דף נוסחאות - סולמות מדידה

סולמות מדידה יעזרו לנו להציג תכונות - בעזרת מספרים.

שם הסולם	מהות הסולם	דוגמאות
סולם שמי	**<u>זהות בלבד</u>** של פרט או איבר – ללא מידע נוסף	מספר ת.ז , מספר חולצה בכדורסל , ארץ מוצא , מספר אישי בצבא
סולם סדר	**<u>זהות + סדר</u>** - מי יותר ומי פחות, אך לא בכמה יותר או פחות.	סקר דירוג מ1-7, דרגות בצה״ל, שביעות רצון מקורס באוניברסיטה.
סולם רווחים (לא נשתמש)	**<u>זהות + סדר + הבדל</u>** - ההפרש בין המספרים מלמד על ההבדל ביניהם אבל אין מידע על היחס שבין הערכים. כלומר, ערכים גבוהים מציינים מיקום גבוה יותר מערכים נמוכים . כמו כן ניתן לדעת עבור ערך מסוים בכמה הוא גדול/קטן מערך אחר. **נדגיש, כאן למספר 0 יש משמעות.**	טמפרטורה
סולם מנה	**<u>זהות + סדר + הבדל + יחס</u>** - כלומר, התוספת כאן היא שאפשר לקבוע את גודל הפער בין המשתתפים ומהו היחס ביניהם. **נדגיש, המספר 0 מציין היעדר מוחלט של התכונה.**	משקל, גובה, משכורת, שטח ועוד. ״מספר ה...״ , נסיעה במהירות 120 קמ״ש. נסיעה במהירות 0 לא קיימת.

<u>לנושא זה אין שאלה במבחן בפני עצמה, נשתמש בו בכדי לפתור שאלות שנושא מדדי קשר (בהמשך הקורס).</u>

דף נוסחאות שיעור מבוא והצגת נתונים

נושא עיקרי בקורס.

1. **שאלה פתוחה במבחן – 25 נקודות – (99%)**
2. **סעיף או שניים בשאלות "נכון/לא נכון" – 5-10 נקודות**

הסבר על הפרק:

סטטיסטיקה תיאורית = ענף העוסק ביצירה והשוואה של מדדים **לתיאור תמציתי וקל לתפישה של נתונים**.

מטרת הסטטיסטיקה התיאורית היא לסייע בארגון וסיכום הנתונים שנאספו מכלי מדידה או הערכה שונים.

בפועל - נקבל נתונים של מדגם שנבדק, מסודרים בטבלה, ונתבקש לחשב על הנתונים האלה מדדים שונים.

טבלת שכיחויות

טבלה שבה עבור כל ערך במדגם מצוין מספר הפעמים שהוא מופיע בו.

מושגים לטבלת מחלקות

- **גבולות מחלקה** – עליון ותחתון
- **גבולות אמיתיים** – הגבול העליון של מחלקה שווה לגבול התחתון של המחלקה מתחתיו.
- **רוחב מחלקה** – גבול עליון פחות גבול תחתון

מושגים נוספים

- **שכיחות יחסית -** מספר הפעמים שערך כלשהוא חוזר על עצמו במדגם מסוים מתוך סך כל המדגם.
- **שכיחות יחסית באחוזים -** נכפיל את השכיחות היחסים ב 100.
- **שכיחות יחסית מצטברת -** לכל ערך במדגם השכיחות שלו + השכיחות של כל הערכים שקטנים ממנו.

הצגה גרפית של נתונים

- **דיאגרמת מקלות להצגת משתנה בדיד** - על ציר ה- X ערכי המשתנים, ציר ה- Y הוא תדירויות הופעת משתנה בערך מסוים.
על כל ערך אפשרי של המשתנה ניצב "מקלי" בגובה פרופורציוני למספר הפעמים שערך זה מופיע במדגם.
- **היסטוגרמה להצגת משתנה רציף** - על ציר ה- X קטגוריות של ערכי משתנים (למשל טווח ערכים).
מעל כל קטגוריה ניצב מלבן בעל שטח פרופורציוני למספר הפעמים שערכי הקטגוריה מופיעים במדגם.

---

* **כאשר רוחב המחלקות לא זהה:** נוסיף עמודת $\boxed{\textbf{צפיפות} = \frac{f(X)}{\text{רוחב}}}$

דף נוסחאות שיעור מדדי מרכז

דף נוסחאות של הקורס – עמוד ראשון למעלה :

מדדי מרכז:	$MR = \frac{x_{\max} + x_{\min}}{2}$; $\bar{x} = \frac{\sum_x xf(x)}{n}$; $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

רדו למטה לעמוד הבא :

שכיח = הערך שמופיע במדגם יותר פעמים מכל ערך אחר.

הסימון הוא = Mo. יכולים להיות שני שכיחים.

נשים לב: כאשר ישנה שכיחות שווה לכל הערכים, נאמר שאין שכיח או שכל ערך הוא שכיח.

אמצע טווח = הממוצע בין שני הערכים הקיצוניים ביותר במדגם/בהתפלגות.

הסימון הוא: MR.

$$\boxed{MR = \frac{X_{max} + X_{min}}{2}}$$

אמצע טווח מושפע מערכים קיצוניים.

חציון = הערך ש50% מהנתונים נמצאים מעליו ו50% מהנתונים נמצאים מתחתיו.

הסימון הוא = MD. נוסחא לחישוב **מיקום** החציון:

גודל המדגם

$$\boxed{MD = \frac{n+1}{2}}$$

נשים לב:
רשימת התצפיות חייבת להיות בסדר עולה כאשר מחשבים חציון.

אין צורך לחשב בקורס חציון בטבלת שכיחויות מקובצת.

עבור מספר שלם, החציון הוא הערך במיקום זה.
עבור מספר לא שלם, החציון הוא הממוצע בין שני האיברים שנמצאים סביב מיקום זה.

מציאת חציון כאשר נתונה לנו טבלת שכיחויות:

1. נוסיף עמודה לטבלה של שכיחות מצטברת ונמלא אותה.
2. כעת, נמצא את מיקום החציון ע"י הנוסחא: $MD = \frac{n+1}{2}$
3. נסתכל בעמודה של שכיחות מצטברת ונבדוק מתי אנו "נתקלים" במיקום שמצאנו קודם או במספר גבוה ממנו ונעצור.
4. נענה תשובה לפי הערך שבשורה שבה עצרנו.

הערה: אם קיבלנו מספר לא שלם, נעשה אותו דבר עבור האיבר הנוסף סביב המיקום שיצא לנו ונחשב ממוצע בין שני הערכים.

- החציון מושפע מסדר הנתונים ולא מערך הנתונים.
- אם נוסיף כמות זהה של ערכים שגדולים מהחציון ושל ערכים שקטנים ממנו, החציון לא ישתנה.

ממוצע

מקרה ראשון: כאשר נתונה לנו רשימת תצפיות (סדרת נתונים), הנוסחא היא:

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

מקרה שני: כאשר נתונה לנו טבלת שכיחויות, הנוסחא היא:

$$\boxed{\bar{x} = \frac{\sum x \cdot f(x)}{n}}$$

- הממוצע מושפע מכל הנתונים. כולל ערכים קיצוניים.
- סכום הסטיות מהממוצע הוא 0.

דף נוסחאות שיעור מדדי פיזור

דף נוסחאות של הקורס – עמוד ראשון למעלה:

מדדי פיזור:

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

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

$$S_x=\sqrt{S_x^2}$$

רדו למטה לעמוד הבא:

מבוא לנושא:

מדדי פיזור מספקים לנו מידע על **גודל השוני בין הנתונים**.

תכונות של מדדי הפיזור

1. הערכים של מדדי פיזור יכולים להיות רק חיוביים או 0.
2. כאשר כל הנתונים בהתפלגות שווים, כל מדדי הפיזור הם 0.
3. ככל שיש יותר שוני בין הנתונים, כך מדדי הפיזור יהיו גדולים יותר.
4. מדדי הפיזור אינם מושפעים מהוספת קבוע לכל הנתונים.

אחוז השגיאות

אחוז השגיאות = **אחוז התצפיות שבהן מתקבל ערך שונה מהערך השכיח.**

טווח = המרחק בין הערך הגדול ביותר לערך הקטן ביותר.

מסומן באות R – Range

$$\boxed{R = X_{max} - X_{min}}$$

שימו לב: נתונים חריגים יכולים להשפיע מאוד על הטווח. כלומר - הטווח מושפע מאוד (!!) מערכים קיצוניים.

טווח בינרבעוני = טב"ר = המרחק בין הרבעון הראשון לבין הרבעון השלישי.

$$\boxed{IQR = Q_3 - Q_1}$$

Q_1 – רבעון תחתון: הערך שעד אליו נמצאים 25% מהתצפיות.
Q_3 – רבעון עליון: הערך שעד אליו נמצאים 75% מהתצפיות.

הנוסחא לחישוב **מיקום** של Q_1 היא : $\frac{n+1}{4}$

הנוסחא לחישוב **מיקום** של Q_3 היא : $\frac{3(n+1)}{4}$

תכונות הטווח בינרבעוני:

1. אינו מושפע כלל מערכים קיצוניים.
2. בהתפלגות סימטרית הממוצע של הרבעון העליון והרבעון התחתון שווה לחציון.

<u>שונות = מדד לפיזור ערכים באוכלוסייה נתונה ביחס לממוצע שלה.</u>

<u>השונות היא מדד הפיזור הנפוץ והשימושי ביותר</u> והיא מסומנת כך : s_x^2.

1. כאשר נתונה רשימת תצפיות/סדרת ערכים נשתמש בנוסחא הבאה :

$$\boxed{s_x^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}=\frac{\sum_{i=1}^{n}{x_i}^2}{n}-\overline{x}^2}$$

2. כאשר נתונה טבלת שכיחויות של משתנה בדיד : (לא נלמד בקורס זה לחשב שונות של משתנה רציף) :

$$\boxed{s_x^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\cdot f(x)}{n}=\frac{\sum_{i=1}^{n}{x_i}^2 f(x)}{n}-\overline{x}^2}$$

- -

<u>סטיית תקן = השורש הריבועי החיובי של השונות</u> ולכן חישוב סטיית התקן הוא : $s_x=\sqrt{s_x^2}$

<u>תכונות השונות וסטיית תקן:</u>

1. ככל הסטיות גדולות יותר, כך השונות/סטיית התקן גדולה יותר.
2. הוספה של קבוע לרשימת התצפיות (כל התצפיות), לא משפיעה על השונות/סטיית התקן.
3. הכפלה של כל ערך במספר קבוע, תכפיל את השונות במספר הקבוע בריבוע וסטיית התקן תוכפל במספר הקבוע.

דף נוסחאות שיעור ממוצע משוקלל ושונות מצורפת

דף נוסחאות של הקורס – עמוד ראשון למעלה:

ממוצע משוקלל ושונות מצורפת:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j n_j}{N} \quad ; \quad N = \sum_{j=1}^{k} n_j \quad ; \quad s_c^2 = \frac{\sum_{j=1}^{k} n_j s_j^2}{N} + \frac{\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2}{N}$$

רדו למטה לעמוד הבא:

מבוא לנושא:

כשרוצים לבצע חישובים סטטיסטיים על כמה מקורות **בנפרד** ואז לדעת מה הממוצע והשונות של המקורות **ביחד**.

נכיר נוסחאות לחישוב הממוצע / השונות הכלליים בעזרת הממוצעים והשונויות של הקבוצות השונות" שנתונים לנו.

נוסחת ממוצע משוקלל:

מספר הקבוצות

ממוצע הקבוצה ה-j

גודל הקבוצה ה-j

ממוצע משוקלל

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j n_j}{N}$$

גודל כל הקבוצות יחד

נוסחת שונות מצורפת:

גודל הקבוצה ה-j

שונות הערכים בקבוצה ה-j

מספר הקבוצות

ממוצע הקבוצה ה-j

$$s_2^2 = \frac{\sum_{j=1}^{k} n_j s_j^2}{N} + \frac{\sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2}{N}$$

ממוצע משוקלל
שמצאנו קודם

גודל כל הקבוצות יחד

דף נוסחאות צורות התפלגות

מבוא לנושא:

הצגה גרפית של נתונים סטטיסטיים שעוזרת לנו להבין את פיזור הנתונים ולקבל "מושג" לגביהם.

ציר x – יהיה הערכים עצמם.

ציר y – שכיחות הערכים.

התפלגות סימטרית - כשמה כן היא, סימטרית!

בהתפלגות זו הממוצע והחציון שווים בדיוק.

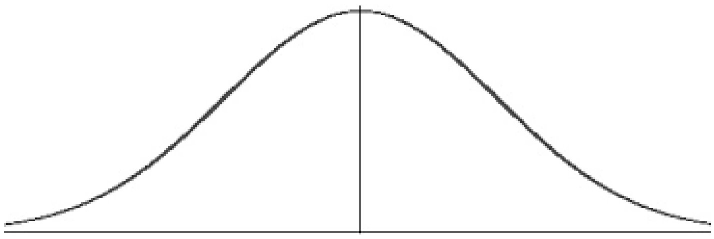
ואם נחצה/נקפל אותה לשניים בנקודת הממוצע והחציון, אזי צידה הימני יהיה תמונת מראה של צידה השמאלי.

התפלגות סימטרית נורמלית

מקרה פרטי של התפלגות סימטרית.

נקראת גם התפלגות פעמון.

בהתפלגות זו- כל מדדי המרכז שווים זה לזה ונמצאים במרכז ההתפלגות.



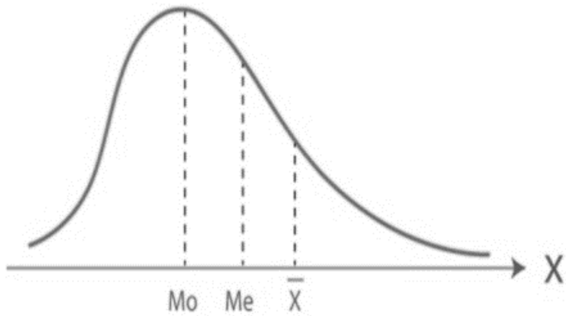
$$M_o = \overline{X} = md = MR$$

שימו לב: לא כל התפלגות סימטרית היא נורמלית.

התפלגות א-סימטרית חיובית (ימנית)

להתפלגות זו יש "זנב" ימני ארוך. כלומר, רוב הערכים (ציר X) הם נמוכים.

היחס בין שלושת מדדי המרכז הוא: **ממוצע** > **חציון** > **שכיח**

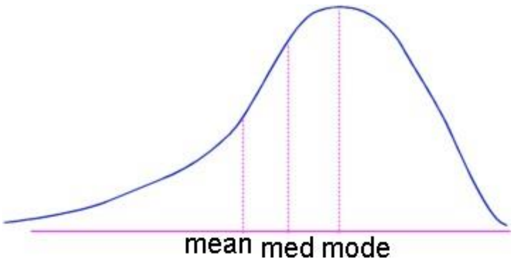


לדוגמא: התפלגות השכר במדינת ישראל / ציונים במבחן קשה.

התפלגות א-סימטרית שלילית (שמאלית)

להתפלגות זו יש "זנב" שמאלי ארוך. כלומר, רוב הערכים (ציר X) הם גבוהים.

היחס בין שלושת מדדי המרכז הוא: **שכיח** > **חציון** > **ממוצע**



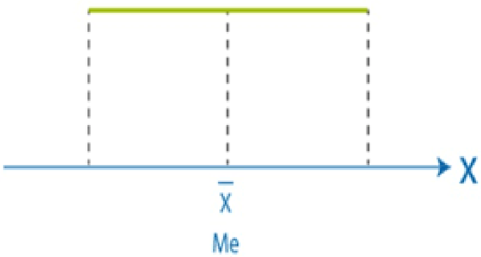
לדוגמא: ציונים במבחן קל. הרוב יקבלו ציונים גבוהים והמיעוט יקבלו ציון נמוך.

התפלגות אחידה

התפלגות זו היא גם מקרה פרטי של התפלגות סימטרית.

משום כך, **הממוצע והחציון שווים.**

בהתפלגות זו השכיחות של כל הערכים שווה, ולכן אין לנו שכיח.



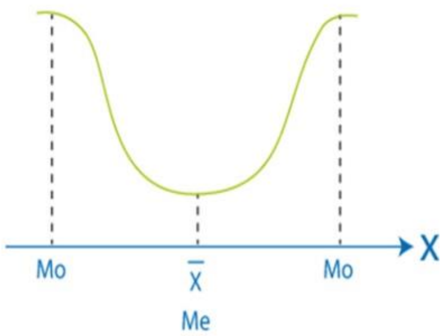
התפלגות U

התפלגות זו, גם היא, מקרה פרטי של התפלגות סימטרית.

והיא גם דו שיאית.

יש לה שני שכיחים וכן, הממוצע והחציון שווים ונמצאים במרכז ההתפלגות.

שכיח אחד קטן מהממוצע והחציון, והשכיח השני גדול מהממוצע והחציון.



דף נוסחאות מדדי מיקום יחסי

מבוא לנושא:

מדדים אלו מעריכים את המיקום **היחסי** של הנתונים במדגם. כלומר נסתכל על ערך מסוים **ביחס לערכים האחרים**.

לדוגמא: גובה משכורתו של אדם **ביחס למשכורות האחרות במשק**.
דוגמא נוספת: מה יותר טוב- לקבל 85 בסטטיסטיקה או 90 בפסיכולוגיה?

קיימים שני מדדים למיקום יחסי:

1. המאון (אחוזון)
2. ציון תקן

המאון במשתנה בדיד

המאון של תצפית שערכה x - מתאר את **אחוז התצפיות עד ל-x, כולל x**. המאון ה-C מסומן ב- X_c.

כלומר, זהו **הנתון ש $C\%$ מהנתונים קטנים או שווים לו.**

דוגמאות:

- חציון הוא המאון ה-50 – כי 50% מהנתונים קטנים או שווים לו.
- רבעון תחתון הוא המאון ה-25 – כי 25% מהנתונים קטנים או שווים לו.
- רבעון עליון הוא המאון ה-75 – כי 75% מהנתונים קטנים או שווים לו.
- עשירון תחתון הוא המאון ה-10 – כי 10% מהנתונים קטנים או שווים לו.
- עשירון עליון הוא המאון ה-90 – כי 90% מהנתונים קטנים או שווים לו.

נוסחא למציאת הערך במאון ה-C במשתנה בדיד:

המאון ה- C נמצא תמיד **במקום** ה- $\frac{c(n+1)}{100}$

לאחר שקיבלנו את המיקום נחפש בעמודת השכיחות המצטברת ונעצור במקום הראשון בו נתקלנו במיקום זה או במיקום גדול ממנו.

אם קיבלנו מיקום כמספר שלם, אזי הערך שנמצא במיקום זה הוא האחוזון.
אם קיבלנו מיקום שאינו מספר שלם, כלומר, מיקום בין 2 ערכים, נחשב את הממוצע בין ערכים אלו.

