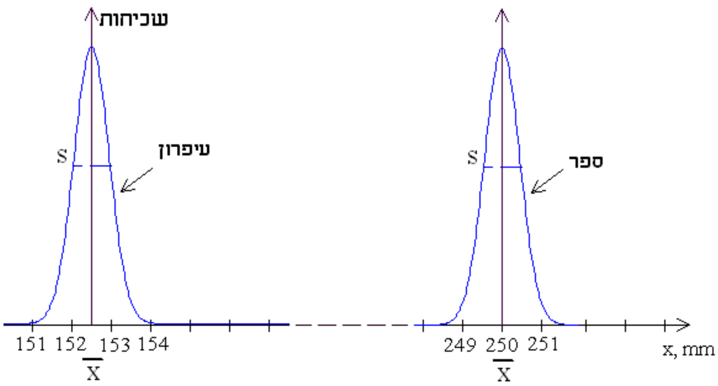
152.4, 152.8, 153.7, 151.9



מהשוואת שני הגרפים, של התלמידים ושל הנגרים, רואים, שמרכזם מתלכד אולם רוחבם שונה: פיזור הנתונים של הנגרים היה קטן מזה של התלמידים.

בניסוי אחר, קיבלו הנגרים את המשימה למדוד את אורך הספר ; שני הגרפים שלפניכם מתארים את התוצאות שקיבלו הנגרים :

מרכזי הגרפים ממוקמים במקומות שונים (השווים לממוצע של אורך העיפרון והספר, בהתאמה), ואילו רוחב העקומה שווה לשני הגרפים (מכיוון ששגיאות המדידה לא תלויות באורך הנמדד).



לסיכום: א. ממוצע הנתונים קובע את מרכז הגרף,

או במילים אחרות : שיעור מרכז הגרף שווה לממוצע הנתונים.

ב. רוחב העקומה נקבע על-ידי סטיית התקן של כל הנתונים.

שיעור y של כל נקודה x על גרף ההתפלגות מבטא את השכיחות היחסית של הופעת הנתון x.

אולם, **שכיחות הופעת הנתון שערכו הוא <u>בדיוק x</u> שווה לאפס!**

הדוגמה שלהלן מוכיחה:

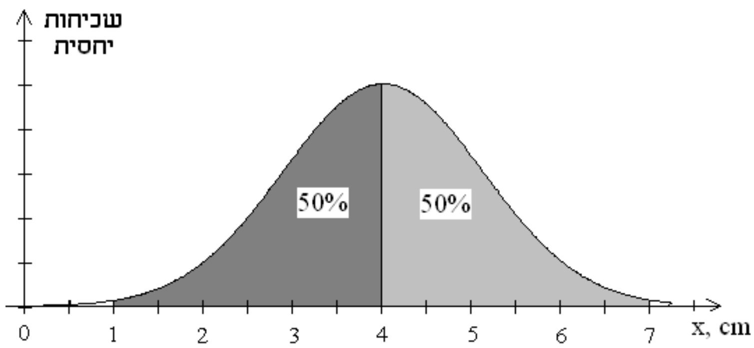
אורך מלפפונים בזמן הקטיף נע בין 3 ל- 5 ס״מ.

מה השכיחות של הופעת מלפפון שאורכו 4 ס״מ בדיוק?

◄ מספר כל הערכים האפשריים של אורך מלפפון בגבולות שבין 3 ל- 5 הוא **אינסופי:**
הרי יכולים להיות ערכים כמו 4.001, 4.000000235, 3.9999875003222 ס״מ וכד׳, כלומר, יכול להופיע **כל מספר** רציונלי (מהסוג של שבר פשוט $\frac{3}{25}$) או אפילו אי-רציונלי (כמו $\sqrt{2}$). יש אינסוף מספרים כאלה בכל קטע של ציר מספרים.
לכן השכיחות של הופעת נתון מסוים (נגיד, 4.00 ס״מ) שווה ל- 0 ($\frac{1}{\infty} = 0$).

<u>ובכן, אין טעם בשאלה מה הסיכוי (ההסתברות) להופעת ערך מסוים מתוך אינסוף ערכים אפשריים של המשתנה בעל התפלגות נורמלית.</u>

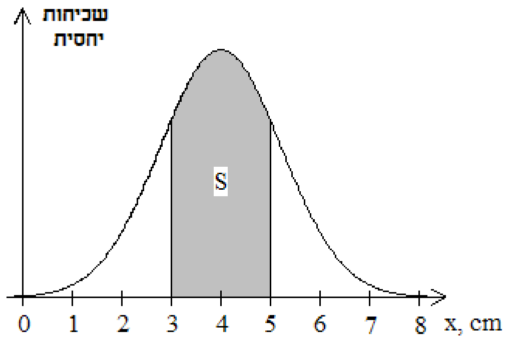
יחד עם זאת ברור, שאם אורכי המלפפונים מתפלגים נורמלית סביב ממוצע של 4 ס״מ, הסיכוי שאורך המלפפון הנבחר באקראי יהיה גדול מ- 4 ס״מ (כלומר, $4 < x < \infty$) יהיה שווה לסיכוי שאורכו יהיה קטן מ- 4 ס״מ ($0 < x < 4$).



כלומר, יש משמעות רק לשכיחות הופעת הנתון <u>בתוך גבולות מסוימים</u>,

לדוגמה: שכיחות הופעת מלפפון שאורכו בין 3.8 ל- 4.2 ס״מ ($3.8 < x < 4.2$).

אפשר להוכיח **ששכיחות הופעת המשתנה המפולג בין גבולות מסוימים שווה לשטח שבין עקומת גרף ההתפלגות לבין ציר x בגבולות אלה.**



כך, למשל, שכיחות הוצאת מלפפון שאורכו בין 3 ל- 5 ס״מ מערימה גדולה של מלפפונים שווה לשטח S המסומן בגרף ההתפלגות של אורכי המלפפונים.

מכיוון ששכיחות הוצאת מלפפון בכל אורך שהוא $(0 < x < \infty)$ שווה ל- 1 (הרי בטוח שבכל מקרה נוציא איזה מלפפון שהוא!), אפשר להסיק, **שהשטח מתחת לגרף בכל ציר המספרים שווה ל- 1**.

בהסתמך על מסקנה זו ועל תכונת הסימטריות של התפלגות נורמלית, אפשר להסביר את הגרף שבדף הקודם (שכיחות הוצאת מלפפון שאורכו בין 0 ל- 4 ס״מ שווה ל- 0.5): $P(0 < x < 4) = P(4 < x < \infty)$ (בגלל שהגרף הוא סימטרי),

ו- $P(0 < x < 4) + P(4 < x < \infty) = 1$,

כאשר **פירוש הביטוי $P(a < x < b)$ הוא ההסתברות שהמשתנה האקראי x מקבל ערכים בתחום בין a ל- b.**

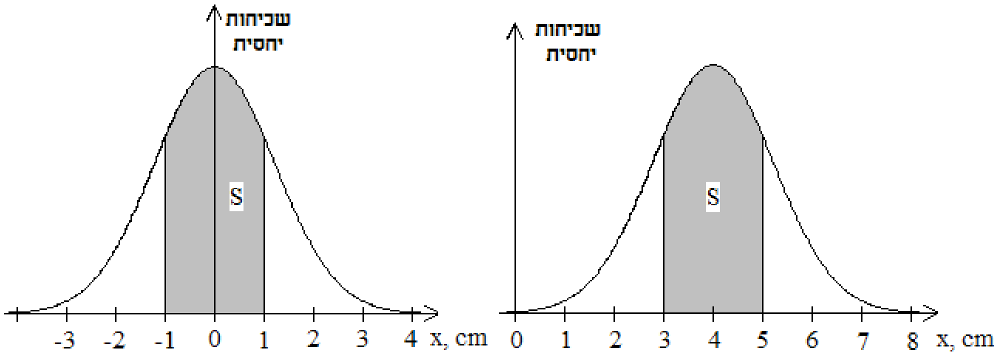
כדי לחשב את **מספר** המלפפונים N, שאורכם נמצאים בגבולות בין a ו- b, עלינו לדעת את כמות כל המלפפונים M ואת ההסתברות שאורכם יהיה בין הגבולות הנתונים:

$$N(a < x < b) = M \cdot P(a < x < b) \qquad (1)$$

נניח שידועה כמות כל המלפפונים: $M = 1000$ והגבולות $a = 3.5$ ו- $b = 4.5$ ס״מ.

כיצד לחשב את ההסתברות $P(3.5 < x < 4.5)$?

למדנו כבר שהסתברות זו שווה לשטח שמתחת לעקומת הגרף בקטע נתון.
האם שטח זה תלוי במיקומו של הממוצע?
מהשוואת שני הגרפים אפשר לראות שמכיוון שהגרף הוא סימטרי יחסית לממוצע, השטח הנדרש תלוי רק במרחק הגבולות מהמרכז:



כלומר, בחישובי השטח, אפשר להחליף את המשתנה x למשתנה חדש x_1:

$$x_1 = x - \bar{x} \qquad (2)$$

ברור ש- $x_1 = 0$ כאשר x שווה לממוצע.

כמובן, לאחר שנמצא את הערכים של x_1, נעבור חזרה למשתנה x:

$$x = x_1 + \bar{x} \qquad (3)$$

אם נמצא שערכי-x_1 נמצאים בתחום $[a, b]$, נוכל למצוא את ערכי- x המתאימים באמצעות (3).

דוגמה: האורך הממוצע של מלפפון שווה ל- 4 ס״מ.
לאחר סדרה של מדידות נמצא של- 90% מהמלפפונים ערכי האורך נמצאים בתחום ±1 ס״מ **סביב הממוצע**. מה אורך המלפפונים מתחום זה?

◄ נרשום נתונים: $\bar{x} = 4,\ a = -1,\ b = 1$

נעבור מהמשתנה x:

$$a = -1: \quad x = -1 + 4 = 3; \quad b = 1: \quad x = 1 + 4 = 5$$

תשובה: ל- 90% מהמלפפונים האורך נמצא בתחום $\underline{3 < x < 5}$ ס״מ. ▷

החלפת המשתנה למשתנה סטנדרטי

ברוב המדידות, סטיית התקן נמצאת ביחס ישר לגודל הממוצע: כשמודדים אורך של עיפרון, שגיאת המדידה היא ± 0.5 מ״מ, במדידת אורך של שולחן, השגיאה היא ± 0.5 ס״מ, במדידת אורך של שדה כדורגל השגיאה היא ± 0.5 מ׳.

אי-לכך, הגיוני למדוד את התפלגות הנתונים לא ביחידות **המוחלטות** (מ׳, ס״מ, ק״ג וכד׳) אלא **ביחידות יחסיות של סטיית תקן**.

לדוגמה, נתון:

האורך הממוצע של מלפפונים הוא 4 ס״מ, סטיית התקן היא 4 מ״מ.

מה אחוז המלפפונים שאורכם נמצא בתחום בין 3.6 ל- 4.4 ס״מ?

את אותה שאלה אפשר לנסח אחרת:
מה אחוז המלפפונים שאורכם נמצא בתחום של שתי סטיות תקן סביב הממוצע?

היתרון בניסוח השני הוא בכך, שהוא אינו תלוי בערך הממוצע ובערכה של סטיית תקן בנפרד, אלא ביחס ביניהם.

כלומר, יחידות מידה של הנתונים הופכות מיחידות המוחלטות (מ׳, ס״מ וכד׳) ליחידות של סטיית תקן.

שיטה זו מאפשרת לתאר תופעות שונות ולחשב מדדים סטטיסטיים עבורן בדרך דומה.

