

# סטטיסטיקה והסתברות

מבוא

# מה נלמד בחלק הזה

- סוגי סטטיסטיקה
- דגימה
- סוגי נתונים
- סולמות מדידה
- שכיחות

## מדע איסוף, ארגון וניתוח נתונים



- בעשרות השנים האחרונות הסטטיסטיקה הייתה בשימוש נרחב בתחומים רבים בחיינו, אולם בשנים האחרונות, עם הפיכתו של העולם ל- **Data Driven World**. חשיבות הסטטיסטיקה רק גברה.
- בעקבות ריבוי ה**נתונים**, ידע בסטטיסטיקה הוא ידע נדרש, ולא רק סטטיסטיקאים צריכים לדעת סטטיסטיקה.
- יש המון **מידע**- האתגר הוא לעבוד איתו נכון כדי להגיע להחלטות טובות יותר

## עובדות או פיסות מידע שאפשר למדוד

למשל:

- קוטר צמיג הרכב
- גיל משתתפי המרוץ למליון
- IQ משתתפי האחוזון העליון
- גובה עצי אלון בחורשה
- מגדר זכרות בסוואנה



# סוגי סטטיסטיקה

```
graph TD; A[סוגי סטטיסטיקה] --> B[סטטיסטיקה היסקית]; A --> C[סטטיסטיקה תיאורית];
```

סטטיסטיקה  
היסקית

סטטיסטיקה  
תיאורית

# דגימה - שלבי המחקר המדעי:





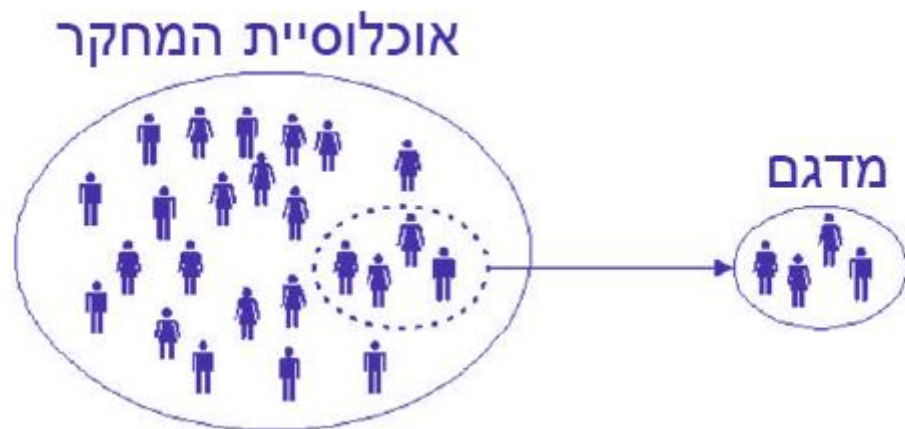
# דגימה - אוכלוסייה ומדגם

בכל מחקר מדעי שנבצע, נגדיר את אוכלוסיית המחקר. לעיתים, בבואנו לבצע את המחקר, נוכל להתבונן על הפריטים באוכלוסייה. לדוגמא, אם ברצוננו לדעת מהי התפלגות הציונים בקורס ציור בחוג לאומנות חזותית, יהיה זה קל מאוד להסתכל על הציונים של כלל הסטודנטים, ולהסיק את המסקנה הנכונה והמדויקת.

אולם לרוב, לא נוכל להתבונן בכל הפריטים באוכלוסייה. לדוגמא, אם ברצוננו לדעת האם תרופה חדשה מסייעת בהורדת לחץ דם, ברור כי לא נוכל לנסות אותה על כל האנשים באוכלוסייה הסובלים מלחץ דם גבוה (לא אפשרי לוגיסטית וכנראה מאוד יקר). במקרה זה, נבחר **מדגם** מתוך האוכלוסייה.

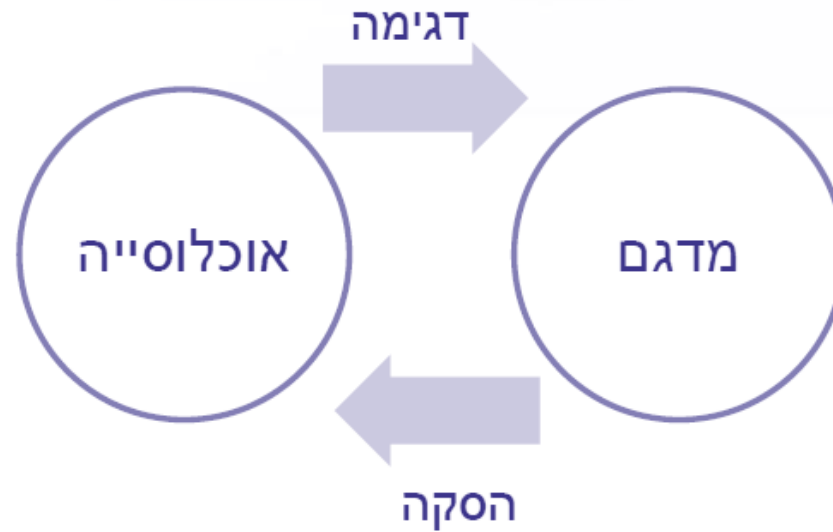
## מדגם הוא קבוצה חלקית של האוכלוסייה.

