<http://u.math.biu.ac.il/~vishne/courses/88165/88165.html>

הקורס הוא איחוד של הקורסים 89-262,88-161,88-162, שגם מופיעים באתר.

כדי לעזור לנו להעריך סיכויים וסיכונים, במקום להסתכל על הסיכוי עצמו p עדיף להסתכל על .

פרק 2: סטטיסטיקה תיאורית

תאור מתומצת של נתונים רבים..

## משתנה X

יכול להיות:

* איכותי: לא ניתן לכימות או לסידור – פשוט חלוקה לקבוצות. למשל מין, צבע עיניים וכו'.
* אורדינלי: מהמילה Order. ניתן לסידור אבל לא לכימות. למשל השכלה: יסודית<תיכונית<על תיכונית.
* אינטרוולי: יש מרווחים קבועים, אבל אין נקודת אפס – למשל שנת לידה, או טמפרטורה . אפשר למדוד אפשרים ביניהם, אבל אין משמעות לכפל או לחילוק.
* מנתי: למשל משכורת, משקל. אפשר לחבר, לחסר, לחלק, להכפיל.

חלוקה נוספת למשתנים:

* משתנה בדיד: סדרה סופית או אינסופית בת מניה של אפשרויות.
* משתנה רציף: למשל

## אוכלוסיה

מכלול הישויות הכפופות לניתוח הסטטיסטי שבו אנחנו מעוניינים.

## מדגם

קבוצת הנתונים שבידינו.

## תאור גרפי

לפעמים חותכים בנקודה מסויימת, כלומר לא מתחילים מ0.

עושים את זה כשיש לנו משתנה אחד איכותי – למשל צבע במקרה שלנו.

אם יש זוג משתנים אפשר להשתמש בתרשים נקודות:

מדדי מרכז

* ממוצע:
* חציון: מספר שמחצית הנתונים גדולה ממנו ומחצית קטנים ממנו. אם מספר הנתונים אי זוגי החציון יכול להיות כל מספר בין שני המספרים באמצע: למשל עבור החציון יכול להיות כל מספר ב
* שכיח: נתון שמופיע מספר פעמים מקסימלי.
* אמצע הטווח: אם מסדרים כך ש אז אמצע הטווח הוא

מדד מרכז = פונקציה של הנתונים שצריכה לקיים:

1. *– סימטריות, אם מחליפים את הסדר הפונקציה לא משתנה.*

## תכונות של ממדי המרכז האלה

* הממוצע הוא נקודת המינימום של הפונקציה :  
  הממוצע הוא נקודת האפס של הנגזרת -> נקודת האפס של הפונקציה..
* אם היינו לוקחים את סכום הערכים המוחלטים של השגיאות במקום את סכום הריבועים שלהם:  
  רוצים להקטין את t כל עוד . ולהגדיל אחרת ⇦ המינימום הוא החציון.
* . השכיח הוא המינימום של הפונקציה  
  רוצים למזער את מספר השגיאות – כלומר המקומות בהם המספר הוא לא השכיח.
* אמצע הטווח ממזער את הפונקציה – כלומר הנקודה שבה השגיאה המקסימלית היא מינימלית.

|  |  |  |
| --- | --- | --- |
|  |  | ממוצע |
|  |  | חציון |
|  |  | שכיח |
|  |  | אמצע הטווח |

# הכללה

אפשר להגדיר מדד מרכז : לכל , בתור המינימום של

מדדי פיזור

מודד עד כמה הנתונים מפוזרים. צריך להיות:

* סימטרי
* הומוגני: הכפלת הנתונים ב מכפילה את המדד ב
* אדיש להזזות: תוספת קבוע אינה משנה את המדד

# דוגמאות

* שונות המדגם:   
   תרגיל:   
  המדד הזה לא מדד פיזור שכן אינו הומוגני – אם מכפילים ב זה יוכפל ב. לכן מגדירים את ***סטיית התקן*** – זה כן מדד פיזור.  
   ⬄ כל הים שווים.
* רוחב הטווח:
* אפשר להגדיר "פרופורציון": ,
* . . = מרחק בין רבעונים.  
  אפשר גם לחלק לעשירונים: