סיכום בחשיבה הסתברותית בחיי יום-יום - שאלון 005

פרק זה הינו חלק מסיכום כולל לשאלון 005 שנכתב על-ידי מאיר בכור

הסתברות: - הסיכוי שמאורע מסויים יקרה

טבלה דו-מימדית (לפי פרופורציה):

	A	$\overline{A}$	
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
	$P(A)$	$P(\overline{A})$	1

לפי השורות בטבלה מתקיים (מלמעלה למטה):

$$P(B) = P(\overline{A} \cap B) + P(A \cap B)$$
$$P(\overline{B}) = P(\overline{A} \cap \overline{B}) + P(A \cap \overline{B})$$
$$1 = P(\overline{A}) + P(A)$$

לפי העמודות בטבלה מתקיים (מימין לשמאל):

$$P(A \cap B) + P(A \cap \overline{B}) = P(A)$$
$$P(\overline{A} \cap B) + P(\overline{A} \cap \overline{B}) = P(\overline{A})$$
$$P(B) + P(\overline{B}) = 1$$

בטבלה דו-מימדית לפי פרופורציה ניתן להשתמש בשברים או בשבר עשרוני אך לא מומלץ השימוש באחוזים.

ניתן לבנות טבלה דו-מימדית לפי שכיחויות, כאשר במקום P יופיע הסימן N, במקום 1 (שהוא $P(S) = 1$) יופיע $N(S)$ (מרחב המדגם).

בשכיחות מתקיים:

$$0 \le N(A) \le N(S)$$
$$N(\overline{A}) = N(S) - N(A)$$
$$N(A \cup B) = N(A) + N(B) - N(A \cap B)$$

בפרופורציה מתקיים:

$$0 \le P(A) \le 1$$
$$P(\overline{A}) = 1 - P(A)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

<u>הסתברות מותנה</u>

1. מהי ההסתברות לקבל בזריקת קוביה את הספרה 2 ?
   מרחב המדגם הוא $\{1, 2, 3, 4, 5, 6\}$ ולכן הסיכוי הוא $\frac{1}{6}$.

2. מהי ההסתברות לקבל בזריקת קוביה את הספרה 2 <u>אם ידוע</u> שהמספר קטן מ- 5?
   כעת מרחב המדגם הצטמצם ל- $\{1, 2, 3, 4\}$ ולכן הסיכוי הוא $\frac{1}{4}$.

לפיכך, ברגע שמשהו ידוע (כמו בדוגמא) מצטמצם מרחב המדגם לידוע בלבד ואת היתר (בדוגמא – הספרות 5 , 6) ניתן למחוק; כלומר, עוברים כעת לקבוצת ייחוס אחרת – חדשה.

קבוצת הייחוס החדשה כעת היא B. $P(A/B) = \frac{P(A \cap B)}{P(B)}$.
(קבוצת הייחוס היא תמיד בצד ימין של ההתניה).

יש לשים לב שגם בהתניה יש משלים: $P(\overline{A}/B) = 1 - P(A/B)$.

נוסחאת בייס: $P(A \cap B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(A)$ (ניתן לפתור תרגילים גם בלי להשתמש בה)

<u>איך מזהים שזו התניה ?</u> (A/B)
כאשר בשאלה יש את המילים:- "אם ידוע", "מבין", "מתוך", "בתנאי", "מ-", "בהינתן" (מילים אלו מכוונות לקבוצת הייחוס החדשה).

<u>איך מזהים שזה חיתוך?</u> $(A \cap B)$
כאשר בשאלה יש את המילים: - "וגם", "ו-".

<u>איך מזהים שזה איחוד?</u> $(A \cup B)$
כאשר בשאלה יש את המילה "או".

כאשר מבקשים תרגום מילולי לפרופורציה מסויימת מאד מומלץ להשתמש במילים המזהות שלעיל בהתאם לסוג הפרופורציה (פעולה הפוכה לשאלה "איך מזהים שזו התניה, חיתוך, איחוד?)
למשל כאשר מבקשים תרגום מילולי ל- $P(A/B)$ נשתמש במילים המזהות <u>התניה</u> כמו:
"...<u>מבין</u>...." , "...אם ידוע ש..."

<u>כאשר פותרים שאלה בחשיבה הסתברותית יש:-</u>

א. להגדיר בצורה מילולית את שתי הקבוצות/תכונות והמשלימים שלהן, כולל את מרחב המדגם (S):

S:- ....
A:- ....
$\overline{A}$:- ....
B:- ....
$\overline{B}$:- ....