לחישוב אחוז שמקיים תנאי מסוים: (נקפיד לתת תשובה באחוזים %)

$$\frac{\text{נחשב את מס' התצפיות שמקיימות את התנאי}}{\text{סה"כ התצפיות}} \ X\ 100$$

<u>מאון במשתנה רציף (מחלקות מקובצות)</u>

במחלקות מקובצות נחשב אחוזונים רק <u>בגבול מחלקה</u> או <u>באמצע מחלקה</u>.

אחוזון באמצע מחלקה ניתן לחשב רק אם נתון לנו ש<u>**ההתפלגות בתוך המחלקה אחידה**</u>. (זה צריך להיות נתון בשאלה) אחרת, לא ניתן לחשב במדויק את האחוזון.

נוסחא למציאת הערך במאון ה-C במשתנה רציף:

המאון ה C נמצא תמיד **במקום** ה- $\frac{n\cdot c}{100}$

אחרי שקיבלנו את המיקום <u>**נחפש בעמודת השכיחות המצטברת**</u> ונעצור במקום הראשון בו נתקלנו במיקום זה או במיקום גדול ממנו.

כאשר נתון לנו שההתפלגות בתוך המחלקה היא <u>**אחידה**</u>, נוכל לחלק את המחלוקת בהתאם ל-מה ששאלו.

לחישוב אחוז שמקיים תנאי מסוים: (נקפיד לתת תשובה באחוזים %)

$$X\ 100\ \ \frac{\text{נחשב את מס' התצפיות שמקיימות את התנאי}}{\text{סה"כ התצפיות}}$$

כאשר בשאלה <u>**לא מדובר על כך שההתפלגות בתוך המחלקה היא אחידה**</u>:

נוסיף עמודה של שכיחות יחסית מצטברת באחוזים ונוכל לענות בקלות...

משכורת יומית X	**מס עובדים $f(x)$**	$F(x)$	שכיחות יחסית	שכיחות יחסית באחוזים	שכיחות יחסית באחוזים מצטברת
300-400	40	40	0.25	25%	25%
400-600	20	60	0.125	12.5%	37.5%
600-800	80	140	0.5	50%	87.5%
800-1200	20	160	0.125	12.5%	100%

<u>ציון תקן</u>

מסומן כ- Z_x.

ציון תקן עונה על השאלה **<u>"בכמה סטיות תקן מרוחק הנתון מהממוצע".</u>**

כלומר, מה הסטייה של ערך נתון בהשוואה לממוצע ביחידות של סטיית תקן.

הנוסחא למציאת ציון תקן:

$$\boxed{z_x = \frac{x-\bar{x}}{s_x}}$$

ממוצע של המדגם

ערך x

ציון התקן של ערך x

סטיית התקן של המדגם

<u>לדוגמא:</u>

אם לאחר הצבה בנוסחה נקבל את הערך 1, אזי נאמר שהנתון נמצא סטיית תקן אחת **<u>מעל</u>** הממוצע.

אם נקבל את הערך -1, אזי נאמר שהנתון נמצא סטיית תקן אחת **<u>מתחת</u>** לממוצע.

<u>בדף הנוסחאות זה נראה כך:</u>

מדדי מיקום יחסי: $Z_x = \frac{x-\bar{x}}{S_x}$

<u>תכונות ציון התקן</u>

1. כאשר $x = \bar{x}$ ציון התקן שווה לאפס.
2. כאשר $x > \bar{x}$ ציון התקן חיובי.
3. כאשר $x < \bar{x}$ ציון התקן שלילי.
4. ממוצע ציוני התקן של התפלגות נתונים תמיד שווה לאפס.
5. השונות וסטיית התקן של התפלגות ציוני התקן שווים תמיד ל- 1.

דף נוסחאות טרנספורמציות

מבוא לנושא:

נניח שכבר חישבנו מדדי מרכז ומדדי פיזור על התפלגות נתונים כלשהי.

כעת – **עושים שינוי כלשהו בנתונים** (טרנספורמציה = שינוי צורה)

נשאלת השאלה : **"מהם המדדים כעת, לאחר השינוי שנעשה? האם נצטרך לחשב שוב הכל מחדש?"**

והתשובה.... **לא!**

מה נעשה?

1. נאפיין את השינוי שבוצע בנתונים
2. בהתאם לשינוי נבצע התאמות במדדי המרכז והפיזור שכבר חישבנו לפני השינוי
3. נקבל את התשובה ישירות, כאמור – ללא חישוב מחדש!

טרנספורמציות מסוג שונים

נסתכל על שלושה סוגים של שינויים שניתן לבצע בנתונים :

1. **שינוי בכלל הנתונים באופן אחיד -** הוספה/החסרה/הכפלה/חלוקה בקבוע מסוים עבור כלל הנתונים.
2. **שינוי רק בחלק מהנתונים -** הוספה/החסרה/הכפלה/חלוקה בקבוע מסוים עבור חלק מהנתונים בלבד.
3. **שינוי במספר/כמות הנתונים -** הוספה או הסרה של נתונים ממדגם.

שינוי בכלל הנתונים באופן אחיד

1. **מוסיפים/מחסירים** ערך קבוע לכלל הערכים במדגם
2. **מכפילים/מחלקים** בערך קבוע את כלל הערכים במדגם.

מה קורה למדדים בכל מצב?

1. **מוסיפים/מחסירים** ערך קבוע לכלל הערכים במדגם
   - מדדי המרכז ישתנו באותו הערך בדיוק.
   - מדדי הפיזור לא ישתנו כלל. (שהרי צורת ההתפלגות לא השתנתה וכן ההבדל בין הערכים לא השתנה)
2. **מכפילים/מחלקים** בערך קבוע את כלל הערכים במדגם.
   - מדדי המרכז יגדלו/יקטנו בדיוק פי ערך הקבוע.
   - מדדי הפיזור יגדלו/יקטנו בדיוק פי ערך הקבוע. (חוץ מהשונות שתגדל פי ערך הקבוע **בריבוע**)

שינויים בנתונים והשפעותיהם	מדדי מרכז	מדדי פיזור
מוסיפים/מחסירים	**משתנה יחד עם השינוי**	**ללא שינוי**
מכפילים/מחלקים	**משתנה יחד עם השינוי**	**משתנה יחד עם השינוי (חוץ משונות שמשתנה בריבוע)**

שינוי בכלל הנתונים באופן אחיד - המשך

המדדים החדשים, אותם אנחנו רוצים לחשב, ניתנים לייצוג באמצעות הנוסחה הבאה :

$$\boxed{X' = b \cdot X + a}$$

המדד החדש

מקדם ההכפלה

המדד הישן

הערך שהתווסף
או שהוחסר

בדף הנוסחאות זה נראה ככה :

טרנספורמציות: אם $x' = b \cdot x + a$ אזי:

$Mo' = b \cdot Mo + a$ $\quad$ $Md' = b \cdot Md + a$ $\quad$ $MR' = b \cdot MR + a$

$\bar{x}' = b \cdot \bar{x} + a$

$s_{x'}^2 = b^2 s_x^2$ $\quad$ $s_{x'} = |b|\, s_x$

שינוי רק בחלק מהנתונים

כאשר משנים (הוספה/הפחתה/הכפלה/חילוק) חלק מהערכים במדגם.

לדוגמא : הוספת 5 נקודות רק לעשרת התלמידים שקיבלו את הציון הנמוך ביותר.

שינוי במספר/כמות הנתונים

הוספה של נתונים למדגם או הסרה של נתונים מהמדגם.

שינויים אלו הם אינם אחידים ולכן לא ניתן לקבוע כיצד יושפעו מדדי המרכז והפיזור.

במקרים אלה נאלץ **לנתח כל מדד לגופו**.

דוגמא ראשונה: אם נוספו למדגם שני תלמידים שציונם הוא גבוה מאוד או גבוה מהממוצע, נוכל לומר שהממוצע יעלה, אך כדי לדעת את ערך הממוצע החדש במדויק נצטרך לחשב מחדש.

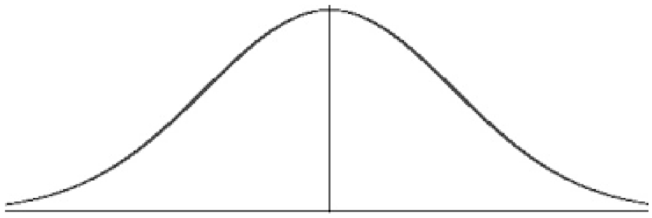
דוגמא שנייה: אם הסרנו ממדגם מסוים 5 נבדקים שערכם הוא לא הגבוה ביותר וגם לא הנמוך ביותר, נוכל לומר בוודאות שהטווח לא ישתנה.

דוגמא שלישית: אם ניקח במדגם מסוים את 2 הערכים הגבוהים ביותר ואת 2 הערכים הנמוכים ביותר ונשנה את ערכם לערך הממוצע, נוכל לומר כי השונות וסטיית התקן יקטנו כיוון שהפיזור קטן.

דף נוסחאות מבוא להתפלגות נורמלית

מבוא לנושא:

ההתפלגות הנורמלית היא ההתפלגות של תופעות בטבע כלומר, אם נבדוק שכיחות של תופעות מסוימות ונבנה גרף של ערכים אל מול שכיחות זו, נקבל צורת "פעמון" כזו:



דוגמאות:

- משקל תינוקת בלידה
- גובה ילדים בכיתה ג'
- ציוני מבחנים באוניברסיטאות
- ויש עוד המון...

התפלגות נורמלית היא התפלגות **סימטרית**. כלומר, כל מדדי המרכז בהתפלגות זו שווים זה לזה.

ממוצע= שכיח =חציון=אמצע טווח

ולכן תמיד 50% מהנתונים יהיו מעל הממוצע ו- 50% מהנתונים יהיו מתחת לממוצע. (חשוב!)

כיצד נזהה שאנחנו נמצאים בנושא הזה?