כדי לעבור ליחידות של סטיית תקן, נגדיר משתנה חדש X, המכונה **משתנה סטנדרטי**:

$$X = \frac{x - \bar{x}}{s} \qquad (4)$$

דוגמה 1

ציוני הבחינה בבית-ספר מתפלגים נורמלית, כאשר הציון המוצע הוא 68 וסטיית התקן היא 10. השאלות הינן:

א. מה אחוז התלמידים שציונם גבוה מ- 88?

ב. מה אחוז התלמידים שציונם נמוך מ- 58?

רשמו את נתוני השאלה באמצעות משתנה סטנדרטי.

◄ <u>השלב החשוב ביותר בפתרון כל בעיה מילולית הוא רישום נתונים בשפה אלגברית</u>.

בשאלה הנ"ל נתון: $\bar{x} = 68,\ s = 10$.

צ"ל: $P(88 < x < \infty) = ?$, $P(0 < x < 58) = ?$

נעבור למשתנה הסטנדרטי: $X = \frac{x - 68}{10}$

נמצא את גבולות התחומים החדשים:

$$x = 88: \quad X = \frac{x - 68}{10} = \frac{88 - 68}{10} = 2$$

$$x = 58: \quad X = \frac{x - 68}{10} = \frac{58 - 68}{10} = -1$$

לכן, במונחים של משתנה חדש, עלינו למצוא את <u>$P(2 < X < \infty)$</u> ואת <u>$P(X < -1)$</u>.

דוגמה 2

ציוני הבחינה בבית-ספר מתפלגים נורמלית, כאשר הציון המוצע הוא 68 וסטיית התקן היא 8. בוחרים תלמיד באקראי.

צ"ל: את ההסתברות שציונו יהיה בין 56 ל- 84.

רשמו את נתוני השאלה באמצעות משתנה סטנדרטי.

◄ נרשום תחילה את הנתונים במונחים של ציונים:

$$\bar{x} = 68,\ s = 8\ ,\ P(56 < x < 84) = ?$$

נעבור למשתנה הסטנדרטי: $X = \frac{x - 68}{8}$.

נמצא את גבולות התחום החדשים:

$$x = 56: \quad X = \frac{56 - 68}{8} = -\frac{12}{8} = -1.5$$

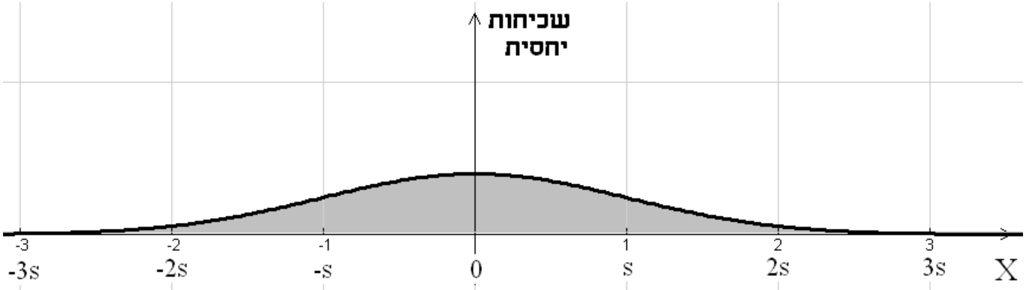
$$x = 84: \quad X = \frac{84 - 68}{8} = 2$$

תשובה: צ"ל את <u>$P(-1.5 < X < 2)$</u> ▷

חישוב השטח הכלוא מתחת לגרף התפלגות נורמלית

כאשר משתנה אקראי x מתפלג נורמלית סביב הממוצע $\bar{x}$ עם סטיית תקן s.

המשתנה X, המבוטא ביחידות של סטיית תקן: $X = \frac{x - \bar{x}}{s}$ מתפלג גם הוא נורמלית סביב הממוצע 0 וסטיית תקן 1. התפלגות מסוג זה מכונה **התפלגות נורמלית סטנדרטית**.



מגרף ההתפלגות הסטנדרטי אפשר לראות ש:

א. הגרף הוא סימטרי יחסית ל- $X = 0$;

ב. השטח הכלוא בין עקומת הגרף לבין ציר x בגבולות שבין $X = -3$ ל- $X = 3$ שווה בקירוב לשטח של משבצת אחת, כלומר, ל- 1. אפשר להוכיח שהשטח הכלוא מתחת לעקומה בין הגבולות $[-\infty, +\infty]$ שווה בדיוק ל- 1.

ג. הסתברות הופעת הערכים המרוחקים מהמרכז ביותר מ- 3 סטיות תקן היא זניחה, כלומר, רוב הערכים מרוכזים בתחום של $\pm 3s$.

הערה: בסטטיסטיקה מקובל לסמן את סטיית התקן באות σ (סיגמה), לכן לעתים קוראים לכלל זה "**כלל של שלוש סיגמה**".

את גודלם המדויק של השטחים הכלואים בין גבולות שונים אפשר לחשב באמצעות חשבון אינטגרלי, ותוצאות החישובים מובאות בטבלה הזאת:

גבול ימין (ביחידות של סטיית תקן s)	0.5	1	1.5	2	2.5	3
שטח S(0<X<b)	0.192	0.341	0.433	0.477	0.494	0.498

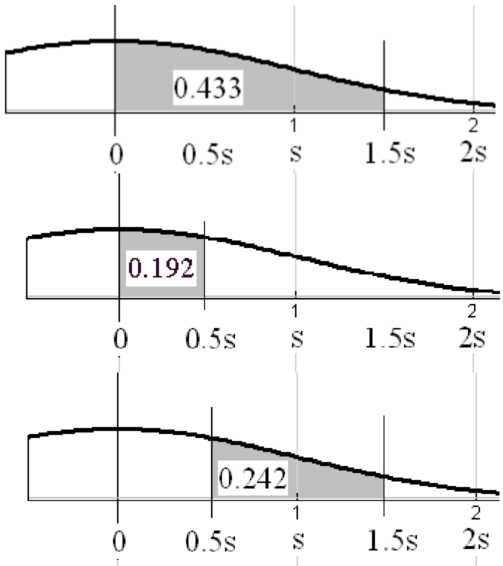
הנתונים שבטבלה מבטאים את השטחים הכלואים בין $X = 0$ (מרכז ההתפלגות) לבין גבול שמימין עד ל- $X = 3s$, במדרגות של $0.5s$.