וזאת גם אם בשאלה הגדירו כבר את הקבוצות/התכונות A ו- B (או באותיות אחרות).

אם A ו- B אינן מוגדרות בשאלה, על-מנת להקל עלינו לזהות מה הן מייצגות, מומלץ "לשלוף אדם" ממרחב המדגם ולשאול אותו שתי שאלות המתייחסות לנתוני השאלה.
(האם אתה שייך ל..., האם אתה בעד....).
"תשובותיו" ייתנו גם את $\overline{A}$ ו- $\overline{B}$.
לפעמים, מתוך סעיפי השאלה ניתן לזהות את הקבוצות השונות.

ב. לרשום את נתוני השאלה בצורה מתמטית (בפרופורציה ו/או בשכיחות).
למשל - $P(A)=0.7$, $P(A/B)=0.3$, $N(A)=25$.

ג. למלא את הטבלה בהתאם.
יש להראות את הדרך, כולל ההצבה בנוסחאות והחישובים המתאימים.
(אם רק ממלאים את הטבלה ולא מפרטים את הדרך זה עלול להיחשב <u>כ"חשד להעתקה"!)</u>

ד. לענות על הסעיפים השונים שבשאלה.
כאשר אתם משתמשים בנתון הלקוח מתוך הטבלה יש לרשום לידו <u>"לפי הטבלה"</u>.

<u>שימו לב!</u>
יש שאלות שבהן נתונות שתי התניות בלבד – בשאלות מסוג זה יהיה שימוש בנעלם או נעלמים.
לפעמים שימוש בנוסחא: $P(A/B)=1-P(\overline{A}/B)$ מצמצמת את הנעלמים.

כאשר שואלים לדוגמא:-
הראו בעזרת חישובים מתאימים למי יש סיכוי גדול יותר לקבל אחזקת פלאפון ("התכונה"), לגבר (קבוצה A) או לאישה (קבוצה $\overline{A}$) ?:

ההשוואה תמיד תהיה – אותה "תכונה" (אחזקת פלאפון) בקבוצות השונות (גבר, אישה):
P(אישה/אחזקת פלאפון) <u>לעומת</u> P(גבר/אחזקת פלאפון)

בצד שמאל של ההתניה תהיה "התכונה", בצד ימין של ההתניה תהיינה הקבוצות השונות.

<u>ניסוח התשובה יכול להיות באחת מהאפשרויות הבאות:</u>

א. אחוז / שיעור "התכונה" <u>מבין</u> קבוצה A (גדול/קטן/שווה) מאשר אחוז / שיעור "התכונה" <u>מבין</u> קבוצה $\bar{A}$.

ב. <u>מבין</u> קבוצה A אחוז / שיעור "התכונה" (גדול/קטן/שווה) מאשר <u>מבין</u> קבוצה $\bar{A}$.

ג. פרופורציית "התכונה" <u>מבין</u> קבוצה A (גדולה/קטנה/שווה) מאשר פרופורציית "התכונה" <u>מבין</u> קבוצה $\bar{A}$.

<u>שימו לב</u> לשימוש במילה הרלוונטית להתניה – "מבין".

<u>קשר סטטיסטי</u>

הקדמה:- בחיי היום-יום אנו מתעניינים לא אחת בשאלות מסוג:

א. האם יש קשר בין הכנת שיעורי בית לבין הצלחה בבחינות?
ב. האם יש קשר בין הצלחה במבחן הפסיכומטרי לבין הצלחה באוניברסיטה?
ג. האם יש קשר בין גובה האב לגובה הבן?

ניקח למשל את דוגמא ג' לעיל:
בדוגמא זו אנו שואלים בעצם האם אנו יכולים לנבא טוב יותר את גובה הבן כאשר ידוע גובה האב.
כלומר, שאלת הקשר ושאלת הניבוי כרוכות זו בזו ומכאן חשיבות הנושא גם בחיי היום-יום.

A קשור סטטיסטית ל- B אם מתקיים אחד מהתנאים הבאים:-

א. $P(A/B) \neq P(A)$
ב. $P(A \cap B) \neq P(A) \cdot P(B)$
ג. $P(A/B) \neq P(A/\bar{B})$

כלומר – אם זה <u>שונה</u> אז <u>יש</u> קשר סטטיסטי.

מתנאים א' ו-ב' ניתן להסיק <u>רק</u> האם יש או אין קשר סטטיסטי.
תנאי ג' נותן גם את משמעות הקשר – את כיוון הקשר (מי גדול יותר).