נהוג לסמן את האוכלוסייה ב  $N$ , ואת המדגם מתוכה ב  $n$ .

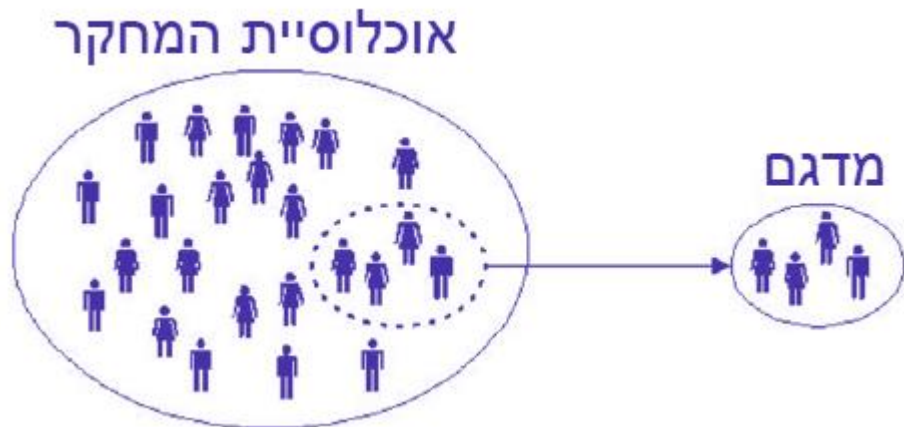


# דגימה - אוכלוסייה ומדגם

כאשר אנו לוקחים מדגם מהאוכלוסייה, אנו רוצים להסיק ממנו מסקנות על האוכלוסייה כולה. יחסי הגומלין בין האוכלוסייה למדגם הם כאלה:



מכיוון שהמדגם הוא קבוצה חלקית של האוכלוסייה, תמיד תהיה מידה של אי ודאות במסקנות על האוכלוסייה, ויש סיכוי לטעות.





Sunrise Extra

Nation's Verdict  
An Editorial, On Page 10.

# The Washington Post

The Weather  
Today: Rain; much colder in afternoon and night.  
Tomorrow: Fair.  
Yesterday: High 78 low 62.  
Details on Page 4.

NO. 22,056    Entered as Second-Class Matter, Postoffice, Washington, D. C.    WASHINGTON: WEDNESDAY, NOVEMBER 4, 1936 \*\*\*\* XX    Copyright, 1936, By The Washington Post    THREE CENTS

## President Roosevelt Is Reelected in Landslide; Victory May Assure Him All But Two States; Democrats Increase Big Majority in Congress

### Easy Control Of the House For Roosevelt Program Due

New Deal Will Hold Present Membership, May Increase It.

Captures 3 Seats In Pennsylvania

Other Gains Reported From N. Y., Maine, R. I., Connecticut.

By Felix Bruner.

A Democratic majority in the new House of Representatives even bigger than that in the present House was indicated this morning as returns poured in from Congressional districts in all parts of the country. Riding the swift flowing tide of Roosevelt landslide, Democratic members of the House almost without exception were re-elected and gains were indicated in several States. On the basis of incomplete returns, the Democrats apparently had gained at least seven seats and the number may grow.

The only Democratic member of Congress reported defeated for reelection was Fred J. Simon of the thirty-third, New York, District. Simon was a leader in the fight to repeal the "red rider" in the District of Columbia school appropriation bill which provided that teachers

### The Post's Institute Poll Forecast Of Election, Most Nearly Correct

YESTERDAY'S election decided not only a political contest . . . but, also, a poll contest.

AMONG ALL of the polls that have been conducted during the campaign, the Washington Post, publishing the results of the final AMERICA SPEAKS poll Sunday, three days ago, gave to its readers the most accurate forecast of the election.

THIS POLL, conducted by the American Institute of Public Opinion under the direction of Dr. George Gallup, showed only three States . . . Maine, New Hampshire and Vermont . . . as safely in the Landon column.

AS THIS edition of The Washington Post goes to press, latest election returns indicate this prediction as being borne out almost to the letter.

THIS COMING Sunday's AMERICA SPEAKS poll . . . on the basis of polls scientifically conducted throughout the span of the campaign . . . such questions are these: "When was the election won or lost?" "What issues changed the most votes?"

IF YOU are not already a regular Washington Post subscriber, please call NATIONAL 4200 today. AMERICA SPEAKS polls appear exclusively, in Washington, in The Washington Post.

### G. O. P. Faces Loss of Many Senate Seats

Hastings Runs Behind in Delaware; Barbour Trails in Jersey.

Green Leading Metcalf in R. I.; Norris and Borah Ahead.

By Robert C. Albright.