כדי לחשב שטח במקרה שהגבול השמאלי a אינו שווה ל- 0: $S(a < X < b)$, $a \neq 0$, מחסירים מהשטח $S(0 < X < b)$ המתאים לגבול הימיני את השטח $S(0 < X < a)$, המתאים לגבול השמאלי:

$$S(a < X < b) = S(0 < X < b) - S(0 < X < a) \qquad (5)$$

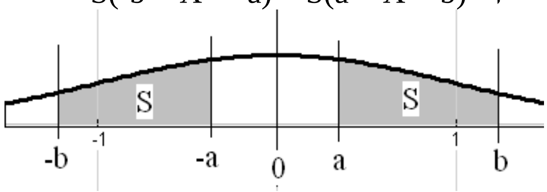
לדוגמה, כדי לחשב את השטח $S(0.5 < X < 1.5)$, נשתמש בערכי הטבלה ונקבל:

$$S(0.5 < X < 1.5) = S(0 < X < 1.5) - S(0 < X < 0.5) =$$
$$= 0.433 - 0.192 = 0.241$$

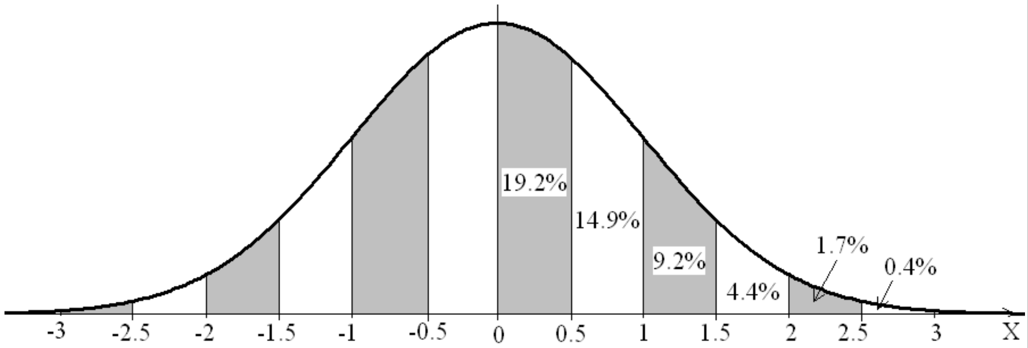


אם אחד או שני גבולות התחום הם שליליים (נמצאים משמאל למרכז), מנצלים את העובדה שהגרף הוא סימטרי יחסית ל- 0, ולכן השטח שמשמאל שווה לשטח הסימטרי שמימין: $S(-a < X < 0) = S(0 < X < a)$

ו- $S(-b < X < -a) = S(a < X < b)$



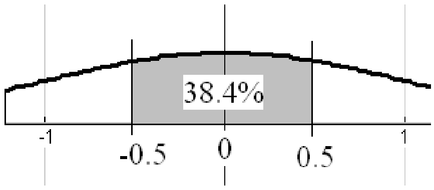
בשיטה זו נבנה את הגרף, שבו מסומנים שטחים (באחוזים) הכלואים מתחת לעקומת גרף ההתפלגות הנורמלית הסטנדרטית, בתחום שמ- 0 עד 3 במדרגות של 0.5. גרף זה, כמו הטבלה, מאפשר להעריך ללא חישובים מסובכים את השטח הכלוא מתחת לעקומה בקטעים שרוחבם מספר שלם של חצאֵי סטיית התקן.



דוגמה 3

למשתנה אקראי X התפלגות נורמלית סטנדרטית.

איזה אחוז מהנתונים שאותם מייצג X נמצאים בתחום של ±0.5?



נשתמש בנתוני הטבלה ובעובדה ששטח בתחום השלילי של X שווה לשטח סימטרי בתחום החיובי:

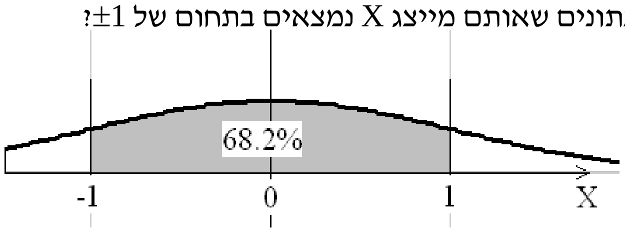
$$S(-0.5 < X < 0.5) = S(-0.5 < X < 0) + S(0 < X < 0.5) =$$

$$= 2 \cdot S(0 < X < 0.5) = 2 \cdot 0.192 = 0.384 = 38.4\%$$

דוגמה 4

למשתנה אקראי X התפלגות נורמלית סטנדרטית.

איזה אחוז מהנתונים שאותם מייצג X נמצאים בתחום של ±1?



$$S(-1 < X < 1) = S(-1 < X < 0) + S(0 < X < 1) =$$

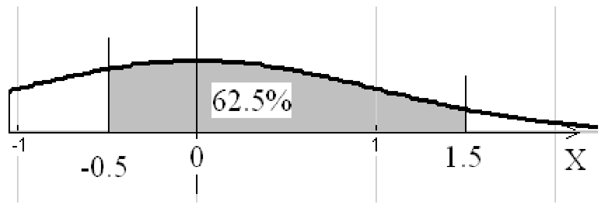
$$= 2 \cdot S(0 < X < 1) = 2 \cdot 0.341 = 0.682 = 68.2\% \quad \triangleleft$$

דוגמה 5

למשתנה אקראי X התפלגות נורמלית סטנדרטית.

איזה אחוז מהנתונים שאותם מייצג X נמצאים בתחום של $-0.5 < X < 1.5$?

◄



$$S(-0.5 < X < 1.5) = S(-0.5 < X < 0) + S(0 < X < 1.5) =$$

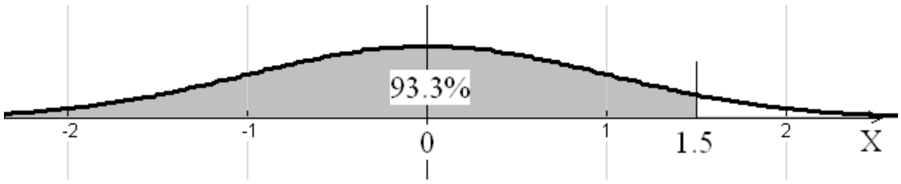
$$= S(0 < X < 0.5) + S(0 < X < 1.5) = 0.192 + 0.433 = 0.625 = 62.5\% \quad \triangleleft$$

דוגמה 6

למשתנה אקראי X התפלגות נורמלית סטנדרטית.

מה אחוז הנתונים שערכם קטן מ- 1.5?

◄



מכיוון שהגרף הוא סימטרי יחסית לממוצע $X = 0$, השטח הכלוא מתחת לעקומה בתחום $-\infty < X < 0$ שווה ל- 0.5. לכן השטח המבוקש שווה ל-

$$S(-\infty < X < 1.5) = S(-\infty < X < 0) + S(0 < X < 1.5) =$$

$$= 0.5 + 0.433 = 0.933 = 93.3\% \quad \triangleleft$$

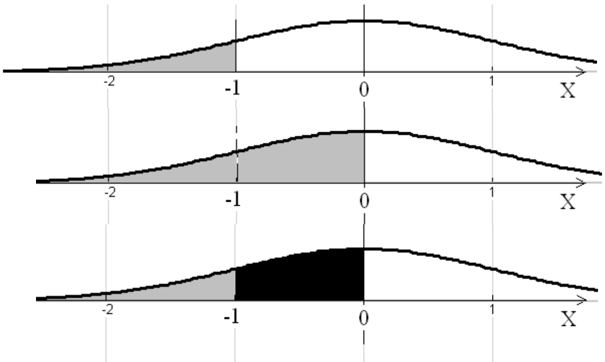
דוגמה 7

למשתנה אקראי X התפלגות נורמלית סטנדרטית.

מה אחוז הנתונים שערכם קטן מ- 1-?

◄ השטח הכלוא מתחת לעקומת ההתפלגות בתחום $X < -1$ שווה להפרש השטחים:

$$S(-\infty < X < -1) = S(-\infty < X < 0) - S(-1 < X < 0)$$



נציב את נתונים מהטבלה:

$$S(-\infty < X < -1) = 0.5 - S(0 < X < 1) = 0.5 - 0.341 = 0.159 = 15.9\%$$

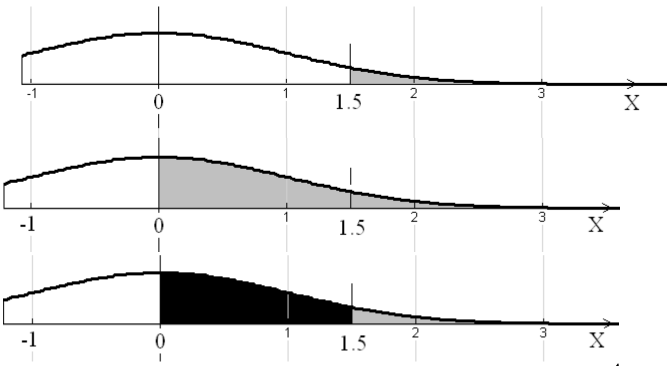
דוגמה 8

למשתנה אקראי X התפלגות נורמלית סטנדרטית.

מה אחוז הנתונים שערכם גדול מ- 1.5?

◄ השטח הכלוא מתחת לעקומת ההתפלגות בתחום $X > 1.5$ שווה להפרש השטחים:

$$S(1.5 < X < \infty) = S(0 < X < \infty) - S(0 < X < 1.5)$$



מציבים את נתוני הטבלה:

$$S(1.5 < X < \infty) = 0.5 - 0.433 = 0.067 = 6.7\%$$ ◁

תרגילים

1. מצאו את השטח הכלוא שבין עקומת ההתפלגות הנורמלית הסטנדרטית לבין ציר – X בתחום $a < X < b$ הזה:

א) $a = 1, b = 1.5$ ב) $a = 1, b = 2$ ג) $a = 1, b = 2.5$

ד) $a = 1, b = 3$ ה) $a = 1, b = 1$ ו) $a = 1, b = \infty$

2. משתנה אקראי X מייצג את ההתפלגות הנורמלית הסטנדרטית של סדרת נתונים מסוימת. מצאו את אחוז הנתונים, שערכם גדול מ:

א) 0 ב) 1 ג) 1.5 ד) 2.5

ה) -0.5 ו) -1.5 ז) -3 ח) $-\infty$

3. משתנה אקראי X מייצג את ההתפלגות הנורמלית הסטנדרטית של סדרת נתונים מסוימת. מצאו את אחוז הנתונים, שערכם קטן מ:

א) 0 ב) 1 ג) 1.5 ד) 2.5

ה) -0.5 ו) -1.5 ז) -3 ח) $-\infty$

4. מצאו את השטח הכלוא בין עקומת ההתפלגות הנורמלית הסטנדרטית לבין ציר – X בתחום $a < X < b$:

א) $a = -1, b = 1.5$ ב) $a = -2, b = 2$ ג) $a = -3, b = 2.5$

ד) $a = -2.5, b = -1.5$ ה) $a = -\infty, b = 1$ ו) $a = -\infty, b = \infty$

§38 חזרה ממשתנה סטנדרטי לנתונים מקוריים

כדי לחזור ממשתנה סטנדרטי לנתונים מקוריים, משתמשים בנוסחה (4) מהסעיף הקודם:

$$X = \frac{x - \bar{x}}{s} \quad (4)$$

ונחלץ ממנה את המשתנה המקורי:

$$x = \bar{x} + X \cdot s \quad (1)$$

☜ שתי הנוסחאות (4) ו- (1) וגרף ההתפלגות הנורמלית הסטנדרטית, המתואר בסעיף הקודם מאפשרות לפתור כל בעיה של התפלגות נורמלית.

