

למי קראת נורמלי



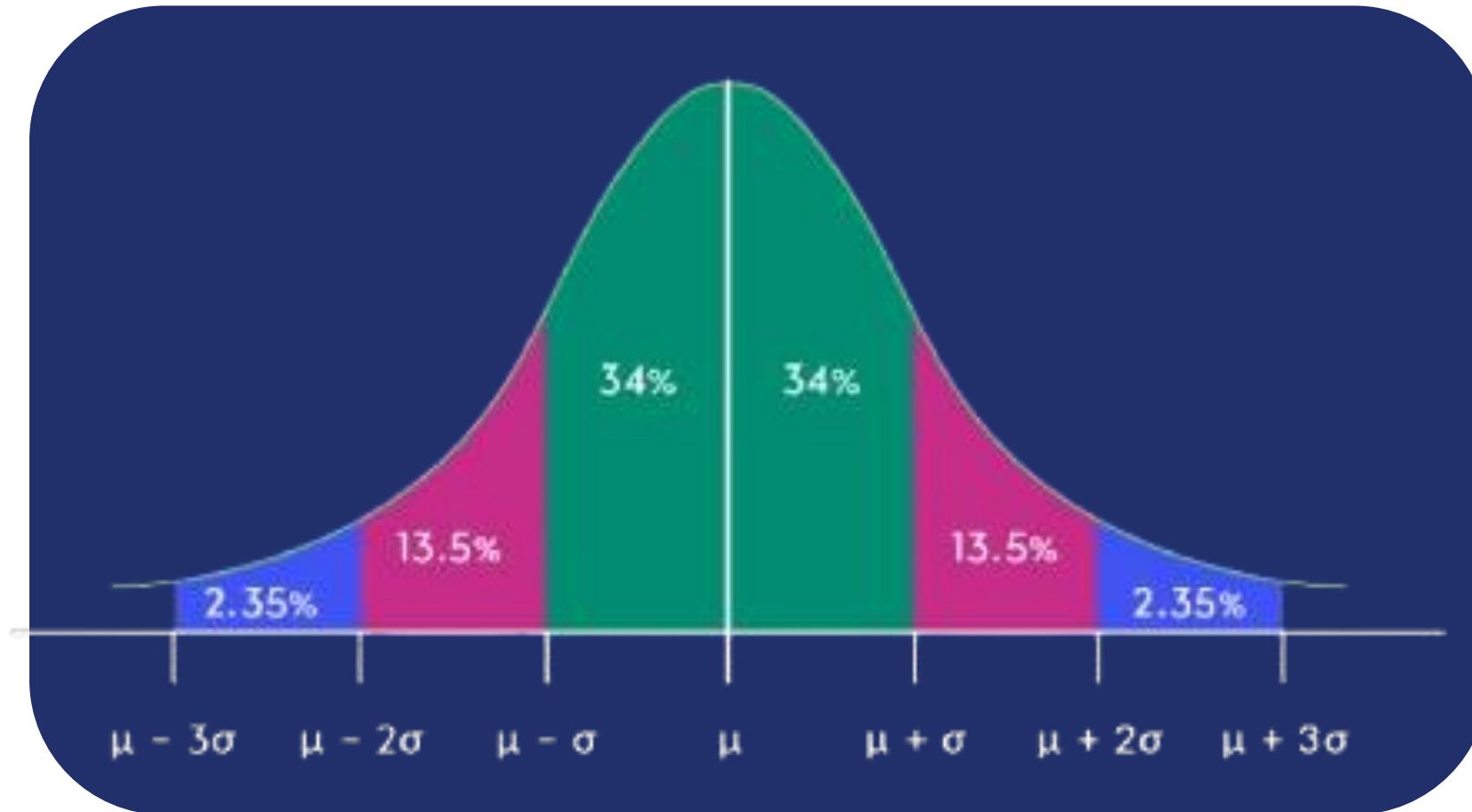
מה במצגת זו?

