



סטטיסטיקה – מבוא וסקירה כללית

1

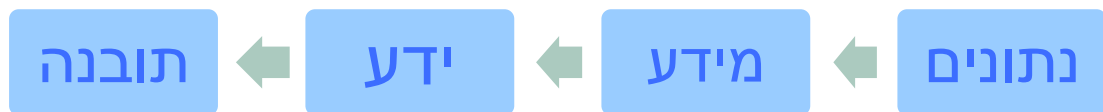


מהי סטטיסטיקה

- **סטטיסטיקה היא מדע העוסק בהסקת מסקנות וקבלת החלטות בתנאי אי וודאות**
 - תנאי אי וודאות = לא כל המידע הנחוץ לנו זמין, ולכן יש צורך להסתמך רק על מידע חלקי
- **סטטיסטיקה איננה מתמטיקה או חלק מהמתמטיקה**
 - עקרונות היסוד של הסטטיסטיקה אינם מתמטיים
 - הסטטיסטיקה משתמשת בכלים מתמטיים, כמו הרבה מדעים אחרים

2

- איסוף נתונים
- עיבוד הנתונים שנאספו
- ניתוח הנתונים שנאספו
- הסקת מסקנות על סמך הניתוח הסטטיסטי
- הצגת התובנות המתקבלות



- הצטברות: נתונים מלאים – אין צורך בסטטיסטיקה
- הרגלי צפיה בנטפליקס, תחלואה בקורונה, חישוב השכר הממוצע במשק
- תצפיות: אין לנו שליטה בפרמטרים המעניינים אותנו, אין אפשרות לאסוף נתונים מלאים
- מחקר פראמינגהם, הרגלי צריכה, השוואת מתחסנים נגד קורונה ללא מתחסנים
- ניסויים: שליטה מלאה בפרמטרים המעניינים אותנו על ידי הקצאה מקרית
- ניסויים קליניים, מחקרים פסיכולוגיים, AB testing
- מדגמים: שליטה מלאה בתנאים מסויימים
- מדגמי בחירות, סקר הכנסות, קבוצות מיקוד

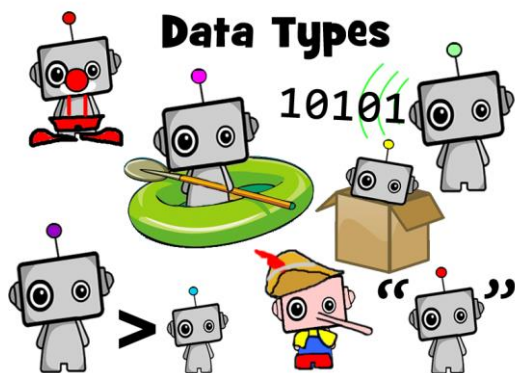
בקורס זה נתמקד בנתונים הנאספים מניסויים ומדגמים, אך לא בנתונים תצפיתיים

- דגימה הסתברותית: לכל הפרטים באוכלוסייה יש אותו סיכוי להיכלל במדגם
- מדגם נוחות: דוגמים את מי שנוח לדגום
- פניה לכל מי שאפשר להגיע אליו
- מדגם מתנדבים
- מדגם "כדור שלג"

רק דגימה הסתברותית מבטיחה מדגם מייצג!

- להרחבה בנושא חפשו
- בגוגל: נסיכת המדעים מדגם לא מייצג
 - ביוטיוב: נסיכת המדעים סקירה על סקרים

מכיוון ששני כלים מרכזיים בהם הסטטיסטיקה משתמשת הם המתמטיקה והמחשב, הנתונים בהם אנו עוסקים הם למעשה נתונים מספריים



- יש שלוש דרכים עיקריות לסווג נתונים
- על פי האופי בו הנתונים מוצגים
 - טקסט/מספר
 - על פי התכונה המתמטית של הנתונים
 - רציפים/בדידים
 - על פי האופן בו הנתונים נמדדים
 - מייד נרחיב בנושא – כי זה מה שבאמת חשוב