שלבי הפִּתרון

א. בהסתמך על הנתונים (או על הנאמר בשאלה), מוצאים את הממוצע $\bar{x}$;

ב. מוצאים את סטיית התקן s;

ג. מעבירים, באמצעות נוסחה (4), למשתנה סטנדרטי X את נתוני הבעיה (בדרך כלל, את גבולות התחומים);

ד. מוצאים באמצעות הגרף או הטבלה של התפלגות נורמלית סטנדרטית את ערכי המשתנה הסטנדרטי X העונים לשאלות הבעיה;

ה. עוברים למשתנים המקוריים באמצעות הנוסחה (1).

דוגמה 1 (מאגר, שאלה מס. 1). ציוני מבחן שכבתי במתמטיקה מתפלגים נורמלית, כאשר הציון הממוצע הוא 68, וסטיית התקן מהממוצע היא 10.

א. מה אחוז התלמידים שציוניהם גבוהים מ- 88?

ב. מה אחוז התלמידים שציוניהם בין מ- 58 ל- 88?

ג. מספר התלמידים שציוניהם בין מ- 58 ל- 88 הוא 902. כמה תלמידים, סך הכול, נגשו למבחן?

◄ א. נרשום את נתוני הבעיה: $P(88 < x) = ?$

נחליף את המשתנה לסטנדרטי ונרשום באמצעותו את הנתונים:

$$\bar{x} = 68,\ s = 10$$

צ"ל: $P(2 < X) = ?$. $X = \frac{x - 68}{10},\ b = 88 \rightarrow B = \frac{88 - 68}{10} = 2$

משתמשים בגרף ומוצאים שאחוז המקרים שעבורם X גדול מ- 2 הוא:

$$1.7\% + 0.4\% = \underline{2.1\%} \quad \triangleleft$$

תשובה: 2.1% ▷

הערה: אין צורך לחזור למשתנה x המקורי: אחוז הנתונים לא ישתנה, ישתנה רק הניסוח: במקום "אחוז המקרים שעבורם $X > 2$ שווה ל- 2.1%" יהיה: "אחוז התלמידים שציוניהם $x > 88$ שווה ל- 2.1%".

ב. נרשום נתונים באמצעות המשתנה הסטנדרטי:

$$X = \frac{x - 68}{10},\ a = 58 \rightarrow A = \frac{58 - 68}{10} = -1$$

צ"ל: $P(-1 < X < 2)$.

נשתמש בגרף או בטבלה ונמצא:

$$P(-1 < X < 2) = P(-1 < X < 0) + P(0 < X < 2) =$$
$$= 14.9\% + 19.2\% + 47.7\% = 81.8\% \quad \triangleleft$$

ד. על-פי הנתון והנמצא בפתְרון הסעיף הקודם, 902 תלמידים מהווים 81.8% מכלל התלמידים.

אם נסמן את מספר התלמידים ב- N, אפשר לרשום:

$$\frac{902}{N} = 81.8\% = 0.812$$

מכאן נמצא N:

◁ $$902 = N \cdot 0.818, \; N = \frac{902}{0.818} \approx 1100$$

☜ *כדי להתמודד בהצלחה בבעיות בסטטיסטיקה, יש לרשום את נתוני הבעיה ולשרטט את סקיצת הגרף המתאים.*

דוגמה 2 (מאגר, שאלה מס. 8).

הגובה של צמחי נוי מסוג מסוים מתפלג נורמלית עם ממוצע של 65 ס"מ.

ידוע שרבע מהצמחים מגיעים לגובה העולה על 75 ס"מ.

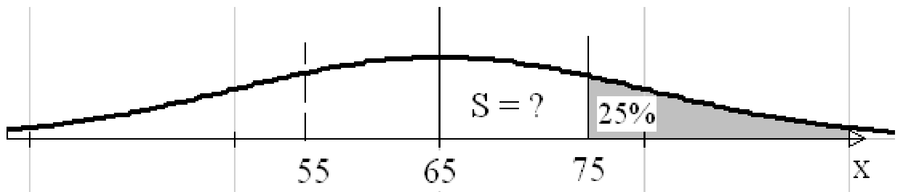
א. מה ההסתברות לבחור באקראי צמח נוי שגובהו מעל הממוצע, אך נמוך מ- 75 ס"מ?

ב. מה אחוז הצמחים שגובהם נמוך מ- 55 ס"מ? נמקו.

◄ א. נרשום את נתוני הבעיה: $\bar{x} = 65, \; P(x > 75) = \frac{1}{4}$

צ"ל: $P(65 < x < 75) = ?$

☜ *במקרים מסוימים ניתן לפתור בעיה בדרכים פשוטות יותר, בהסתמך על תכונות הסימטריה של גרף ההתפלגות. נשרטט סקיצה של הגרף המתאים לנתוני הבעיה:*



מהגרף אפשר להסיק, שמכיוון שהשטח מימינו של הממוצע שווה ל-

$$P(65 < x < \infty) = 0.5$$

אז השטח הנדרש שווה ל-

◁ $$P(65 < x < 75) = P(65 < x < \infty) - P(75 < x < \infty) = 0.5 - 0.25 = \underline{0.25}$$

ב. $P(x < 55) = ?$

מתכונות סימטריה של הגרף רואים שתחום הגבהים $(-\infty < x < 55)$ הוא סימטרי לתחום $(75 < x < \infty)$ יחסית לממוצע $(x = 65)$, לכן גם השטחים מתחת לעקומות הגרף בתחומים אלה הם שווים, וגם ההסתברויות שוות לבחור צמח מתחומים אלה:

$$\triangleleft \quad P(x < 55) = P(75 < x) = \underline{0.25}$$

☜ <u>במקרים, בהם חסר נתון כלשהו, יש לרשום אותו באות, ולהמשיך בפתרון הבעיה. בדרך כלל, הנתון החסר "מתגלה" מתוך נתוני הבעיה.</u>

דוגמה 3 (המאגר, שאלה מס. 19).