- התפלגות נורמלית
- ציון תקן \ Z-score
- סטנדרטיזציה
- נרמול

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

תזכורת: סטיית תקן

יחידה שמודדת מרחק מהממוצע

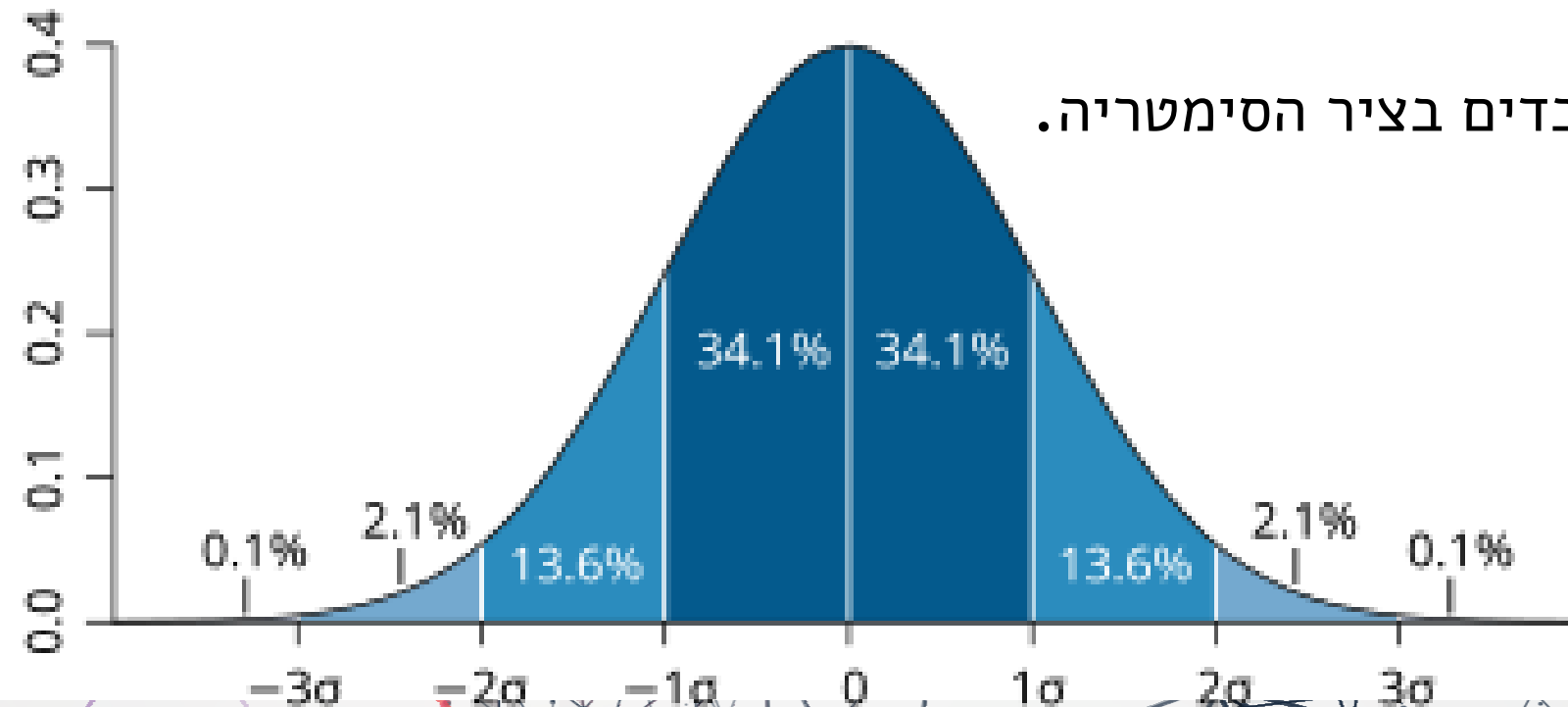


התפלגות נורמלית\גאוס

גרף זה מוכר גם בשם "גרף פעמון" או "פעמון גאוס", שכן כאשר הוא משורטט בתור גרף המציין את השכיחות של כל ערך, מקבלת העקומה צורה דמוית פעמון.

תכונות:

- ההתפלגות היא סימטרית
- חד-שיאית (unimodal).
- התוחלת, החציון והשכיח מתלכדים בציר הסימטריה.
- מתקיים חוק 68-95-99.7%