שים לב שתנאי ג' הוא בעצם "אותה תכונה בקבוצות השונות".

ניסוח אחר שאפשרי לגבי קשר סטטיסטי:- האם יש תלות? – כלומר האם יש קשר סטטיסטי.

<u>קשר סיבתי</u>

<u>הגדרות: -</u>

<u>קשר סיבתי:</u>
האם יש הסבר הגיוני (סיבה) לקשר הסטטיסטי בין שני גורמים.

<u>הצמדת גורמים:</u>
התופעה שבה יש לשני גורמים (A ו- B) קשר סטטיסטי עם גורם שלישי (C) ולכן נוצר, בדרך עקיפה, גם קשר סטטיסטי בין שני הגורמים.
הגורם השלישי נקרא <u>גורם מתווך</u>.

הערה:
כאשר יש הצמדת גורמים הקשר הסטטיסטי בין A ו- B הוא קשר <u>שלא עומד בזכות עצמו</u> אלא <u>תלוי</u> בגורם C .

<u>נטרול גורמים</u>:
הפעולה שנעשית כדי להתגבר על הצמדת גורמים.
מנטרלים את השפעתו של הגורם החשוד כגורם מתווך (C) על-ידי בדיקה האם קיים קשר סטטיסטי בין A ל- B בכל אחת מהקבוצות C ו- $\bar{C}$.

שאלה בנושא קשר סיבתי כוללת בדרך כלל מספר סעיפים (סדר הסעיפים יכול להשתנות משאלה לשאלה):

- **סעיף בו תקבלו נתונים (או תתבקשו למצוא נתונים) הנוגעים לקשר הסטטיסטי בין A ל – B .**

- **סעיף בו תועלנה טענות כלשהן לגבי השפעה אפשרית של גורם מתווך ותתבקשו להסביר כיצד הגיעו לטענות הנ"ל, ו/או לבדוק האם הטענות נכונות או לא.**

- **סעיף בו תתבקשו להסביר את הסתירה שהתקבלה (אם התקבלה) בין הקשר הסטטיסטי שהתקבל בין A ל- B לפני נטרול הגורם המתווך ובין הקשר הסטטיסטי שהתקבל בין A ל- B לאחר נטרול הגורם המתווך.**

בכל סעיף עליכם לרשום <u>מסקנה</u> הכוללת <u>תיאור מילולי</u> של הקשר הסטטיסטי בין הגורמים ומשמעותו (הכיוון שלו). בסוף השאלה עליכם להביא <u>סיכום</u> המסביר את התוצאות שהתקבלו תוך שימוש במושגים: "הצמדת גורמים" , "גורם מתווך" , "נטרול גורמים" , "קשר סיבתי" ואם רלוונטי, גם את המושג "היפוך הקשר" ("הפרדוקס של סימפסון") .
תיאור מילולי פירושו לתאר את הקשר הסטטיסטי תוך שימוש מדוייק בשם הקבוצה.

נבהיר את שלבי הפתרון ודרך כתיבת המסקנות בדוגמא כללית:

<u>שלבי הפתרון לשאלה בנושא קשר סיבתי</u>:

1. בדיקת הקשר הסטטיסטי בין A ל- B ומשמעותו.

לדוגמא: נניח שמצאתם בעזרת חישובים, או שנאמר לכם כי בין A ל- B קיים קשר סטטיסטי עם המשמעות: - $P(A/B) > P(A/\bar{B})$.

2. מעלים טענה שקיים גורם C המשפיע על הקשר בין A ל- B.
הינכם מתבקשים לבדוק האם הטענה נכונה וכיצד משפיע הגורם C על הקשר הסטטיסטי בין A ל- B:

א. נבדוק קודם כל שקיים קשר סטטיסטי בין A ל- C ובין B ל- C .

<u>שימו לב</u>!
תוצאות סעיף זה יכולות לשמש אתכם בהסבר התוצאות ובניסוח הסיכום.

ב. <u>בשלב זה עליכם לבצע נטרול גורמים</u> .

בדרך כלל בשלב נטרול הגורמים יהיה עליכם לעשות שימוש בשתי טבלאות המראות את הקשר בין A ל- B בקבוצה C ואת הקשר בין A ל- B בקבוצה $\bar{C}$.