סיווג נתונים על פי אופן המדידה

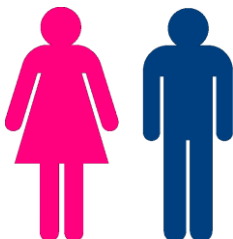
- משתנים איכותיים
 - משתנה שמי (nominal scale)
 - משתנה סודר (ordinal scale)
- משתנים כמותיים
 - משתנה רווח (interval scale)
 - משתנה מנה (ratio scale)

להרחבה בנושא חפשו בגוגל: נסיכת המדעים סקירה על סולמות מדידה

7

משתנה שמי/נומינלי

- לערך המספרי של המשתנה אין שום משמעות כמותית
- הערך משמש לזיהוי בלבד



אשה=2

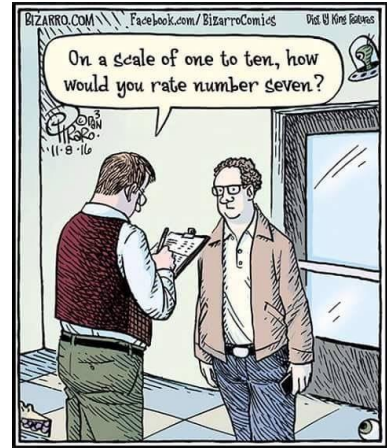
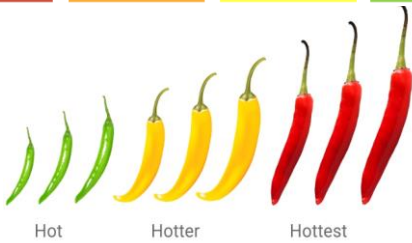
גבר=1



8

יש משמעות לסדר בין הערכים אך לא להפרש ביניהם

פעילות מינימלית	פעילות ממוצמת	פעילות מוגבלת	פעילות מורחבת
מחוצ למבנה לא יותר מ- 20 איש	מחוצ למבנה תפוסה ולא יותר מ- 40% 50	מחוצ למבנה תפוסה ולא יותר מ- 60% 100	מחוצ למבנה תפוסה ולא יותר מ- 80% 250
בתוך מבנה לא יותר מ- 10 איש	בתוך מבנה תפוסה ולא יותר מ- 20% 25	בתוך מבנה תפוסה ולא יותר מ- 40% 50	בתוך מבנה תפוסה ולא יותר מ- 60% 100



יש משמעות להפרשים בין הערכים אך לא ליחס ביניהם

משקעים: 2%
 לחות: 61%
 רוח: 18 קמ"ש

18 °F | °C



שבת	יום ו'	יום ה'	יום ד'	יום ג'	יום ב'	יום א'	שבת
8° 16°	8° 14°	8° 13°	8° 12°	8° 20°	9° 19°	10° 21°	9° 21°
14:00	11:00	8:00	5:00	2:00	23:00	20:00	17:00
שבת	יום ו'	יום ה'	יום ד'	יום ג'	יום ב'	יום א'	שבת
46° 61°	46° 57°	47° 56°	47° 53°	46° 68°	49° 67°	50° 69°	48° 69°



"Wait, it's September already?! Wow, time flies, huh?
OK, then you only have two months to live."



11

- אופן הניתוח הסטטיסטי של נתונים תלוי בסולם המדידה שבהם הם נמדדים
- שיטות הניתוח הסטטיסטיות של משתנה בסולם מדידה נתון תקפות גם למשתנים גבוהים יותר בסולם, אך לא להיפך
- דוגמאות
 - ניתן לחשב את הממוצע עבור משתנה בסולם רווח ובסולם מנה, אך לא עבור משתנה שמי ומשתנה סדר
 - במשתנה שמי, ניתן לחשב את השכיחות של כל ערך, וזה אפשרי גם עבור משתני סדר, רווח ומנה
 - ניתן לחשב את החציון עבור משתנה סדר, ולכן גם עבור משתנה רווח ומשתנה מנה, אך לא עבור משתנה שמי.

12