יהיו חייבים להגיד לנו שהנתונים בשאלה **"מתפלגים נורמלית"** / ניסוח אחר שכולל את הטיות של המילה "התפלגות" ואת המילה נורמלית.

1. "משקל הילדים מתפלג נורמלית...."
2. "גובה הגברים בעיר תל אביב מתפלג נורמלית...."
3. "להלן נתונים על התפלגות הציונים במבחן בהיסטוריה.... נתון כי הנתונים מתפלגים נורמלית..."

צורת הפעמון של ההתפלגות:

צורת ה"פעמון" בהתפלגות הנורמלית נקבעת ע"י שני מדדים: **ממוצע $\bar{x}$ וסטיית תקן S_x.**

תפקיד הממוצע $\bar{x}$ - לקבוע את מרכז ההתפלגות.

תפקיד סטיית התקן S_x - לקבוע את הפיזור סביב הממוצע. (כלומר האם הפעמון יהיה "רחב" או "צר").

<u>כמה הערות:</u>

- כל השטח מתחת לפעמון שווה ל-1 כלומר 100% מהמקרים.
- השטח משמאל לערך מסוים, הוא ההסתברות/הסיכוי לקבל אותו או ערכים קטנים ממנו.
- נהוג לסמן משתנה המתפלג נורמלית כך:

$$X \sim N(\overline{x}, S_x^2)$$

<u>התפלגות נורמלית סטנדרטית</u>

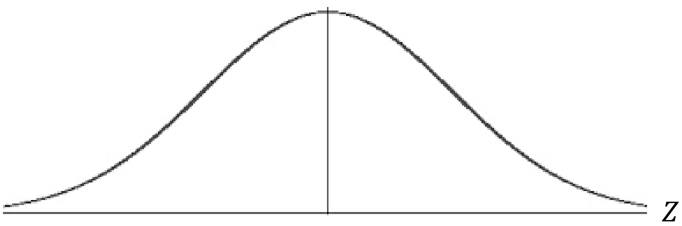
מקרה פרטי של התפלגות נורמלית נקרא <u>התפלגות נורמלית סטנדרטית</u>.

במקרה זה כביכול קיבלנו מקבץ נתונים אשר הממוצע שלהם, מרכז הפעמון, הוא 0 וסטיית התקן/השונות תהיה 1.

נכתוב כך: $Z \sim N(0,1)$

וזה ייראה כך:

ציר x – יתאר כעת ציוני תקן (Z)



שימו לב חשוב - השטח משמאל לנקודה Z יסומן על ידי $\emptyset(Z)$.

$\emptyset(Z)$ – השטח משמאל לנקודה Z

<u>הערה:</u> הטבלה שנלמד בהמשך בנויה על ההתפלגות הנורמלית הסטנדרטית וזה יהיה לנו מאוד נוח לשימוש.

ציון תקן + היכרות עם טבלת ההתפלגות הנורמלית

הדוגמא החשובה מבחינתנו של ערכים המתפלגים נורמלית סטנדרטית הם **ציוני התקן**.
נהוג לסמן ציון תקן של משתנה נורמלי סטנדרטי באות Z. ולעמים נסמן כך: Z_x.

טבלאות התפלגות

פונקציית ההתפלגות המצטברת של משתנה נורמלי סטנדרטי, $\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

הערכים שבשוליים הם ערכי ציוני התקן

הערכים בפנים מייצגים את ההסתברויות המתאימות לציוני תקן שבשוליים.

זה בעצם השטח שמתחת לעקומה ומשמאל לציון התקן.

טבלת עזר: z כפונקציה של $\Phi(z)$

$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z
.50	0	.91	1.341	.995	2.576
.55	.126	.92	1.405	.999	3.090
.60	.253	.93	1.476	.9995	3.291
.65	.385	.94	1.555	.9999	3.719
.70	.524	.95	1.645	.99995	3.891
.75	.674	.96	1.751	.99999	4.265
.80	.842	.97	1.881	.999995	4.417
.85	1.036	.98	2.054	.999999	4.753
.90	1.282	.99	2.326	.9999999	5.199

דף נוסחאות מעבר מהתפלגות נורמלית להתפלגות נורמלית סטנדרטית

התהליך נקרא תקנון או סטנדרטיזציה. (הפיכה לסטנדרטי...)

למה צריך אותו?

הנתונים שלנו בתרגילים ייתנו בהתפלגות נורמלית... אבל אנחנו נרצה להשתמש בטבלה... והטבלה מופיעה בנורמלית סטנדרטית...

לשם כך נהיה חייבים לבצע **מעבר/המרה מ"נורמלית" ל"נורמלית סטנדרטית"**.

נכיר נוסחא חשובה ביותר:

על מנת לעבור ממשתנה נורמלי X לציון תקן נעזר בנוסחה הבאה:

ערך התצפית

ממוצע

ציון תקן

$$\boxed{Z=\frac{x-\bar{x}}{s_x}}$$

סטיית תקן

תכונות ההתפלגות הנורמלית הסטנדרטית

תכונות ההתפלגות הנורמלית הסטנדרטית:

$$P(Z<z)=\phi(z)$$

$$P(Z>z)=1-\phi(z)$$

$$\phi(-z)=1-\phi(z)$$

לכל $a<b$:

$$P(a<Z<b)=\phi(b)-\phi(a)$$

$$P(-a<Z<a)=\phi(a)-\phi(-a)=2\cdot\phi(a)-1$$

תזכורת + הערות חשובות:

- בהתפלגות נורמלית תמיד 50% מהנתונים נמצאים מעל הממוצע ו- 50% מהנתונים נמצאים מעל הממוצע.
- בהתפלגות נורמלית **תמיד** במרחק של סטיית תקן אחת מהממוצע $\bar{x}\pm S_x$ נמצאים כ- 68% מהנתונים. (הכי מדויק: 68.26%)
- אם השטח מתחת לעקומה קטן מ-50% אזי Z שלילי.
- טווח בינריבעוני בהתפלגות נורמלית: $\boxed{2\cdot 0.674\cdot S_x}$

דף נוסחאות איך נפתור שאלות בהתפלגות נורמלית

1. זיהוי הנושא – בנתוני השאלה יהיה כתוב "מתפלג נורמלית".
2. נזהה את תת הנושא עליו שואלים ונפעל בהתאם – נחלק לשני מקרים :

מקרה 1:

נתון: ערך/ציון.

צריך למצוא: שטח/ההסתברות/הסיכוי/האחוז או שכיחות יחסית.

נראה דוגמא :

"מה אחוז התלמידים שגובהם גדול מ-155 ס"מ?"
"מה השכיחות היחסית של התינוקות השוקלים פחות מ 3 ק"ג?"
"מה אחוז הסטודנטים המשקיעים במטלות בין 10-20 שעות שבועיות?"

נבחן 3 מקרים :

מציאת $P(X<a)$ כאשר a הוא הערך הנתון:

1. נמצא את ציון התקן של הערך ע"י הצבה בנוסחא $Z=\frac{a-\bar{x}}{s_x}$.
2. נחפש בטבלה את ϕ ששייך לציון התקן.
3. נכתוב את התשובה שקיבלנו. (אם ביקשו אחוז, נכפיל ב100).

מציאת $P(X>a)$ כאשר a הוא הערך הנתון:

1. נמצא את ציון התקן של הערך ע"י הצבה בנוסחא $Z=\frac{a-\bar{x}}{s_x}$.
2. נחפש בטבלה את ϕ ששייך לציון התקן.
3. התשובה תהיה : $1-\phi(Z)$. (אם ביקשו אחוז, נכפיל ב100).

מציאת $P(a<X<b)$ כאשר a,b נתונים שניהם:

1. נמצא את השטח של a ע"י הצבת הערך a בנוסחא $Z=\frac{a-\bar{x}}{s_x}$
2. נחפש בטבלה את ϕ ששייך לציון התקן a.
3. נמצא את השטח של b ע"י הצבת הערך b בנוסחא $Z=\frac{b-\bar{x}}{s_x}$
4. נחפש בטבלה את ϕ ששייך לציון התקן b.
5. התשובה תהיה : חיסור השטח a מ b = $\phi\left(\frac{b-\bar{x}}{s}\right)-\phi\left(\frac{a-\bar{x}}{s_x}\right)$. (אם ביקשו אחוז, נכפיל ב100).

מקרה 2:

נתון: שטח/ההסתברות/הסיכוי/האחוז או שכיחות יחסית.

צריך למצוא: ערך/ציון.

נראה דוגמא :

"מהו מספר השעות ש- 80% מהילדים ישנים יותר ממנו?"
"מהו מספר שעות העבודה שרק 0.05 מהאוכלוסייה עובדת פחות ממנו?"

נבחן 2 מקרים :

מציאת $P(X<a)$ כאשר a:

1. נמצא בטבלה (בשטח בפנים) את ϕ שהכי קרוב לשטח הנתון בשאלה.
2. לפי המספר שמצאנו בסעיף 1, נמצא את ציון התקן (שנמצא בשוליים).
3. נציב את ציון התקן בנוסחא $Z=\frac{a-\bar{x}}{s_x}$ כאשר a הוא התשובה שאנחנו מחפשים.

מציאת $P(X>a)$ כאשר a:

1. נמצא בטבלה (בשטח בפנים) את ϕ שהכי קרוב ל-1 פחות השטח הנתון.
2. לפי השטח שמצאנו בסעיף 1, נמצא את ציון התקן (שנמצא בשוליים).
3. נציב את ציון התקן בנוסחא $Z=\frac{a-\bar{x}}{s_x}$ כאשר a הוא התשובה שאנחנו מחפשים.

הערה:
אם $\phi<0.5$ לא ניתן למצוא אותו בטבלה ולכן נמצא את $1-\phi$ וציון התקן יהיה **במינוס**.

דף נוסחאות סולמות מדידה + מדדי למדה וקרמר

מבוא לנושא:

רוצים למצוא האם יש קשר בין שני משתנים. כן/לא. קשר חזק/חלש וכו'.

דוגמאות:

1. קשר בין מין האדם לבין המפלגה לה הוא מצביע
2. קשר בין הקורס אותו סטודנט לומד לבין מידת ההנאה שלו
3. קשר בין מין האדם לבין מידת ה-IQ שלו

כיצד נבדוק זאת?

לכל סוג משתנה יש מדד קשר אחר שאותו נרצה לבדוק...

לכן, בשלב הראשון, נרצה להגדיר את סוג המשתנה.

את זה נעשה באמצעות **סולמות המדידה** שכבר למדנו.

סולמות מדידה יעזרו לנו להציג תכונות - בעזרת מספרים.

שם הסולם	מהות הסולם	דוגמאות
סולם שמי	**זהות בלבד** של פרט או איבר – ללא מידע נוסף. מילולי.	מספר ת.ז , מספר חולצה בכדורסל , ארץ מוצא , מספר אישי בצבא
סולם סדר	**זהות + סדר** - מי יותר ומי פחות, אך לא בכמה יותר או פחות. (גבוה, נמוך, בינוני)	סקר דירוג מ1-7, דרגות בצה"ל, שביעות רצון מקורס באוניברסיטה.
סולם מנה	**זהות + סדר + הבדל + יחס** - כלומר, התוספת כאן היא שאפשר לקבוע את גודל הפער בין המשתתפים ומהו היחס ביניהם. מספרי. כמות. **נדגיש, המספר 0 מציין היעדר מוחלט של התכונה.**	משקל, גובה, משכורת, שטח ועוד. "מספר ה..." , נסיעה במהירות 120 קמ"ש. נסיעה במהירות 0 לא קיימת.

מדדי קשר למשתנים שמיים

כאמור, נתון תסריט ובו שני משתנים. אנו בודקים האם יש קשר ביניהם.

כאשר אחד מהמשתנים לפחות **נתון בסולם שמי** (נתון באופן מילולי, מייצג זהות בלבד),

נשתמש **במדד למדה λ ובמדד קרמר r_c.**

נשים לב שמדד λ מתחלק לשניים: $\lambda_{y/x}$, $\lambda_{x/y}$

שימו לב - בכל שאלה כזו, אם לא נאמר לנו אחרת, **יש לחשב 3 מדדים**:

1. **למדה y לפי x - $\lambda_{y/x}$ (ניבוי y לפי שימוש ב- x)**
2. **למדה x לפי y - $\lambda_{x/y}$ (ניבוי x לפי שימוש ב- y)**
3. **מדד קרמר - r_c**

פתרון תרגילים בנושא זה יבוא לידי ביטוי בטבלת שכיחויות משותפת שנראית כך **(תמיד!)**:

			x \ y

במידה ולמשתנים הנבדקים יש כבר שמות באותיות (x, y) - נשתמש בהם לאורך התרגיל.

אם למשתנים הנבדקים אין שמות - נקרא למשתנה אחד x ולמשתנה האחר y (באופן שרירותי).

חישוב מדד למדה

בדף הנוסחאות:

$$\lambda_{x/y} = \frac{L_x - L_{x/y}}{L_x} \qquad \lambda_{y/x} = \frac{L_y - L_{y/x}}{L_y}$$

למדה מסוג 1	למדה מסוג 2
למדה y לפי x: (למדה y בתנאי ש- x / ניבוי y לפי x)	**למדה x לפי y: (למדה x בתנאי ש- y / ניבוי x לפי y)**
$$\lambda_{y/x} = \frac{L_y - L_{y/x}}{L_y}$$	$$\lambda_{x/y} = \frac{L_x - L_{x/y}}{L_x}$$
L_y – **סך הטעויות בניבוי y <u>ללא</u> שימוש ב- x** $L_{y/x}$ – **סך הטעויות בניבוי y <u>עם</u> שימוש ב- x** **שלבים במציאת $\lambda_{y/x}$:**	L_x – **סך הטעויות בניבוי x <u>ללא</u> שימוש ב- y** $L_{x/y}$ – **סך הטעויות בניבוי x <u>עם</u> שימוש ב- y** **שלבים במציאת $\lambda_{x/y}$:**
1. **נמצא את השכיח של y לפי הסה"כ.** 2. **שכיח – סך כל הנתונים $= L_y$** 3. **נשתמש עכשיו ב-x על מנת לנבא את y:** א. **נניח שמתקיים ערך מסוים של x ונסתכל רק בשורה שלו. נמצא את השכיח בשורה.** ב. **נחשב את סה"כ השורה פחות השכיח שמצאנו. (זו הטעות האפשרית)** ג. **נבצע שוב את סעיף א' עד שנגמרים ערכי x.** ד. **נחבר את כל הטעויות האפשריות שקיבלנו.** 4. **חיבור כל הטעויות שקיבלנו $= L_{y/x}$** 5. **מציבים בנוסחא שלמעלה למציאת $\lambda_{y/x}$** 6. **מוודאים שהתשובה שיצאה היא בין 0 ל-1.**	1. **נמצא את השכיח של x לפי הסה"כ.** 2. שכיח – סך כל הנתונים $= L_x$ 3. **נשתמש עכשיו ב-y על מנת לנבא את x:** א. **נניח שמתקיים ערך מסוים של y ונסתכל רק בעמודה שלו. נמצא את השכיח בעמודה.** ב. **נחשב את סה"כ העמודה פחות השכיח שמצאנו. (זו הטעות האפשרית)** ג. **נבצע שוב את סעיף א' עד שנגמרים ערכי y.** ד. **נחבר את כל הטעויות האפשריות שקיבלנו.** 4. חיבור כל הטעויות שקיבלנו $= L_{x/y}$ 5. **מציבים בנוסחא שלמעלה למציאת $\lambda_{x/y}$** 6. **מוודאים שהתשובה שיצאה היא בין 0 ל-1.**

ערכים אפשריים של λ הם: $0 \le \lambda \le 1$ כאשר:

- אם $\lambda = 0$ נגיד ש<u>**אין קשר**</u> בין המשתנים שבדקנו.
- אם $\lambda = 1$ נגיד ש<u>**יש קשר מלא**</u>. (קשר חזק מאוד) בין המשתנים שבדקנו.
- אם $\lambda = 0.8$ נגיד שזה כמו ציון 80.
- אם $\lambda = 0.6$ נגיד שזה כמו ציון 60.

מדד קרמר r_c

סימונים:

O – הנתונים שהתקבלו בניסוי.

E – הערך שהיינו מצפים לו (מחשבים לבד).

n – מספר הנתונים.

L – המספר הקטן מבין מספר השורות ומספר העמודות.

בדף הנוסחאות:

$$r_c = \sqrt{\frac{1}{n(L-1)}\chi^2} = \sqrt{\frac{1}{n(L-1)}\sum_i \frac{(O_i - E_i)^2}{E_i}}$$

שלבי החישוב:

1. חישוב הערכים הצפויים של כל משבצת בטבלה לפי הנוסחא:

$$E = \frac{\text{סה"כ הטור X סה"כ השורה}}{\text{סך כל הנתונים}}$$

2. חישוב "חי בריבוע" לפי הנוסחא הבאה:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

3. חישוב מדד קרמר לפי הנוסחא הבאה:

$$r_c = \sqrt{\frac{1 \cdot \chi^2}{n(L-1)}}$$

ערכים אפשריים למדד קרמר: $0 \le r_c \le 1$

- אם $r_c = 0$ נגיד ש**אין קשר** בין המשתנים שבדקנו.
- אם $r_c = 1$ נגיד ש**יש קשר מלא**. (קשר חזק מאוד) בין המשתנים שבדקנו.

המשך מדד קרמר r_c

מקרה פרטי של קרמר – מדד $\emptyset$:

אפשר להשתמש **רק כאשר יש טבלת נתונים של $2X2$ (!)**

$$\phi = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{(a \cdot d - b \cdot c)^2}{e \cdot f \cdot r \cdot k}}$$

$$r_c = \emptyset = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{(a \cdot d - b \cdot c)^2}{e \cdot f \cdot r \cdot k}}$$

כאשר:

	קטגוריה	**קטגוריה**	**סה"כ**
קטגוריה	a	b	r
קטגוריה	c	d	k
סה"כ	e	f	**סה"כ נתונים**

*לא צריך כאן לחשב את E_i. מציבים בנוסחא את הנתונים מהשאלה כמו שהם.

דף נוסחאות מדד ספירמן

מבוא לנושא:

משתמשים כאשר אחד המשתנים הוא מסולם סדר (דירוג) והמשתנה השני מסולם סדר או סולם מנה.
(העיקר שלא יהיה משתנה מסולם שמי כי אז נצטרך להשתמש במדדי למדה וקרמר).

שימו לב שמבחינה ויזאולית, הנתונים כאן תמיד יהיו **שתי שורות בלבד או שתי עמודות בלבד**,
(בניגוד לטבלת שכיחויות משותפת שמכוונת אותנו לשימוש במדדי למדה וקרמר).

			x
			y

ערכים אפשריים למדד ספירמן: $-1 \le r_s \le 1$

- אם $r_s = 1$ נגיד ש**יש קשר חיובי חזק מאוד** (קשר ישר) בין המשתנים שבדקנו.
- אם $r_s = -1$ נגיד ש**יש קשר שלילי חזק מאוד**. (קשר הפוך) בין המשתנים שבדקנו.
- אם $r_s = 0$ נגיד ש**אין קשר** בין המשתנים שבדקנו.

הערה:

אנחנו רואים שכאן נבדוק, לא רק האם יש קשר, אלא גם מה סוג הקשר: ישר או הפוך.

מדד ספירמן r_s - המשך

בדף הנוסחאות:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

שלבי החישוב:

1. R_x – דרוגות משתנה x. (מדרגים בסדר עולה)
   R_y – דרוגות משתנה y. (מדרגים בסדר עולה)

* שימו לב - אם יש נתונים זהים עושים ממוצע של הדרגות העוקבות.

נתוני התרגיל	x	50	42	84	50	16
	y	72	17	45	72	21
שלב 1	R_x	3.5	2	5	3.5	1
	R_y	4.5	1	3	4.5	2

2. מחשבים שורה חדשה של הפרשים: $d_i = R_x - R_y$.
3. מחשבים שורה חדשה של d_i^2.
4. מחשבים את $\sum d_i^2$

נתוני התרגיל	x	50	42	84	50	16
	y	72	17	45	72	21
שלב 1	R_x	3.5	2	5	3.5	1
	R_y	4.5	1	3	4.5	2
שלב 2	d_i	-1	1	2	-1	-1
שלב 3	d_i^2	1	1	4	1	1
שלב 4	$\sum d_i^2 =$	8				

5. הצבה בנוסחא: $r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ (כאשר **<u>n הוא מספר הזוגות הנתונים (העמודות)</u>**. אצלנו $n = 5$)

הערה:
אם ידוע ש- $r = 1$, אז $r_s = 1$
אבל לא בהכרח להיפך.

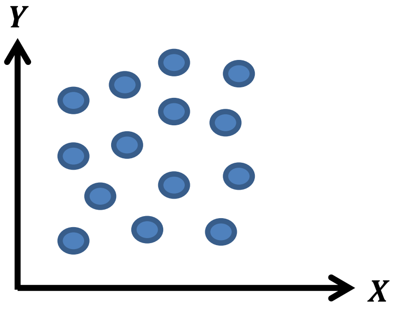
התשובה היא:

$$r_s = 1 - \frac{6 \cdot 8}{5(5^2-1)} = 1 - \frac{48}{120} = 0.6$$

דף נוסחאות מדד לקשר לינארי פירסון

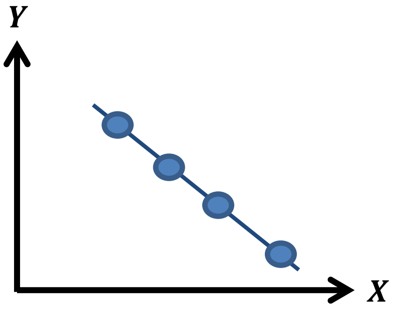
משתמשים כאשר שני המשתנים הם מסדר מנה (כמותי, מספרי).
משתנה אחד ייקרא X ואחד ייקרא Y. (חשוב מאוד להקפיד עליהם)

ערכים אפשריים: $-1 \le r \le 1$



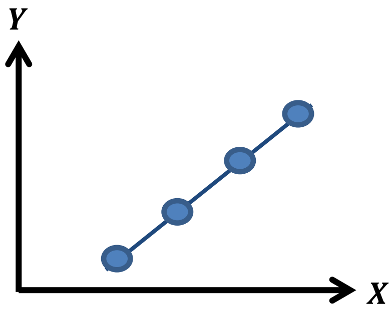
$r = 0$

אין קשר



$r = -1$

קשר שלילי חזק מאוד



$r = 1$

קשר חיובי חזק מאוד

תזכורת נוסחאות שנצטרך בחישובים:

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

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

$$S_x = \sqrt{S_x^2}$$

א. נחשב מתוך הנתונים - חמישה מדדים:

1. $\bar{x}$ - ממוצע של x
2. $\bar{y}$ - ממוצע של y
3. $\overline{x \cdot y}$ – ממוצע של מכפלת x ב- y. (נבנה בעבור זה טור חדש)
4. S_x - סטיית תקן של x
5. S_y - סטיית תקן של y

דוגמא לנתונים שיכולים להופיע:

x	y	
4	8	
7	6	
10	1	

או נתונים מספריים עם סיגמאות

$$\overline{x \cdot y} = \frac{\sum_i x_i \cdot y_i}{n}$$

ב. נציב את המדדים בנוסחאות הבאות ונמצא את r:

שונות משותפת

$$Cov(x, y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y}$$

אם המונה חיובי אז r חיובי
אם המונה שלילי אז r שלילי

$$r = \frac{Cov(x,y)}{S_x \cdot S_y}$$

זה מקדם פירסון
גם כאן וגם בטבלה שבעמוד הבא

סיימנו. מצאנו את המקדם...

קו ניבוי/קו תחזית/קו רגרסיה

קו ניבוי / קו תחזית / קו רגרסיה:

מצאנו שישנו קשר בין שני משתנים.

עכשיו נותנים לנו ערך מסוים של אחד המשתנים, ומבקשים מאיתנו לנבא בעזרתו את הערך של המשתנה השני.

לשם כך אנחנו צריכים לבנות **קו ניבוי** שיתאר את הקשר בין המשתנים:

ערך שרוצים לנבא **ערך שרוצים לנבא**

קו ניבוי ל-y	קו ניבוי ל-x
$\tilde{y} = b \cdot x + a$ $b = \frac{r \cdot S_y}{S_x}$ $a = \bar{y} - b \cdot \bar{x}$	$\tilde{x} = b' \cdot y + a'$ $b' = \frac{r \cdot S_x}{S_y}$ $a' = \bar{x} - b' \cdot \bar{y}$

נזכור שכבר מצאנו את המקדם r וגם את:

1. $\bar{x}$ - ממוצע של x
2. $\bar{y}$ - ממוצע של y
3. $\overline{x \cdot y}$ – ממוצע של מכפלת x ב- y. (נבנה בעבור זה טור חדש)
4. S_x - סטיית תקן של x
5. S_y - סטיית תקן של y

כל מה שנותר הוא לזהות מה המשתנה שמחפשים לנבא ולהציב בנוסחאות המתאימות...

ככה זה נראה בדף הנוסחאות:

$$\tilde{y} = bx + a \quad ; \quad b = \frac{rs_y}{s_x} \quad ; \quad a = \bar{y} - b\bar{x}$$

$$\tilde{x} = b'y + a' \quad ; \quad b' = \frac{rs_x}{s_y} \quad ; \quad a' = \bar{x} - b'\bar{y}$$

שונות הניבויים ושונות הטעויות

השונות הכללית, זו השונות הרגילה שאנחנו מכירים מהקורס, (S_x^2) מתחלקת לשני מרכיבים:

$$\boxed{\text{שונות הטעויות} + \text{שונות הניבויים} = \text{שונות כללית}}$$

S_x^2 ,S_y^2
השונות שאנחנו מכירים עד היום

שונות הניבויים = פיזור הנתונים על קו הניבוי.

שונות הטעויות = פיזור הנתונים סביב קו הניבוי.

שונות הניבויים בניבוי y לפי x	**שונות הניבויים בניבוי x לפי y**
סימון שונות הניבויים - $S_{\hat{y}}^2$ **נוסחא לחישוב שונות הניבויים של y לפי x:** $\boxed{S_{\hat{y}}^2 = r^2 \cdot S_y^2}$ **נוסחא לחישוב שונות הטעויות:** שונות הטעויות = $\boxed{(1-r^2)\cdot S_y^2 = S_y^2 - S_{\hat{y}}^2}$	**סימון שונות הניבויים** - $S_{\hat{x}}^2$ **נוסחא לחישוב שונות הניבויים של x לפי y:** $\boxed{S_{\hat{x}}^2 = r^2 \cdot S_x^2}$ **נוסחא לחישוב שונות הטעויות:** שונות הטעויות = $\boxed{(1-r^2)\cdot S_x^2 = S_x^2 - S_{\hat{x}}^2}$

ככה זה נראה בדף הנוסחאות:

$$\boxed{r^2 = \frac{S_{\hat{y}}^2}{S_y^2}} \qquad \boxed{r^2 = \frac{S_{\hat{x}}^2}{S_x^2}}$$

אם שואלים מהו אחוז השונות המוסברת, נחשב כך: $r^2 \cdot 100$

נניח ש $r = 0.4$ אז $r^2 = 0.16$.

במקרה כזה נגיד ש:

16% מהשונות של y היא שונות הניבויים (השונות המוסברת ע"י קו הניבוי)

84% מהשונות של y היא שונות הטעויות (השונות הלא מוסברת)

דף נוסחאות הסתברות - מושגים פעולות בין מאורעות וריבוע הקסם

נושא עיקרי בקורס.

1. **שאלה פתוחה במבחן – 25 נקודות – (99%)**
2. **סעיף או שניים בשאלות "נכון/לא נכון" – 5-10 נקודות**
3. **סעיף או שניים בתוך שאלות פתוחות אחרות – 5-10 נקודות**

נושא הסתברות מכיל בתוכו מספר נושאים:

- מושגים כלליים בהסתברות
- קומבינטוריקה
- הסתברות מותנית
- התפלגות בינומית
- משתנה מקרי
- משתנה מקרי בינומי

בתוך נושא הסתברות נכיר כלים ויזואליים שיעזרו בפתרון תרגילים כמו: ריבוע הקסם, דיאגרמת ון ועצי הסתברות.

מושגים בהסתברות

הסתברות = "מה הסיכוי שמאורע כלשהו יתרחש?"

מרחב המדגם Ω - אוסף כל התוצאות האפשריות של ניסוי **מקרי.**

ניסוי מקרי - ניסוי שקיימות לו מספר תוצאות אפשריות שונות ולא ניתן לדעת איזו תוצאה אפשרית תתרחש בפועל.

