

ד"ר עינת אביאלי  
הוקלד ע"י דביר חדד

# סיכומי הרצאות

מבוא לסטטיסטיקה  
והסתברות

הוקלד ע"י דביר חדד

מרצה: ד"ר עינת אביאל  
סמסטר קיץ תשע"ב

## 26.7.12-הוצאה 1

### סטטיסטיקה:

מדע העוסק בנתונים כמותיים. את הנתונים אנו אוספים, מעבדים, מציגים ומסיקים מהם מסקנות.

### הגדרות:

משתנה: תכונה המקבלת ערכים שונים. כגון טמפר', גובה, משקל, וכו'..  
קבוע: גודל שערכו יחיד.



### מיון משתנים:

1. מיון לפי מהות משתנה: כמותי ואיכותי.  
כמותי: ערכי המשתנה מציינים כמות- גובה, משקל, וכו'.  
איכותי: ערכי המשתנים נבחנים לפי סוג איכות ולא ע"י ביטוי מספרי- מין, מצב משפחתי, וכו'.  
המשתנה הכמותי מתחלק לרציף ובדיד. רציף הכוונה שבין כל שני ערכים יש  $\infty$  ערכים. (גיל, גובה). בדיד הכוונה שבין כל שני ערכים מספר סופי של ערכים. (מס' ילדים, מס' חדרים בבית)
2. מיון לפי רמת המדידה:
  - א. משתנה נורמלי (שמי) – ערכי המשתנה נבחנים לפי שמות, באשר אין משמעות לסדר בין הערכים. (מצב משפחתי, מוצא, דת...)
  - ב. משתנה אורדינאלי (סדורי) – ניתן לדרג את הערכים מהגבוה לנמוך. (דרגות בצבא, שביעות רצון..)
  - ג. משתנה אינטרוואלי (מרווחי) – ערכי המשתנים הם מספרים ויש משמעות למרווחים ביניהם. (IQ, טמפרטורה..)
  - ד. משתנה יחס- ערכי המשתנה הם במספרים ויש משמעות בחלוקת ערך אחד בשני וציון היחס ביניהם (גיל, גובה, משקל)..
3. מיון לפי כיוון התיאור:
  - משתנה תלוי: (מוסבר) משתנה המתואר על ידי המשנה הבלתי תלוי
  - משתנה בלתי תלוי: (מסביר) נשנה המסביר את המשנה התלוי.

### סטטיסטיקה תיאורטית:

שכיחות: מספר המקרים במדגם עבור קבוצה מסוימת. התחלקות השכיחויות לפי הערכים של המשתנה.  
תסומן ב  $f/F$ .

### דוגמה:

משתנה X	שכיחות $f(x)$	שכיחות יחסית $\frac{f(x)}{n}$	$F(x)$ שכיחות מצטברת	שכיחות מצטברת יחסית $\frac{F(x)}{n}$
5	3	$\frac{3}{22}$	3	$\frac{3}{22}$
$\mu_0=6$	8	$\frac{8}{22}$	11	$\frac{11}{22}$
7	4	$\frac{4}{22}$	15	$\frac{15}{22}$
8	3	$\frac{3}{22}$	18	$\frac{18}{22}$
9	2	$\frac{2}{22}$	20	$\frac{20}{22}$
10	2	$\frac{2}{22}$	22	$\frac{22}{22}$
	$n=22$ (גודל המדגם)	1		

**ד"ר עינת אביאלי**  
**הוקלד ע"י דביר חדד**

**מדדים:**

1. שכיח -  $M_0$ . הערך / נקרה ששכיחותו היא הגבוהה ביותר. דיאגרמת מקלות- הצגה גרפית למשתנה בדיד.

**מדדי מרכז:**

1. חציון -  $Md$ . חציון הוא הערך המשתנה הנחקר שמחצית הנחקרים גדולים ממנו וחצי גדולים ממנו. החציון הוא ערך אמצעי בהתפלגות.  
 כלל: עבור קבוצת ערכים  $\{X_1, \dots, X_n\}$  שמסודרים בסדר עולה. קיימות שתי אופציות : n זוגי או אי זוגי. אם n אי זוגי, אז החציון נמצא במקום ה- $\frac{n+1}{2}$  ז"א :  $Md = X_{\frac{n+1}{2}}$   
 אם n זוגי, נקבל ש :  $Md = \frac{X_{\frac{n}{2}} + X_{\frac{n+1}{2}}}{2}$
2. ממוצע: הממוצע הוא סכום כל ערכי המשתנה לכל הנחקרים חלקי מס' הנחקרים.

ובטבלת שכיחויות :

3. אמצע טווח :  $MR = \frac{X_{min} + X_{max}}{2}$

X משתנה	f(x)	I רוחב קבוצה	אמצע קטע- x חדש	d צפיפות = $\frac{\text{שכיחות}}{\text{קבוצה רוחב}}$
40-50	5	10	45	5/10
50-60	3	10	55	3/10
60-70	8	10	65	8/10
70-80	2	10	75	2/10
80-90	1	10	85	1/10
	n=19			

**דוגמה:**

בהתחלה הנתונים היו רציפים ויצרנו x חדש כך שהמקרה בדיד ואנו יודעים לטפל בו. נמצא את השכיח, הממוצע, אמצע הטווח והחציון.

שכיח =  $M_0 = 65$

טווח אמצע =  $MR = \frac{40+90}{2} = 65$

ממוצע =  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 60.25$

כדי לחשב את החציון ברציף, נשתמש בנוסחה הבאה  $Md = L_0 + \frac{\frac{n}{2} - F(x)}{f(x)} (L_1 - L_0)$

נבהיר מה כל אחד מהסימונים :

$L_0$  - גבול תחתון של קב' החציון

$L_1$  - גבול עליון של קב' החציון

$F(x)$  - שכיחות מצטברת עד קב' החציון (לא כולל)

$f(x)$  שכיחות של קבוצת החציון

n - גודל המדגם.

**ד"ר עינת אביאלי**  
**הוקלד ע"י דביר חדד**

**מדדי פיזור:**

מהו פיזור? מידת השוני בקבוצה או גודל ההבדלים בין הנתונים לבין ערך מרכזי מסויים.

1. טווח:  $R = X_{max} - X_{min}$

2. טווח ברבעוני:  $IQR = Q_3 - Q_1$

3. שונות- ממוצע קיבועי הסטיות בין ערכי הסדרה לבין הממוצע.  $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$ .  
תמיד חיובית.

4. סטית התקן,  $s$ , שורש ריבועי של השונות. משמעות: רוב התצפיות נמצאות  $\pm$  סטיית התקן מהממוצע.

**בטבלת שכיחות:**

בבדיד:  $\bar{x} = \frac{\sum_{i=1}^n x_i f(x_i)}{n}$ ,  $s^2 = \frac{\sum_{i=1}^n f(x_i)(x_i - \bar{x})^2}{n}$  ברציף אותו דבר בדיוק פשוט עם האינסוף החדש (אמצע קטע)

**שאלה:** נתונה קבוצה של 20 תלמידי. הממוצע שלהם הים 90, וסטיית תקן 5. מאוחר יותר נבחנו עוד 2 תלמידים וקיבלו את הציונים הבאים:

ה'	ד'	ג'	ב'	א'	ציונים נוספים
85,95	81,82	97,98	83,97	87,93	
לא השתנה	ירד	עלה	לא השתנה	לא השתנה	ממוצע
לא השתנה	דרוש חישוב נוסף	עלה	עלה	ירד	סטטית תקן

**ציון תקן-** מתאר מיקום יחסי של תצפית ביחס לסדרה הסטטי אליה משתייך.

**דוגמה:** סטודנט נבחן בסטטיסטיקה ובביולוגיה. בסטטי קיבל 85 ובביו 80. בסטטי  $\bar{x} = 80$  וגם  $S_x = 5$   
82

ובביולוגיה  $\bar{x} = 70$  וגם  $S_x = 3$ .

מתקיים שציון התקן של התלמיד בסטטי קטן מציון התקן בביולוגיה. לכן מצטיין יותר בביולוגיה.

**מקדם השתנות:** מסמנים ב.c.v.  $c.v. = \frac{S_x}{\bar{x}}$ . מדד פיזור יחסי שנועד להשוות בין קבוצות מבחינת הומוגניות. ככל ש.c.v. קטן יותר, הקבוצה יותר הומוגנית.

**דוגמה:**

באוכלוסיה מסוימת הגובה הממוצע הוא 174 ס"מ עם סטטית תקן 15 ס"מ. המשקל הממוצע 60 קילו עם סטיית תקן של 10 קילו. היכן הקבוצה יותר הומוגנית?  
ברור שהגובה יותר הומוגני מחישוב פשוט.

**טרנספורמציה ליניארית:**

טרנספורמציה ליניארית זה שינוי קוויץ אם יש לנו משתנה x ומבצעים  $y = ax + b$  אזי:

$Mdy = aMdx + b$  (החציון)

$\bar{y} = a\bar{x} + b$  (הממוצע)

$M_{oy} = aM_{ox} + b$  (השכיח)

$R_y = |a|R_x$  (הטווח)

$MR_y = aMR_x + b$  (אמצע טווח)

$S^2_y = a^2 S^2_x$  (שונות)

$S_y = |a|S_x$  (סטית תקן)

$IQR_y = |a|IQR_x$  (טווח בין רבעוני)