Certainly that the Democrats will maintain their top-heavy majority in the Senate, and a strong possibility that they may even increase their hold on that august body, developed early this morning on the crest of the Roosevelt sweep.

New Deal domination of the Senate, never at any time endangered, may take deeper root in the Seventy-fifth Congress if later returns bear out trends manifest in still scattered counting.

Republicans at best had hoped to pick up four or five new seats in the Senate, but as tallying progressed their optimism waned, and their present minority of 22 was endangered.

**Strongholds Weaken.**

In three normally Republican strongholds, Delaware, New Jersey and Rhode Island, incomplete returns showed Republican Senatorial incumbents trailing in their race for reelection. In Rhode Island, Theodore Francis Green, Democrat, moved ahead of Senator Josiah M. Metcalf, Republican. Senator Daniel O. Hastings, Republican, faltered, while Alvin M. Landon, Democrat, moved ahead of Senator Josiah M. Metcalf, Republican.

Not dark eyes that had sparkled brightly when this morning at the

Re-elected by Landslide



PRESIDENT ROOSEVELT.

### Mother of President Is Happy; Mrs. Landon Smiles Amid Tears

Crowds Cheer Family of the President at Hyde Park.

Hyde Park, N. Y., Nov. 3 (AP)—Mrs. Franklin D. Roosevelt, radiant in a white chiffon evening gown, welcomed jubilant friends tonight at the Roosevelt family home, and said of the election return:

"I am very happy to feel that a great number of the people seem to wish to have the President carry on the work which he has begun in the last five years."

The President's 42-year-old mother, Mrs. James Roosevelt, sat, with almost happy tears, on the terrace in a drizzling rain while he greeted his cheering Hyde Park neighbors, confident "I wasn't worried a bit. I just felt

### New Hampshire Conceded to F.R. By Knox Paper

Landon's Defeat Is Worst Suffered by Candidate Since Taft's in 1912; Roosevelt's Popular Ballot Rises.

Electoral and Popular Vote Computations—Page 2.

**Bulletin**

Manchester, N. H., Nov. 4 (A.P.)—The Manchester Union, of which Col. Frank Knox, Republican Vice Presidential candidate, is president, conceded New Hampshire's four electoral votes to Roosevelt early today.

By Franklyn Waltman.

Franklin Delano Roosevelt and John Nance Garner yesterday were re-elected President and Vice President, respectively, of the United States by a landslide of unprecedented proportions in both the electoral college and in popular votes.

They will commence their new terms of office on January 20, the first time a President and Vice President have been inaugurated on that date instead of March 4.

Gov. Landon conceded his defeat at 1 a. m. today, asserting in a telegram to the President that "the Nation has spoken" and adding that "every American will accept the verdict and work for the common cause of our country."

"That is the spirit of democracy," he continued. "You have my sincere congratulations."

President Roosevelt left his Republican opponent, Gov. Alf M. Landon, and his running mate, Col. Frank Knox, buried in an avalanche of the largest vote cast in the history

### Reprisals 'Out,' Landon Admits Farley Asserts, Defeat, Sends

Harley Voted, Defeat, Sends

בשנת 1936 נערך סקר בחירות בארה"ב. הסקר ניבא שאלפרד לאנדון ינצח את פרנקלין רוזוולט.

הסקר התבסס על מדגם של כשני מיליון בוחרים, שנדגמו מספר הטלפונים בארה"ב.

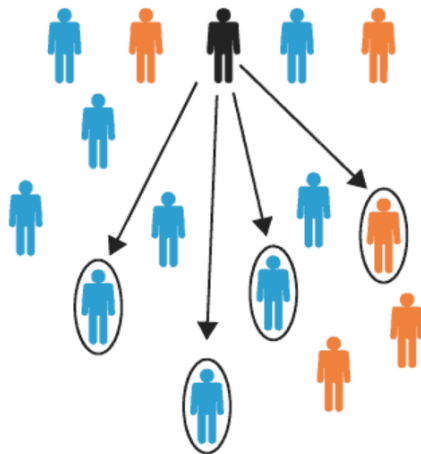
בתום הבחירות -רוזוולט ניצח.

מה ההסבר?

## שיטות דגימה

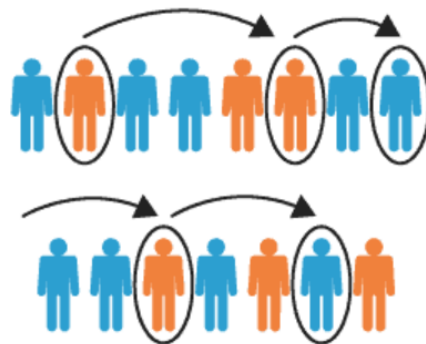
### דגימה ממוקדת מטרה

Purposive sample



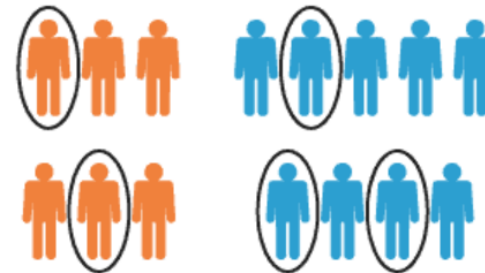
### דגימה שיטתית

Systematic sample



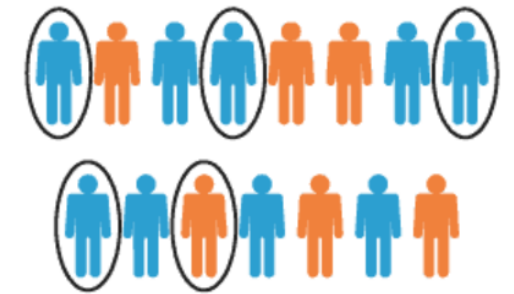
### דגימת שכבות

Stratified sample



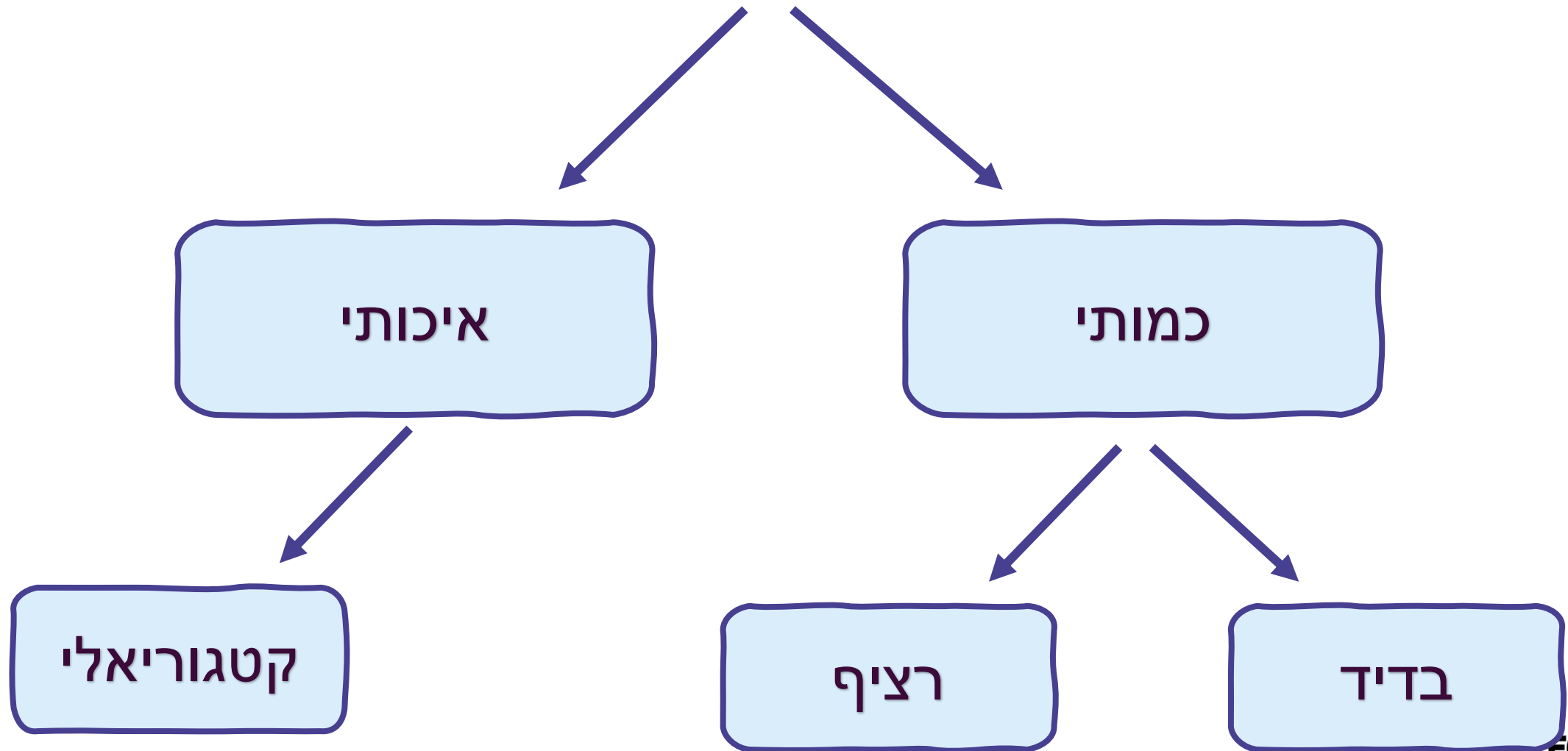
### דגימה מקרית פשוטה

Simple random sample



# משתנים

משתנה הוא תכונה שיכולה לקבל כל ערך



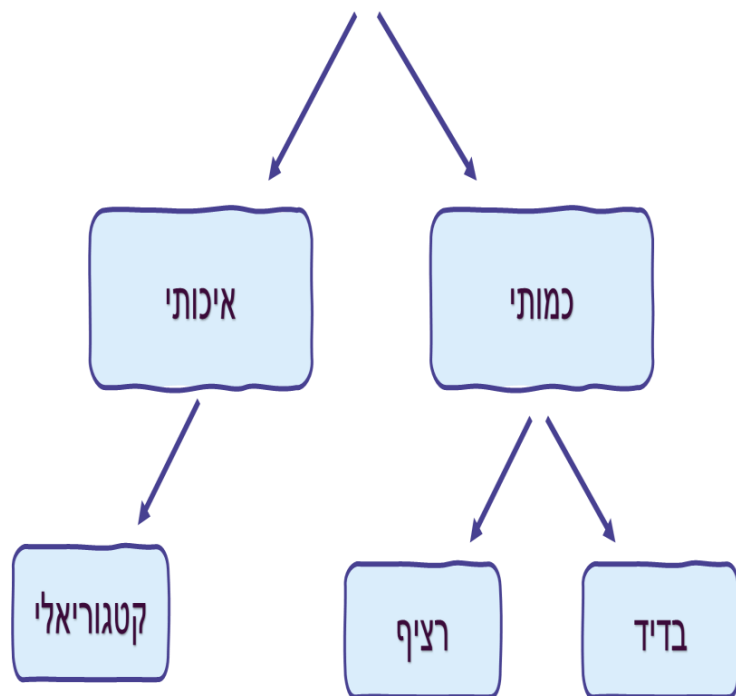
# משתנה

משתנה שערכיו אינם ערכים מספריים, אלא ערכים שניתן לסווגם לקבוצות, נקרא **משתנה איכותי**, או **משתנה קטגוריאלי**. ערכי המשתנה הם הקבוצות, או הקטגוריות.

משתנה שערכיו נמדדים ומוצגים באופן מספרי, נקרא **משתנה כמותי**.

משתנה כמותי שקבוצת הערכים האפשריים שלו היא קבוצה סופית או קבוצה אינסופית בת מנייה, נקרא **משתנה כמותי בדיד**.

משתנה כמותי שקבוצת הערכים האפשריים שלו היא קבוצה אינסופית שאיננה בת מנייה, נקרא **משתנה כמותי רציף**.







## בכל אחד מהסעיפים הבאים, קבעו האם המשתנה הוא איכותי או כמותי:

- המצב המשפחתי של הרופאים בבית החולים איכילוב
- הזמן הדרוש לביצוע ריצת מרתון.
- הצבע של רכבים בחניון של מרכז קניות.
- נפח הגלידה בכוס מילקשייק.
- מספר מקומות הישיבה באצטדיון כדורגל.

# סולמות מדידה:

בסטטיסטיקה, מדידה היא התאמת ערך מספרי לתכונה מסוימת הנמדדת במחקר.  
בכל פעם שאנו מתאימים ערך מספרי לתכונה מסוימת, אנו מבצעים **מדידה**  
**סטטיסטית**.

**דוגמא 1** : אנו רוצים לאגור נתונים על מינם של המשתתפים במחקר. אנו יכולים  
להצמיד לגברים את הערך המספרי " 0 " ולנשים את הערך המספרי " 1 ". כאשר אנו  
עושים זאת, אנו "מודדים" את מינם של המשתתפים במחקר, על ידי התאמת ערך  
מספרי לתכונה.

**דוגמא 2** : בסקר שביעות רצון, נדרשים סטודנטים לדרג את המרצה / מתרגל לפי  
המדרג הבא:

1 – כלל לא מרוצה 2 – לא מרוצה 3 – בסדר 4 – מרוצה 5 – מרוצה מאוד  
התאמת הערך המספרי לקטגוריה -זוהי מדידה



מנה

רווח

סודר

שמי

# סולמות מדידה- שמי

סולם המדידה השמי הוא סולם מדידה **איכותי**, בו הערך המספרי משמש לזיהוי בלבד של התכונה נמדדת. אין חשיבות לערכם של המספרים או לסדר שלהם.

דוגמאות למשתנים איכותיים בעלי סולם מדידה שמי:  
מגדר " 0 - גבר, " 1 - אישה, או " 1 - גבר, " 2 - אישה

צבע עיניים " 1 - חום, " 2 - ירוק, " 3 - כחול

מצב משפחתי " 1 - רווק, " 2 - נשוי, " 3 - גרוש, " 4 - אלמן

המספרים משמשים לזיהוי בלבד, ועל כן לא נוכל לבצע עליהם פעולות חשבון, למעט פעולת ספירה.  
לא נוכל

לבצע חיבור כגון  $1+0$ , כי אין משמעות לביצוע הפעולה גבר+אישה. עם זאת, נוכל לספור כמה גברים וכמה נשים יש לנו בנתונים.

# סולמות מדידה - סודר

סולם מדידה סודר הוא סולם מדידה **איכותי** בו הערכים המספריים מבטאים סדר טבעי של התכונה הנמדדת. דוגמאות למשתנים איכותיים בעלי סולם מדידה סודר:

סקר שביעות רצון " 1 – מאוד לא מרוצה, " 2 – לא מרוצה, " 3 – קצת מרוצה, " 4 – מרוצה, " 5 – מאוד מרוצה.

דרגת קצונה בצבא " 1 – סג"מ, " 2 – סגן, " 3 – סרן, " 4 – רס"ן, " 5 – סא"ל, " 6 – אל"מ, " 7 – תא"ל, " 8 – אלוף, " 9 – רא"ל

גודל כוס השתייה במקדונלדס " 1 – קטן, " 2 – בינוני, " 3 – גדול, " 4 – ענק

למספרים בסולם סודר אין משמעות מלבד הסדר אותו הם מגדירים. את דירוג המרצים של 1 עד 5 נוכל להחליף ל – 0 עד 4, וזה יהיה בסדר. מכאן שאין משמעות למספרים עצמם.

מבחינת פעולות, נוכל לספור כמה מופעים יש מכל קטגוריה, אך נוכל גם לענות על שאלות כמו "כמה אנשים לפחות קצת מרוצים? נוכל לחשב אחוזונים (נלמד בהמשך..).

# סולמות מדידה - רווח

סולם מדידה רווח הוא סולם מדידה **כמותי**, כלומר יש משמעות כמותית למספרים שאנו מודדים. בסולם מדידה רווח, לראשונה המושג מדידה סטטיסטית ומדידה פיסיקלית מתייחסים לאותה הפעולה.

סולם מדידה רווח מתאים מספרים לתכונות שעבורן ה-0 נקבע שרירותית, ואין לו את המשמעות המתמטית של המספר 0.

דוגמאות למשתנים כמותיים מסולם רווח:

טמפרטורה -טמפרטורה של 0 מעלות צלזיוס אין משמעותה שאין טמפרטורה בחוץ.

גובה פני הים - גובה 0 מסמל את גובה פני הים. אין זה אומר שאין בנקודה זו גובה

מבחינת פעולות חשבון, המגבלה שמוטלת על סולם הרווח היא שיש משמעות להפרשים (רווחים) שבין הערכים הנמדדים, אך לא ליחסים (מנות) שבין הערכים.

רוב המדדים הסטטיסטיים הנפוצים ניתנים ליישום עם נתונים שנמדדו בסולם רווח (לממוצע טמפרטורת יש משמעות. עם זאת, יש להיזהר בפירוש התוצאות).

# סולמות מדידה - מנה

סולם מדידה מנה הוא סולם מדידה **כמותי**, כלומר יש משמעות כמותית למספרים שאנו מודדים.

סולם מדידה מנה מתאים מספרים לתכונות שעבורן לערך  $0$  יש משמעות זהה למשמעות המתמטית של המספר  $0$ .

דוגמאות למשתנים כמותיים מסולם מנה:

משכורת - משכורת  $0$  משמעה שאין משכורת

מספר ילדים -  $0$  ילדים אומר שאין ילדים

מבחינת פעולות חשבון, אין מגבלות, גם הפרשים וגם יחסים מותרים על נתונים שנמדדים בסולם מדידה מנה.

כל המדדים הסטטיסטיים ניתנים לחישוב עבור נתונים שנמדדים בסולם מנה.

שם הסולם	משמעות לסדר	משמעות לערך	משמעות לאפס
שמי			
סודר			
רווח			
מנה			



שם הסולם	משמעות לסדר	משמעות לערך	משמעות לאפס
שמי	-	-	-
סודר	+	-	-
רווח	+	+	-
מנה	+	+	+

בכל אחד מהסעיפים הבאים, קבעו את סולם המדידה של המשתנה. נמקו את הבחירה:

- מספר העמודים בספרי לימוד של בני גורן.
- הדירוג העולמי של שחקני טניס.
- הטמפרטורה בתוך תנורים לאפיית פיצה.
- משקלם של מחשבים ניידים.
- משוב לקוחות על איכות הקפה (5-מעולה 4-סבבה 3- בסדר 2- לא עפתי 1- קטסטרופה)
- נושא המגזינים בחדר המתנה של רופא שיניים (ספורט, אופנה, חדשות, טבע וכו')

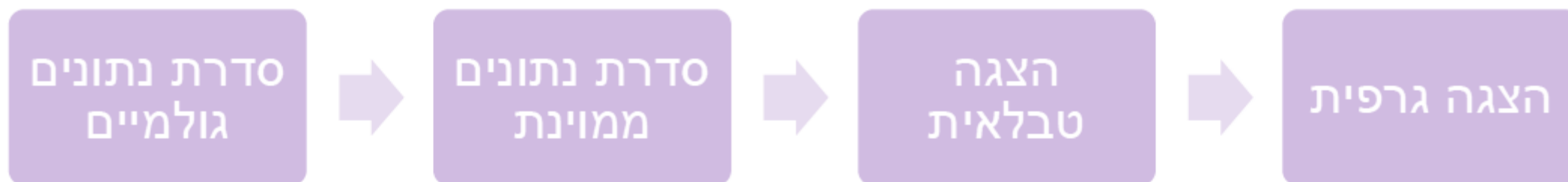
# הצגת נתונים-

השלב הראשון ואולי החשוב ביותר במחקר סטטיסטי, הוא להציג את הנתונים שנאספו במחקר באופן נכון וברור לקורא.

את הנתונים שנאספו במחקר ניתן להציג באופן גולמי כסדרה, אך הצגה זו תהיה ברורה אך ורק כאשר מספר הנתונים מועט מאוד. כאשר מספר הנתונים גדול יותר, נציג את הנתונים בשתי חלופות: **הצגה טבלאית**, ו**הצגה גרפית**.

הבחירה בסוג הטבלה או סוג הגרף, תלויה באופי הנתונים: בסוג המשתנה ו/או בסולם המדידה.

כאשר נקבל את סדרת הנתונים הגולמיים, נמיין אותה, ואז נבנה טבלאות. מהטבלאות נוכל לבנות גרפים מתאימים.



# הצגת נתונים - שכיחות $f(x)$ ושכיחות יחסית

להלן נתונים שנאספו במחקר על מצבם המשפחתי של 22 הנחקרים:  
רווק, נשוי, רווק, נשוי, גרוש, נשוי, רווק, נשוי, גרוש, נשוי, רווק, נשוי, גרוש, נשוי, אלמן,  
גרוש, נשוי, נשוי,  
רווק, רווק.

ברור כי הצגה זו אינה נוחה לקריאה. הדרך הנכונה להציג את הנתונים היא באמצעות **טבלת שכיחויות**.

שכיחות יחסית $f\%(X)$	שכיחות $f(X)$	מצב משפחתי (X)
40.9%	9	רווק
36.4%	8	נשוי
18.2%	4	גרוש
4.5%	1	אלמן
100%	22	סה"כ

# הצגת נתונים - שכיחות מצטברת

כאשר לפנינו משתנה איכותי עם סולם מדידה סודר, נציג את הנתונים באמצעות טבלת שכיחויות. הפעם, מכיוון שישחשיבות לסדר, עלינו לדאוג ולסדר את הערכים של המשתנה לפי הסדר הנכון בטבלת השכיחויות.

דוגמא: מחקר בא לבדוק מהו גודל כוס השתייה המועדף על ילדים האוכלים ברשת מזון מהיר. להלן התוצאות:

שכיחות מצטברת $F(X)$	שכיחות יחסית $f\%(X)$	שכיחות $f(X)$	גודל כוס השתייה $(X)$
5	20%	5	קטן
15	40%	10	בינוני
22	28%	7	גדול
25	12%	3	ענק
25	100%	25	סה"כ

כאשר סולם המדידה הוא סודר, יש היגיון בחישוב שכיחות מצטברת. אנו יכולים לומר כי 15 ילדים הזמינו כוס שתייה בינונית ומטה.

# הצגת נתונים - חלוקה למחלקות

X- Blood Cholesterol	f(X)
170-180	3
180-190	7
190-200	13
200-210	8
210-220	5
220-230	4
Total	40

כאשר לפנינו משתנה כמותי בדיד עם מספר רב של ערכים, או משתנה כמותי רציף, לא נוכל להשתמש בטבלת השכיחויות שזה עתה הכרנו על מנת להציג את הנתונים.

הפתרון לבעיה הוא **טבלת שכיחויות עם מחלקות**.

את הנתונים שנרצה להציג נחלק לקבוצות, מחלקות עם **גבול תחתון וגבול עליון**, ונמנה את השכיחות של כל מחלקה, ולא של כל ערך בודד.

יש יותר מדרך אחת לחלק את הנתונים למחלקות. בבואנו לחלק את הנתונים, עלינו להקפיד על שני כללי אצבע חשובים:

**לא יהיה נתון שלא תהיה לו מחלקה**

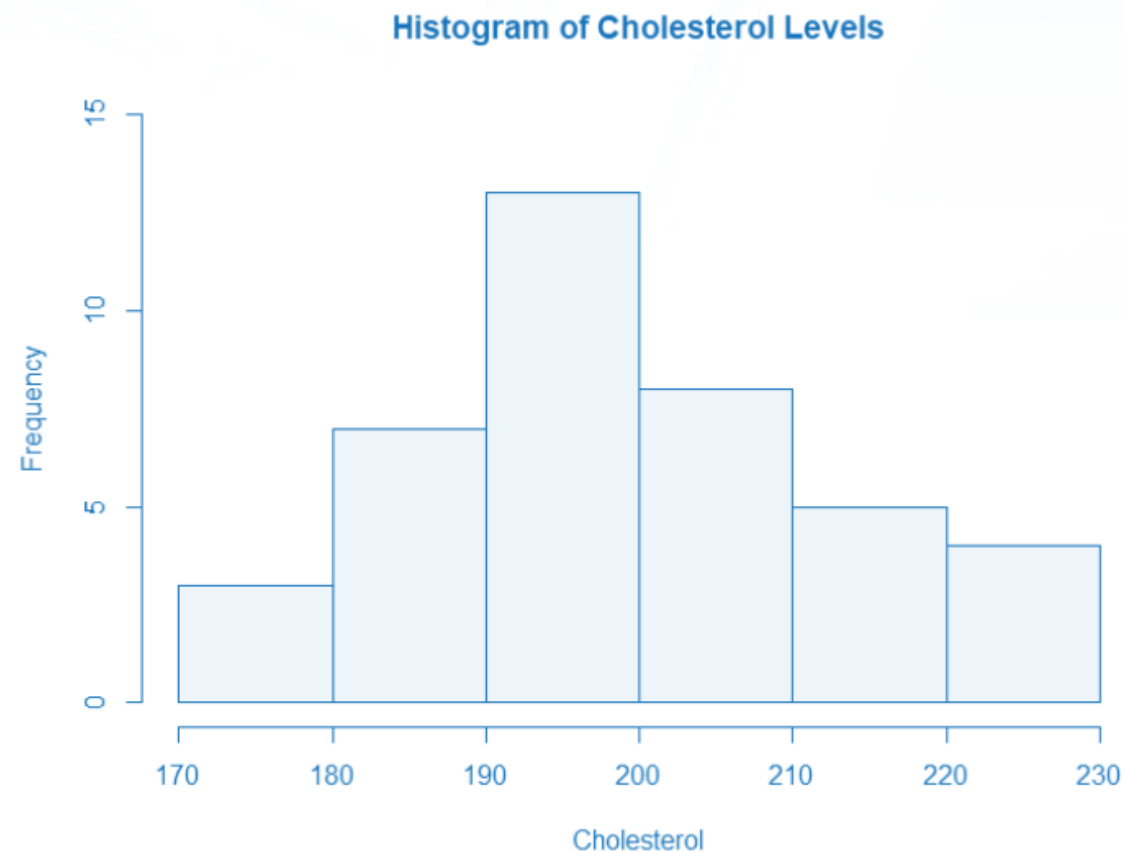
**לא יהיה נתון שיהיו לו שתי מחלקות.**



# הצגת נתונים - חלוקה למחלקות. היסטוגרמה

171, 174, 178, 181, 183, 183, 183, 184, 187, 188, 191, 191, 191, 192, 193, 193, 193, 194, 194, 195, 196, 196, 199, 200, 200, 200, 202, 204, 204, 205, 206, 211, 212, 213, 213, 216, 220, 221, 221, 227

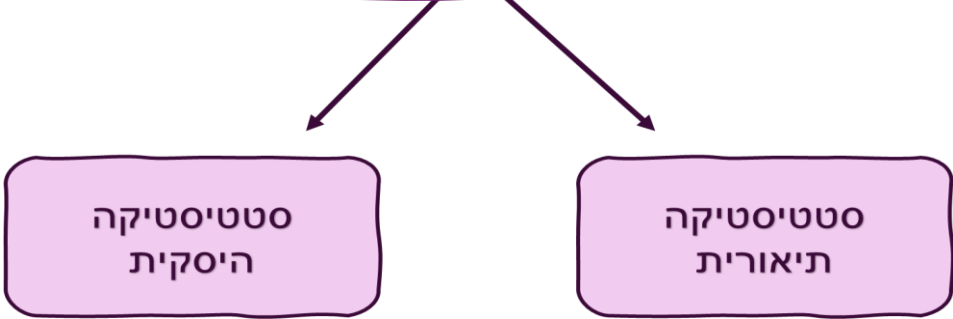
X- Blood Cholesterol	f(X)
170-180	3
180-190	7
190-200	13
200-210	8
210-220	5
220-230	4
Total	40



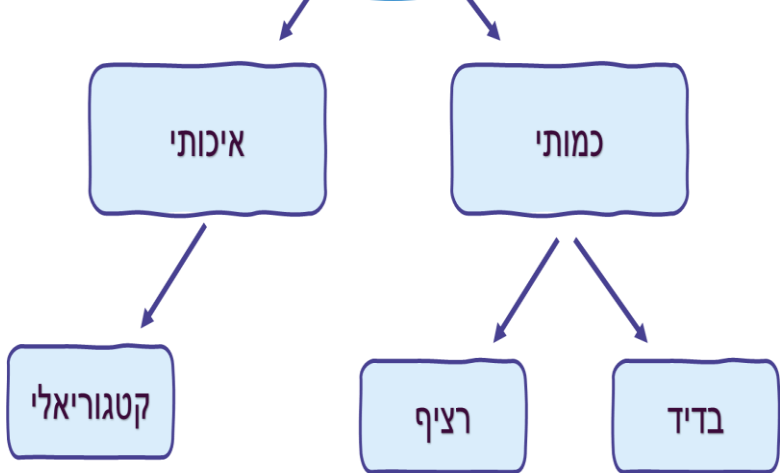
## שיטות דגימה



## סטטיסטיקה



## משתנה



שם הסולם	משמעות לסדר	משמעות לערך	משמעות לאפס
שמי	-	-	-
סודר	+	-	-
רווח	+	+	-
מנה	+	+	+