מאורע - קבוצה חלקית של מרחב המדגם. נהוג לסמן ב- A, B, C...

מאורע ריק - מאורע שאינו מכיל איברים ממרחב המדגם ומסמנים אותו ב: $\emptyset$

חישוב הסתברות של מאורעות

הסתברות של מאורע A נסמן כך: $P(A)$

$$\boxed{P(A)=\frac{\text{מספר התוצאות שמכיל מאורע } A}{\text{מספר התוצאות שמכיל מרחב המדגם}}=\frac{n(A)}{n(\Omega)}}$$

התוצאה של חישוב ההסתברות חייבת, כמובן, להיות בין 0 ל 1.

כלומר נזכור תמיד ש- $\boxed{0\le P\le 1}$

<u>פעולות בקבוצות/מאורעות</u>

מאורע משלים – בהינתן המאורע A, אזי המאורע המשלים של A הוא קבוצה המכילה את האיברים במרחב המדגם שאינם ב-A.

המשלים של מאורע A יסומן כך : A^C או $\bar{A}$.

לדוגמא :

$$\Omega = \{1,2,3,4,5,6,7,8,9,10\}$$
$$A = \{1,3,5,7,9\}$$
$$A^c = \{2,4,6,8,10\}$$

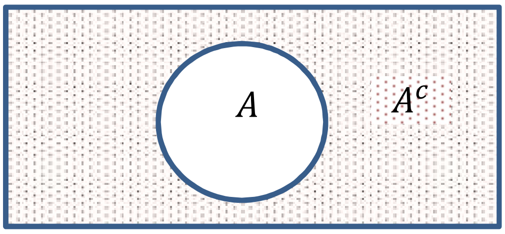
מכאן נסיק כי: $P(A) + P(A^c) = 1$

$$P(A^c) = 1 - P(A)$$

נהוג להציג את הפעולות בקבוצות בדיאגרמת ון – דרך ויזואלית לפתור שאלות בהסתברות.

עבור כל פעולה שנלמד, נציג אותה ויזואלית ע"י דיאגרמת ון.

דיאגרמת ון:



חיתוך בין מאורעות – בהינתן מאורעות A ו B אזי חיתוך המאורעות הוא קבוצה המכילה את האיברים הנמצאים <u>**גם**</u> <u>**ב-A וגם ב-B**</u> ורק אותם.

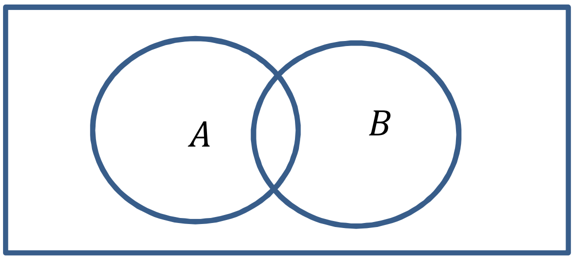
כלומר, האיברים המשותפים לשתי הקבוצות. ניתן לומר גם A וגם B התרחשו יחד.

חיתוך בין מאורעות יסומן כך : $A \cap B$

לדוגמא:

$$A = \{1,2,3,4,5\}$$
$$B = \{1,6,8,10\}$$
$$A \cap B = \{1\}$$

דיאגרמת ון:



מאורעות זרים (מאורעות "מוציאים" זה את זה)

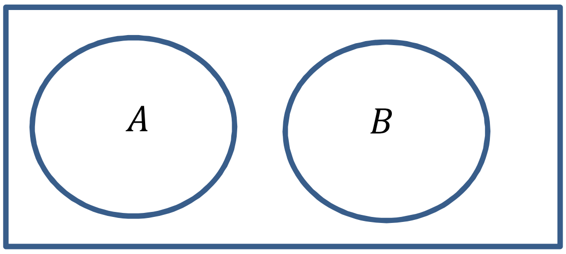
אם לשני מאורעות אין איברים משותפים, אזי **המאורעות זרים**. כלומר, אינם יכולים להתרחש ביחד.

נסמן זאת כך: $\boldsymbol{A \cap B = \emptyset}$

לדוגמא:

$$\boldsymbol{A = \{1,2,3,4\}}$$
$$\boldsymbol{B = \{5,6\}}$$
$$\boldsymbol{A \cap B = \emptyset}$$

דיאגרמת ון:



איחוד בין מאורעות – בהינתן מאורעות A ו-B אזי איחוד המאורעות הוא קבוצה המכילה את האיברים הנמצאים או בA או בB או בשתיהן.

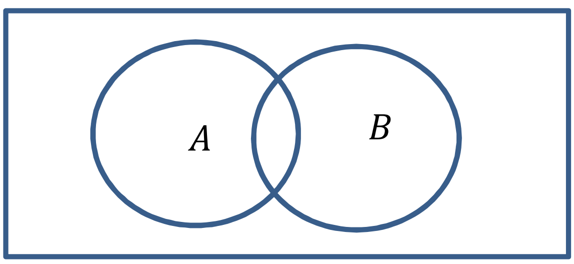
כלומר, <u>**האיברים ששייכים לפחות לאחת מהקבוצות**</u>. ניתן לומר גם ש **או** A **או** B או שניהם התרחשו.

איחוד בין מאורעות יסומן כך: $\boldsymbol{A \cup B}$.

לדוגמא:

$$\boldsymbol{A = \{1,2,3\}}$$
$$\boldsymbol{B = \{1,4,5,6\}}$$
$$\boldsymbol{A \cup B = \{1,2,3,4,5,6\}}$$

דיאגרמת ון:



<u>כלל החיבור</u>

$$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

ריבוע הקסם

	מאורע A	מאורע A^c	
מאורע B			$P(B)$
מאורע B^c			$P(B^c)$
1	$P(A)$	$P(A^c)$	**1**

ארבעת התאים שבתוך הטבלה מייצגים ארבעה מאורעות זרים שכולם מרכיבים את מרחב המדגם כולו ולכן חיבור של ההסתברויות שלהם שווה 1.

דף נוסחאות שיעור קומבינטוריקה

מבוא לנושא + חוק המכפלה

קומבינטוריקה = דרך מקוצרת למנייה מהירה של מספר אפשרויות. קומבינציות... שילובים... אופציות סידור...

חוק המכפלה (עקרון הכפל)

כאשר יש ניסוי המורכב מ-K שלבים,

ובשלב 1 יש n_1 תוצאות אפשריות,

ובשלב 2 יש n_2 תוצאות אפשריות,

ובשלב 3 יש n_3 תוצאות אפשריות,

וכך עד לשלב ה-k ובו n_k תוצאות אפשריות,

אזי מספר התוצאות הכולל האפשרי הוא מכפלת מספר התוצאות האפשריות בכל אחד מהשלבים.

$$\boxed{\boldsymbol{n = n_1 \cdot n_2 \cdot n_3 \cdot \ldots \ldots \ldots . . \cdot n_k}}$$

הערה: בכל פעם שנגיד את התחילית "ו.." זו תהיה פעולת **כפל**.

עוד הערה: בכל פעם שנגיד את התחילית "או.." זו תהיה פעולת **חיבור** (נראה בהמשך).

תכונות של מדגם

בפתרון שאלות בקומבינטוריקה אנו בעצם דוגמים K איברים מתוך קבוצה בת N איברים.

על מנת לפתור כמו שצריך נהיה חייבים לשים לב לשתי תכונות של המדגם:

1. **סדר** - האם יש חשיבות לסדר האיברים או לא.
   מדגם סדור – מדגם שבו **יש** משמעות וחשיבות לסדר:
   ספרות במספר טלפון, מספרים בעלי 3 ספרות (527 לא שווה ל-572)
   מדגם לא סדור – מדגם שבו **אין** חשיבות לסדר: אם דגמנו פעם 12 ופעם 21 זה נחשב לתוצאה אחת.
   (אותה התוצאה)

2. **החזרה** - האם איבר שנדגם, מוחזר חזרה לקבוצה ויכול להדגם בשנית, או שאיבר שנדגם נשאר בחוץ ולא יכול להידגם שנית.
   עם החזרה - איבר שנדגם חוזר חזרה לקבוצה ויכול להדגם שנית.
   בלי החזרה - איבר שנדגם לא חוזר חזרה לאוכלוסייה ונשאר בחוץ ולכן לא יכול להדגם פעמיים.

מדגם סדור עם החזרה

דוגמים איבר מאוכלוסייה בת N איברים, רושמים את התוצאה **ומחזירים בחזרה את האיבר לאוכלוסייה** ודוגמים שוב ולכן אותו איבר שיצא פעם קודמת יכול לצאת שוב.

במדגם סדור עם החזרה של K איברים מתוך קבוצה/אוכלוסייה בת N איברים מסי׳ האפשרויות הוא:

$$\boxed{n = n \cdot n \cdot \ldots . n = n^k}$$

מדגם סדור ללא החזרה

דוגמים איבר מאוכלוסייה בת N איברים **ולא מחזירים אותו בחזרה לאוכלוסייה**. לאחר מכן דוגמים איבר שני מאוכלוסייה בת $N-1$ איברים, וגם אותו לא מחזירים לאוכלוסייה חזרה ודוגמים איבר שלישי מאוכלוסייה בת $N-2$ איברים, וכך חוזרים על הניסוי K פעמים.

במדגם סדור ללא החזרה של K איברים מתוך קבוצה/אוכלוסייה בת N איברים מסי׳ האפשרויות הוא:

$$\boxed{N_K = N \cdot (N-1) \cdot (N-2) \cdot (N-3) \cdot \ldots (N-K+1)}$$

חלק ב: הסתברות

קומבינטוריקה:	$(N)_k = N * (N-1) * (N-2) * \ldots \ldots (N-K+1)$

מדגם לא סדור ללא החזרה

דוגמים K איברים מתוך אוכלוסייה בת N איברים ללא חשיבות לסדר וללא החזרה.