תנובת החלב היומית של פרות מתפלגת נורמלית.

ידוע ש- 16% מהפרות מניבות פחות מ- 20 ליטר ביום, ו- 2% מהפרות מניבות פחות מ- 10 ליטר ביום.

א. חַשבו את הממוצע ואת סטיית התקן של תנובת החלב היומית של הפרות.

ב. מה אחוז הפרות המניבות יותר מ- 30 ליטר ליום?

ג. מה אחוז הפרות המניבות יותר מ- 15 ליטר ליום?

◄ א. נסמן ב- x את התנובה היומית של פרה, ונרשום את הנתונים:

$$P(x < 20) = 0.16, \quad P(x < 10) = 0.02$$

$$\bar{x} = ? \qquad s = ?$$

נעבור למשתנה סטנדרטי X: $X = \frac{x - \bar{x}}{s}$.

כאשר גבול התחום של משתנה x שווה ל- $a_1 = 20$,

הגבול המתאים של X שווה ל- $A_1 = \frac{20 - \bar{x}}{s}$,

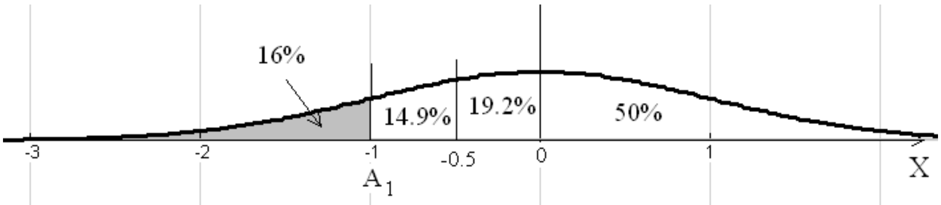
וכאשר גבול התחום של משתנה x שווה ל- $a_2 = 10$,

הגבול המתאים של X שווה ל- $A_2 = \frac{10 - \bar{x}}{s}$.

נרשום נתונים במונחים של משתנה חדש X:

$$P(X < A_1) = 0.16, \quad P(X < A_2) = 0.02$$

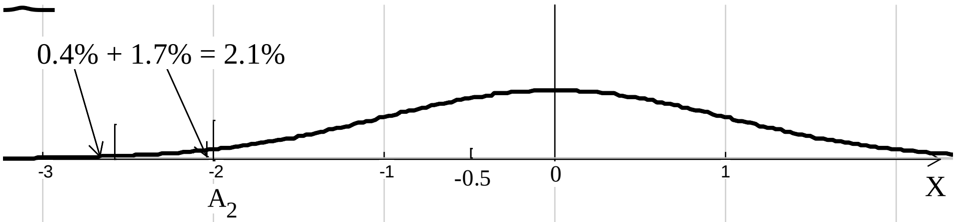
נתבונן בגרף ההתפלגות הנורמלית הסטנדרטית, ונבדוק מה גבולות התחומים שעבורם מתקיימים שוויונות אלה.



נמצא תחילה ערך A_1 כזה שעבורו מתקיים תנַאי $P(X < A_1) = 0.16$.

מהגרף רואים ש- $\underline{A_1 = -1}$ (הבדיקה: השטח מימינו מ- $X = -1$ שווה בערך ל- 84%, כלומר, אחוז הפרות המניבות **יותר** מ- 20 ליטר ליום הוא 84%, לכן אחוז הפרות שמניבות **פחות** מזה שווה ל- 16% $(100\% - 84\% = 16\%)$.

בדומה לכך, מהתבוננות בגרף ובחיבור שטחים, נמצא את A_2 : $\underline{A_2 = -2}$.



נרשום את A_1 ו- A_2 באמצעות $\overline{x}$ ו- s:

$$\frac{20 - \overline{x}}{s} = A_1 = -1$$

$$\frac{10 - \overline{x}}{s} = A_2 = -2$$

נפתור את המערכת:

$$20 - x = -s$$
$$10 - \overline{x} = -2s$$

$$10 - \overline{x} = 2 \cdot (20 - \overline{x}),\quad 10 - \overline{x} = 40 - 2\overline{x},\quad \overline{x} = 30,\ s = 10$$ ◁

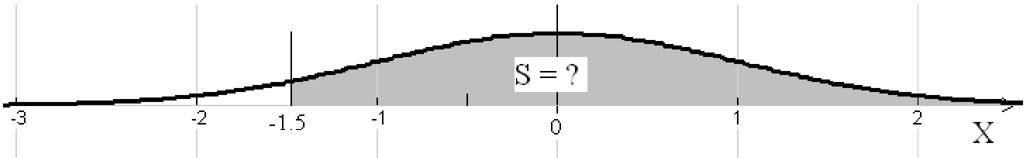
ב. מכיוון שממוצע המניבה היומית שווה ל- 30 ליטר, אחוז הפרות המניבות יותר מ- 30 ליטר שווה ל- 50%. ▷

ג. נעבור למשתנה סטנדרטי ונרשום נתונים:

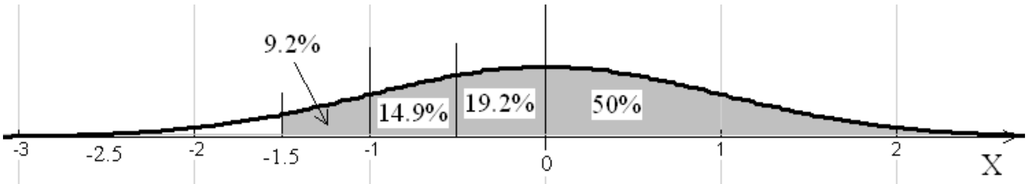
$$X = \frac{x - 30}{10},\ A = \frac{15 - 30}{10} = -1.5$$

צ״ל: $P(X > -1.5)$

נשרטט סקיצה של גרף ההתפלגות, ונסמן את השטח הנדרש:



נשווה עם גרף ההתפלגות הנורמלית הסטנדרטית:



נסכם את השטחים ונקבל תשובה:

◁ $P(X > -1.5) = 9.2\% + 14.9\% + 19.2\% + 50\% = \underline{93.3\%}$

תרגילים

1. סקר של אורך חייהם של אנשים באוכלוסייה בת 50,000 איש הראה כי ממוצע אורך החיים בקבוצה זו הוא 76 שנה, וכי סטיית התקן היא 14 שנה.
בהנחה שהתפלגות השכיחויות היא נורמלית, מִצאו:
א) כמה מאנשי הקבוצה יחיו פחות מ- 48 שנה?
ב) כמה מהם יחיו מעל 90 שנה?
ג) כמה מהם יחיו עד הגיל הנע בין 62 ל- 83 שנה?

2. הגיל הממוצע של 600 הצירים שבבית נבחרים מסוים הוא 62 שנה, וסטיית התקן היא 16 שנה. בהנחה, שהתפלגות הגילים היא נורמלית, מִצאו את מספר הצירים שגילם:
א) למעלה מ- 38 שנה
ב) למטה מ- 46 שנה
ג) נע בין 45 לבין 78 שנה
ד) נע בין 70 שנה לבין 86 שנה

3. הציון הממוצע השנתי של 800 תלמידי בית ספר תיכון מסוים במתמטיקה הוא 72. ציונם של 127 מהתלמידים הוא פחות מ- 50.

בהנחה שהתפלגות הציונים היא נורמלית, מִצאו את:

א) סטיית התקן

ב) מספר התלמידים שציונם במתמטיקה הוא גבוה מ- 83

ג) מספר התלמידים שציונם גבוה מ- 60 וקטן מ- 84.

4. הגובה הממוצע של 4000 המגויסים שהתייצבו בחודש מסוים הוא 175 ס"מ, וסטיית התקן היא 10 ס"מ. בהנחה שהתפלגות ערכי הגובה היא נורמלית, מִצאו:

א) את גובהו המְרַבִּי (מקסימלי) של מגויס מתוך 636 המגויסים הנמוכים

ב) את גובהו המזערי (מינימלי) של מגויס מתוך 1232 המגויסים הגבוהים

ג) כמה מהמגויסים קומתם קטנה מ- 170 ס"מ?

5. נערך ניסוי רפואי בקרב 500 גברים צעירים. במסגרת הניסוי נמנו מספר הכדוריות האדומות בממ"ק של כל נבדק.

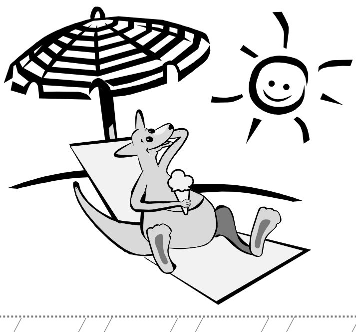
נמצא, כי הממוצע הוא 5.4 מיליוני כדוריות, וסטיית-תקן היא 0.8 מיליוני כדוריות בממ"ק.

מניחים, כי התפלגות המדידות היא נורמלית.

א) מצאו, בכמה מהגברים שבמדגם נע מספר הכדוריות האדומות בין 5 ל- 5.8 מיליונים בממ"ק;

ב) מתוך 500 הגברים הללו נבחרו 192, אשר מספר הכדוריות האדומות בממ"ק אחד בדמם קרוב יותר לממוצע.

מה היה מספר הכדוריות המקסימלי בממ"ק דם אצל הגבר שבקבוצה שנבחרה?



§37

1. א) 0.092 ב) 0.136 ג) 0.153 ד) 0.157 ה) 0 ו) 0.159
2. א) 50% ב) 15.9% ג) 6.7% ד) 0.6%
   ה) 69.2% ו) 93.3% ז) 99.8% ח) 100%
3. א) 50% ב) 84.1% ג) 93.3% ד) 99.4%
   ה) 30.8% ו) 6.7% ז) 0.2% ח) 0
4. א) 77.4% ב) 95.4% ג) 99.2% ד) 6.1%
   ה) 84.1% ו) 100%

§38

1. א) 1150 ב) 7950 ג) 26,650
2. א) 600 ב) 95 ג) 409 ד) 144
3. א) **הדרכה**: תחילה חַשבו, מה אחוז התלמידים שקיבלו ציון פחות מ- 50:

$$\frac{127}{800} = 0.159 = 15.9\%$$

לאחר מכן חַשבו את אחוז התלמידים שקיבלו ציון בין הממוצע לבין 50, ובעזרת הטבלה שבעמ' 309 מִצאו, בכמה סטיות תקן מהממוצע מרוחק הציון 50. **תשובה**: s = 22.

ב) 246 ג) 307

4. **הדרכה**: תחילה חַשבו, כמה אחוז מכלל המגויסים מהווים 636 המגויסים הנמוכים ו- 1232 הגבוהים; אחר-כך בעזרת הטבלה מִצאו את גבולות התחומים ביחידות של s.

א) **פתרון**: $\frac{x - \bar{x}}{s} = -1,\ x = \bar{x} - s = 175 - 10 = 165$

ב) **פתרון**: $\frac{x - \bar{x}}{s} = 0.5,\ x = \bar{x} + 0.5 \cdot s = 175 + 5 = 180$

ג) 1232

5. א) 192 ב) 5.8