68-95-99.7% Rule

The Empirical Rule

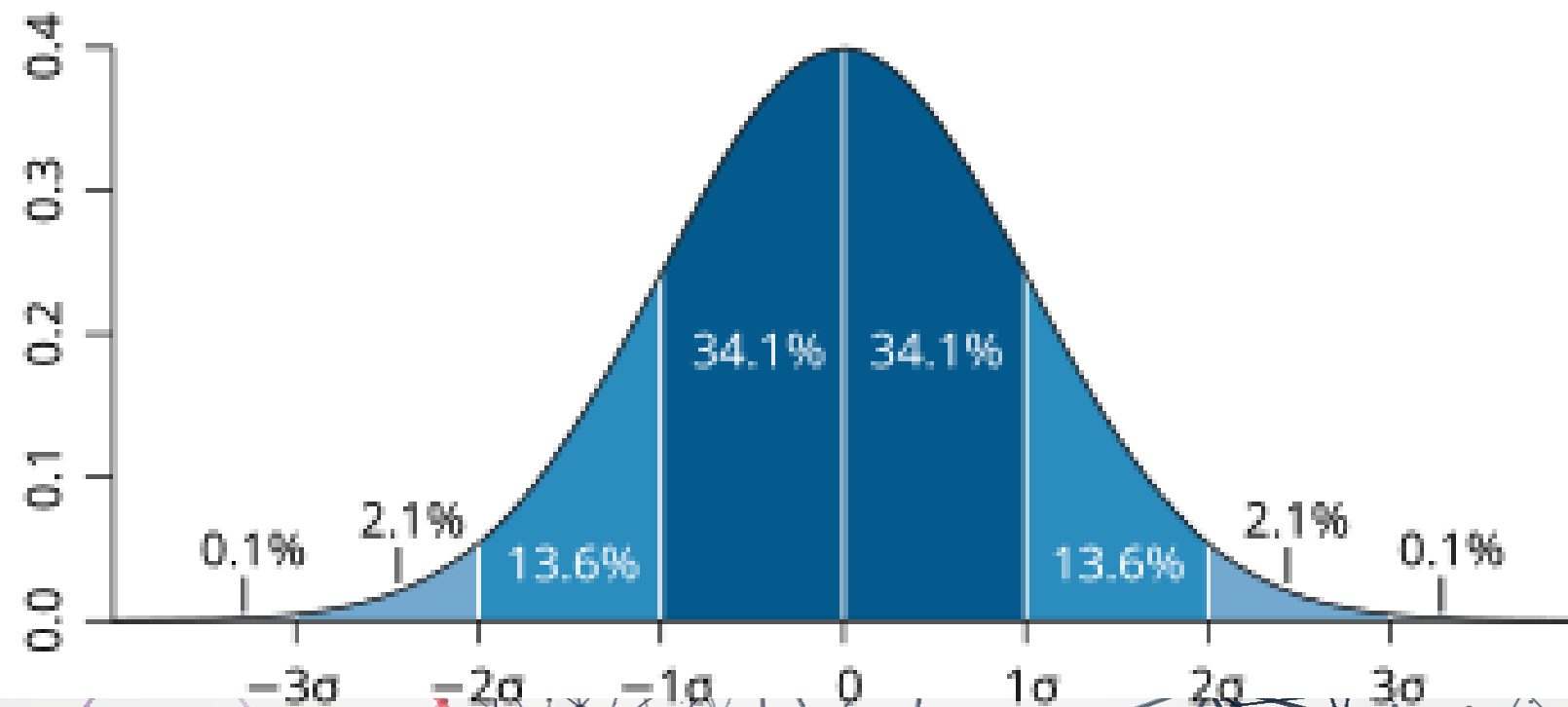
בהתפלגות נורמלית

68%* מהנתונים נמצאים בטווח 1std

95%* מהנתונים נמצאים בטווח 2std

99.7%* מהנתונים נמצאים בטווח 3std

*בקירוב



כמה סטיות תקן מהממוצע – נמצא ערך מסויים

ציון התקן מאפשר להשוות באופן יחסי ערכים מהתפלגויות שונות.

כאשר התצפית גדולה מהממוצע (תוחלת), ציון התקן יהיה חיובי.
כאשר התצפית קטנה מהממוצע (תוחלת), ציון התקן יהיה שלילי.
כאשר התצפית שווה לממוצע, ציון התקן יהיה אפס.

$$Z = \frac{x_i - \mu}{\sigma}$$

אוכלוסיה

$$Z = \frac{x - \bar{x}}{s}$$

מדגם

ציון תקן Z score - דוגמא

עובד מחברה א' מרוויח 9,700 ₪ בחודש. הממוצע בחברה א' הוא 9,300 ₪ עם סטיית תקן 1,500 ₪.

עובד מחברה ב' מרוויח 8,300 ₪ בחודש. הממוצע בחברה ב' הוא 7,100 ₪ לחודש עם סטיית תקן 900 ₪.