כדי להבין נתחיל בדוגמא: בכמה דרכים ניתן לבחור 4 חולצות שונות מתוך 10 דגמים?

אין משמעות לסדר. זה לא מעניין איזה חולצה תהיה ראשונה וכמובן שאין החזרה כי אנו לא רוצים לבחור באותה חולצה פעמיים?

במקרה זה, מספר האפשרויות הוא:

$$\boxed{\binom{n}{k} = \frac{n!}{k!\,(n-k)!}}$$

תמורות

סידור של N איברים **בשורה או טור** נקרא תמורה.

מספר האפשרויות לסידור כזה של N איברים בשורה או טור:

$$\boxed{n \cdot (n-1) \cdot (n-2) \ldots \cdot 1 = n!}$$

דף נוסחאות הסתברות מותנית+עצי הסתברות+מאורעות בלתי תלויים

הסתברות מותנית = ההסתברות למאורע, בהתחשב בכך שמאורע אחר כבר התרחש.

הסתברות מותנית מצמצמת את מרחב המדגם כיוון שכבר ידוע על מאורע שהתקיים.

הגדרה:

בהינתן שני מאורעות A ו- B, ההסתברות של מאורע B כאשר ידוע שמאורע A כבר קרה היא:

$$\boldsymbol{P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{n(A\cap B)}{n(A)}}$$

ידוע

עצי הסתברות

דיאגרמת עץ היא כלי עבודה. בעזרתו נוכל לתאר בצורה ויזואלית הסתברויות של מקרים שונים.

השאלה הנשאלת - מתי נשתמש בדיאגרמת עץ?

והתשובה - כאשר נתון לנו ניסוי בעל, לפחות, שני שלבים. כמו"כ כאשר יש לנו הסתברויות מותנות.

בניית העץ:

1. משורש העץ נוציא מספר ענפים כמספר האפשרויות בשלב הראשון של הניסוי.
   מכל ענף שבשלב א' נוציא ענפים כמספר האפשרויות השונות בשלב ב' וכן הלאה עבור כל שלבי הניסוי.
   גובה העץ הוא כמספר שלבי הניסוי.
2. על כל ענף נרשום את ההסתברות לקבלת התוצאה המתאימה.
3. סך ההסתברויות בכל התפצלות של ענף שווה ל-1.
4. כל המסלולים בעץ זרים זה לזה.
5. כל מסלול מתחילת העץ ועד סופו שווה למכפלת כל ההסתברויות הרשומות על הענפים לאורך המסלול המתאים לתוצאה.
6. הסתברות של מאורע מסוים היא סכום של כל המסלולים המתאימים למאורע זה.

סוגי שאלות:

א. הסתברות לאורך ענף (כפל) – המילה "וגם"
ב. הסתברות שמופיעה בסוף העץ, במספר ענפים שונים (חיבור) – המילה "או"
ג. הסתברות מותנית

מאורעות בלתי תלויים

מאורעות בלתי תלויים הם מאורעות שההסתברות של מאורע אחד אינה משפיעה על ההסתברות של המאורע האחר.

לדוגמא: הוצאת שני כדורים מהכד עם החזרה, הטלת קובייה...

בהינתן שני מאורעות A ו- B בלתי תלויים זה בזה... כלומר, $P(B/A)=P(B)$

אזי מתקיים $\boldsymbol{P(A\cap B)=P(A)\cdot P(B)}$

דף נוסחאות התפלגות בינומית

מבוא+מתי נשתמש:

על מנת להשתמש **בנוסחת ההתפלגות הבינומית** לחישוב הסתברויות חייבים להתקיים מספר תנאים.

תנאים אלה יהיו, בעצם, נתוני השאלה עצמה.

אם כך, זיהוי הנושא "התפלגות בינומית" יתבצע לאחר שקראנו את השאלה, וזיהינו התאמה בין הנתונים לבין התנאים.

תנאי השאלה, שחייבים להתקיים, על מנת שנפנה לנושא "התפלגות בינומית" הם:

1. **מבוצעות מספר חזרות על אותו ניסוי**. כלומר, סדרה של n ניסויים. לדוגמא: יגידו לנו שמטילים קובייה 5 פעמים.
2. **בכל ניסוי יש שתי תוצאות אפשרויות בדיוק**: "הצלחה" בהסתברות של p או "כישלון" בהסתברות של $1-p$.
3. **החזרות על הניסוי הן בלתי תלויות זו בזו.**

נוסחת הבינום:

$$\boxed{p(x=k)=\binom{n}{k}P^k(1-p)^{n-k}}$$

$$\boxed{\binom{n}{k}=\frac{n!}{k!\,(n-k)!}}$$

תפקידי הפרמטרים:

n- מספר החזרות על הניסוי. זה יכול להיות גם גודל המדגם.

x- מספר "ההצלחות". $x=0,1,2\ldots,n$.

p- ההסתברות "להצלחה"

q- ההסתברות "לכישלון"

דף נוסחאות משתנה מקרי

מאפיינים של משתנה מקרי

1. גודל מספרי . לדוגמא: הטלת מטבע לא יכול להיות.
2. תוצאה של ניסוי מקרי.

נהוג לסמן את המשתנה המקרי באותיות ... X, Y, Z

את המשתנה המקרי מציגים בעזרת פונקציית הסתברות:

פונקציית הסתברות של משתנה מקרי

נציג כך:

רשימת הערכים ש X יכול לקבל	X	$X_1, \ldots\ldots, X_n$	
ההסתברות לכל ערך X	P	$P(X_1), \ldots\ldots, P(X_n)$	סכום ההסתברויות חייב להיות שווה ל1 $\Sigma_{p_i} = 1$

תוחלת של משתנה מקרי

תוחלת (ממוצע) של משתנה מקרי - מסומנת ב-$E(X)$ או μ.

הנוסחה:

$$\boxed{\boldsymbol{E(x) = \sum x_i p(x_i) = \mu}}$$

שונות/סטיית תקן של משתנה מקרי

מסומנת ב-$V(X)$ או σ^2.

הנוסחה:

$$\boxed{\boldsymbol{V(x) = \sum x_i^2 p(x_i) - \mu^2}}$$

נוסחה נוספת:

$$\boxed{\boldsymbol{V(x) = \sum (x_i - \mu)^2 p(x_i)}}$$

סטיית תקן של משתנה מקרי - כידוע: $\boldsymbol{\sigma = \sqrt{V(X)}}$

תכונות מדדים של משתנה מקרי

תכונות התוחלת:

בדומה לממוצע שלמדנו בהתפלגות שכיחויות של נתונים:

1. אם המשתנה מקבל ערך קבוע a תמיד, אזי $E(a) = a$
2. $E(X+a) = E(X) + a$ – הוספת הקבוע a לכל ערכי המשתנה, תגדיל את התוחלת בa
3. $E(bX) = bE(X)$- הכפלת כל ערכי המשתנה בקבוע b, תגדיל את התוחלת פי b.
4. $E(X+Y) = E(X) + E(Y)$- תוחלת של שני משתנים מקריים שווה לסכום התוחלת של שניהם.

תכונות השונות:

בדומה לשונות שלמדנו בהתפלגות שכיחויות של נתונים:

1. עבור כל משתנה מקרי X , $V(X) \geq 0$.
2. אם המשתנה מקבל ערך קבוע a תמיד, אזי $V(a) = 0$
3. הוספת הקבוע a לכל ערכי המשתנה, לא תשנה את השונות.
4. הכפלת כל ערכי המשתנה בקבוע b, תגדיל את השונות פי b^2 ואת סטיית התקן פי b.
5. אם X וY משתנים מקריים בלתי תלויים, אזי $V(X+Y) = V(X) + V(Y$

משמעות התוחלת במשחקי מזל

נגדיר את X- הרווח במשחק / הרווח הנקי

$E(X)$- תוחלת הרווח

$E(X) > 0$ – כדאי לשחק

$E(X) < 0$- לא כדאי לשחק

$E(X) = 0$- משחק הוגן

דף נוסחאות משתנה מקרי בינומי

ריענון - התפלגות בינומית

על מנת להשתמש **בנוסחת ההתפלגות הבינומית** לחישוב הסתברויות חייבים להתקיים מספר תנאים.

1. **מבוצעות מספר חזרות על אותו ניסוי**. כלומר, סדרה של n ניסויים. לדוגמא: יגידו לנו שמטילים קובייה 5 פעמים.
2. **בכל ניסוי יש שתי תוצאות אפשריות בדיוק**: "הצלחה" בהסתברות של p או "כישלון" בהסתברות של $1-p$.
3. **החזרות על הניסוי הן בלתי תלויות זו בזו**.

נוסחת הבינום:

$$p(x=k)=\binom{n}{k}P^k(1-p)^{n-k}$$

$$\boxed{\binom{n}{k}=\frac{n!}{k!\,(n-k)!}}$$

תפקידי הפרמטרים:

n- מספר החזרות על הניסוי. זה יכול להיות גם גודל המדגם.

x- מספר "ההצלחות". $x=0,1,2....,n$.

p- ההסתברות "להצלחה"

q- ההסתברות "לכישלון"

חישוב תוחלת, שונות/סטיית תקן

מגדירים את X - מספר ההצלחות בניסוי. אם X מתפלג בינומית:

התוחלת:

$$\boxed{E(X)=np}$$

השונות:

$$\boxed{V(X)=npq}$$

סטיית התקן:

$$\boxed{\sigma=\sqrt{V(X)}}$$

כלומר, אין צורך לחשב את פונקציית ההסתברות וניתן ישירות להציב בנוסחאות לעיל.