מי מהעובדים מרוויח טוב יותר באופן יחסי ?
איזה שכר בחברה א' שקול לשכר של 8,000 ₪ בחברה ב' ?

ציון תקן Z score - פתרון

עובד מחברה א' מרוויח 9,700 ₪ בחודש. הממוצע בחברה א' הוא 9,300 ₪ עם סטיית תקן 1,500 ₪.
עובד מחברה ב' מרוויח 8,300 ₪ בחודש. הממוצע בחברה ב' הוא 7,100 ₪ לחודש עם סטיית תקן 900 ₪.
מי מהעובדים מרוויח טוב יותר באופן יחסי ?
איזה שכר בחברה א' שקול לשכר של 8,000 ₪ בחברה ב' ?

נחשב את ציוני התקן:

$$Z_A = \frac{9700 - 9300}{1500} = 0.266 \quad ; \quad Z_B = \frac{8300 - 7100}{900} = 1.333$$

העובד מחברה ב' מרוויח טוב יותר באופן יחסי למקום העבודה.

$$Z_{8000} = \frac{8000 - 7100}{900} = 1 \rightarrow 1 = \frac{X - 9300}{1500} \rightarrow X = 10,800$$

השכר בחברה א' ששקול ל – 8,000 ₪ בחברה ב' הוא 10,800 ₪.

תרגול

תרגול

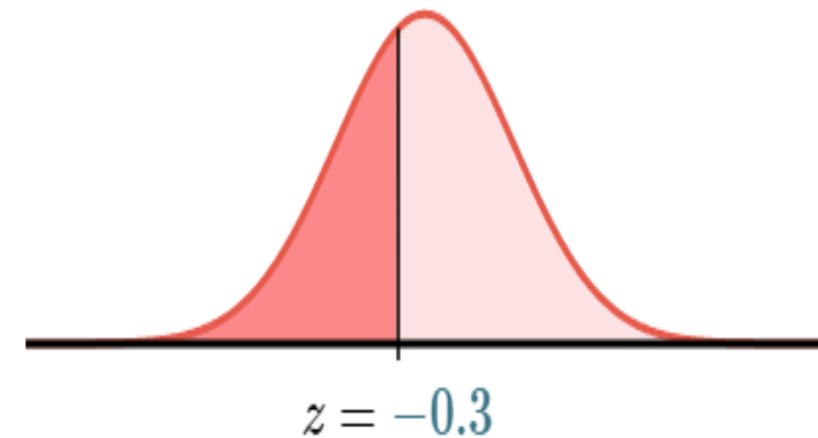


שימוש ב Z-table \טבלת ההתפלגות הנורמלית הסטנדרטית

כמות מים בליטרים שצרכו חקלאים במושב מתפלגת נורמלית עם ממוצע של 81 ליטר וסטיית תקן של 10 ליטרים. חנן הגנן צרך 78 ליטרים בחודש. איזה אחוז מהחקלאים צרכו **פחות** מים מחנן הגנן?

z	0	0.01	0.02	0.03	0.04	0.05
-0	.50000	.49601	.49202	.48803	.48405	.48006
-0.1	.46017	.45620	.45224	.44828	.44433	.44034
-0.2	.42074	.41683	.41294	.40905	.40517	.40129
-0.3	.38209	.37828	.37448	.37070	.36693	.36317
-0.4	.34458	.34090	.33724	.33360	.32997	.32636
-0.5	.30854	.30503	.30153	.29806	.29460	.29116
-0.6	.27425	.27093	.26763	.26435	.26109	.25785
-0.7	.24196	.23885	.23576	.23270	.22965	.22663
-0.8	.21186	.20897	.20611	.20327	.20045	.19766
-0.9	.18406	.18141	.17879	.17619	.17361	.17106
-1	.15866	.15625	.15386	.15151	.14917	.14686

$$Z = \frac{x - \mu}{\sigma} = \frac{78 - 81}{10} = -0.3$$



<https://www.ztable.net/>

תשובה: 0.382 או 38.2%

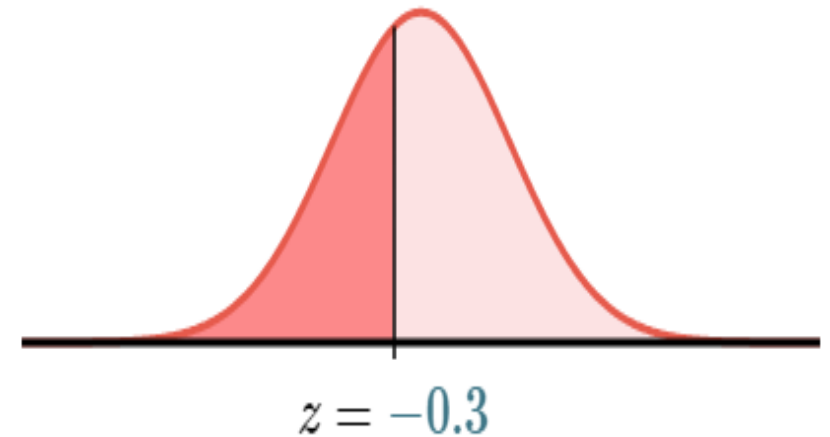


שימוש ב Z-table \טבלת ההתפלגות הנורמלית הסטנדרטית

כמות מים בליטרים שצרכו חקלאים במושב מתפלגת נורמלית עם ממוצע של 81 ליטר וסטיית תקן של 10 ליטרים. חנן הגן צרך 78 ליטרים בחודש. איזה אחוז מהחקלאים צרכו יותר מים מחנן הגן?

z	0	0.01	0.02	0.03	0.04	0.05
-0	.50000	.49601	.49202	.48803	.48405	.48006
-0.1	.46017	.45620	.45224	.44828	.44433	.44034
-0.2	.42074	.41683	.41294	.40905	.40517	.40129
-0.3	.38209	.37828	.37448	.37070	.36693	.36317
-0.4	.34458	.34090	.33724	.33360	.32997	.32636
-0.5	.30854	.30503	.30153	.29806	.29460	.29116
-0.6	.27425	.27093	.26763	.26435	.26109	.25785
-0.7	.24196	.23885	.23576	.23270	.22965	.22663
-0.8	.21186	.20897	.20611	.20327	.20045	.19766
-0.9	.18406	.18141	.17879	.17619	.17361	.17106
-1	.15866	.15625	.15386	.15151	.14917	.14686

$$Z = \frac{x - \mu}{\sigma} = \frac{78 - 81}{10} = -0.3$$



<https://www.ztable.net/>

תשובה: $100\% - 38.2 = 61.8\%$

תרגול

תרגול

standardization סטנדרטיזציה

בדרך כלל בסט נתונים יש **יחידות מידה שונות** לנתונים שונים (ק"ג למשקל, ש" למחיר, סמ' לאורך וכו') – מה שמייצר קושי בהשוואה ביניהם.
תהליך הסטנדרטיזציה הופך את כל הנתונים לבעלי **יחידות מידה זהות**

Subject	Age (years)	Occupation	Height (cm)	Weight (kg)	Length of car ride (hours)	Date of car ride	Sex
P	21	Student	175	62.5	5	21. 3. 2009	Male
L	22	Student	185	85	5	21. 3. 2009	Male
Š1	23	Student	170	74	2	19. 12. 2009	Male
LI1	20	Student	185	74	2	19. 12. 2009	Male
Š2	23	Student	170	74	5.5	20. 12. 2009	Male
LI2	20	Student	185	74	5.5	20. 12. 2009	Male

אם נחשב לכל נתון את ציון התקן שלו, נקבל קבוצה חדשה של ציוני תקן.

**הממוצע של קבוצה זו הוא תמיד 0 ,
והשונות תמיד 1**

תהליך נקרא סטנדרטיזציה

standardization סטנדרטיזציה

הממוצע של קבוצה זו הוא תמיד 0, והשונות תמיד 1.

Index	x	z
1	160	-1.345
2	163	-0.905
3	167	-0.318
4	170	0.122
5	175	0.856
6	180	1.589
Mean	169.17	0
SD	6.82	1

לדוגמא בטבלה הבאה נתונים גבהים של 6 סטודנטים. לאחר שחושב הממוצע וחושבה סטיית התקן של גבהי הסטודנטים, חושב ציון התקן של כל אחד מגבהים. לבסוף, חושב ממוצע וחושבה סטיית תקן לציוני התקן

לפעמים יש צורך להכניס את כל
הנתונים לטווח בין 0 ל1.
פעולה זו נקראת נירמול.
למשל - בקלאסיפיקציה של תמונות
CNN.

יש שפונקציות מובנות שעושות את זה
למשל minmaxscaler של skitlearn



- ✓ התפלגות נורמלית
- ✓ ציון תקן Z-score
- ✓ סטנדרטיזציה
- ✓ נרמול