3. ניתוח התוצאות, מסקנות וסיכום:

א. אם אחרי הנטרול נמצא <u>שאין קשר סטטיסטי</u> בין A ל- B בכל אחת מהקבוצות C ו- $\bar{C}$, <u>המסקנה</u> תהיה:
אין קשר סיבתי בין A ל- B.
C היה גורם מתווך שגרם להצמדת גורמים ויצר רושם כי קיים קשר סטטיסטי בין A ל- B .
<u>הסיכום יהיה</u>:
התברר כי אין קשר סטטיסטי בין A ל- B וגם אין קשר סיבתי ביניהם.
C הוא גורם מתווך שגרם להצמדת גורמים והוא שיצר את הרושם שיש קשר סטטיסטי בכיוון מסויים בין A ל- B.

ב. אם אחרי הנטרול נמצא <u>שיש קשר סטטיסטי</u> בין A ל-B <u>והקשר שומר על המשמעות</u> שהייתה לו לפני נטרול הגורמים (אי-השוויון עדיין גדול: $P(A/B) > P(A/\bar{B})$) ,
<u>המסקנה</u> תהיה:
יש קשר סיבתי בין A ל- B ואין שינוי במשמעות הקשר.
מכאן ש- C <u>לא</u> היה גורם מתווך (כלומר, C לא השפיע על הקשר בין A ל- B).

<u>הסיכום יהיה:</u>

התברר כי הקשר הסטטיסטי בין A ל- B נשמר גם לאחר הנטרול, כלומר <u>לכאורה</u> יש קשר סיבתי בין A ל- B והגורם C לא היה גורם מתווך ולכן לא שינה את משמעות הקשר, אך ייתכן <u>שיש גורמים מתווכים אחרים</u> (שאותם לא בדקנו).

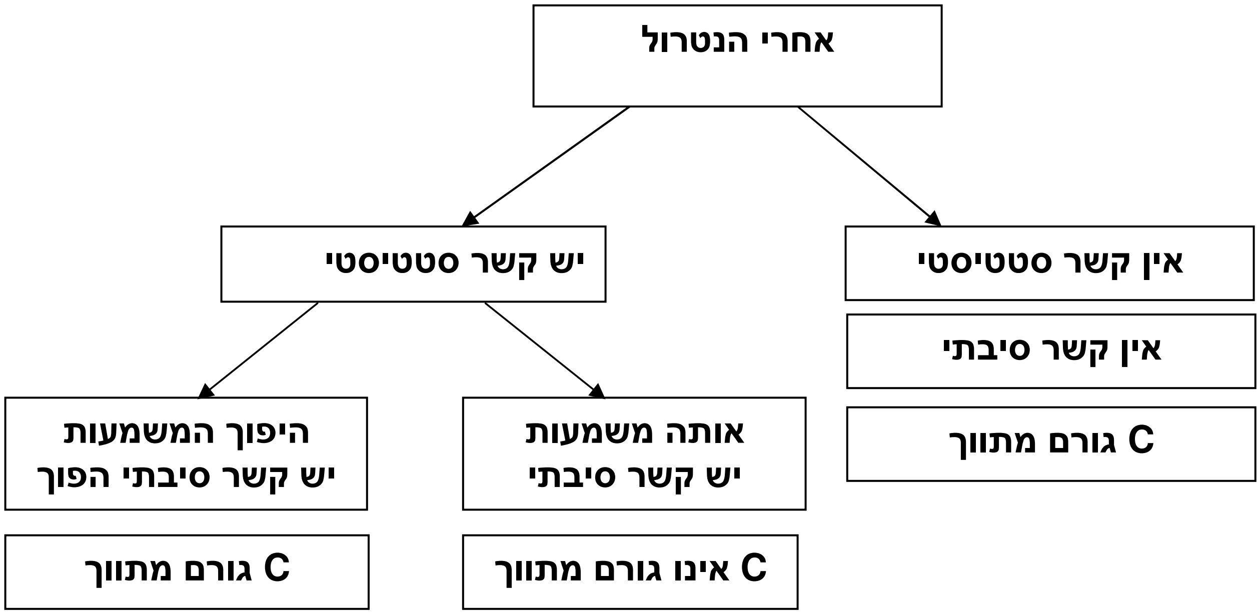
ג. אם אחרי הנטרול נמצא <u>שיש קשר סטטיסטי</u> בין A ל- B אך <u>עם משמעות הפוכה</u> (אי-השוויון התהפך: $(P(A/B) < P(A/\bar{B})$) <u>המסקנה</u> תהיה:
יש קשר סיבתי בין A ל- B אך בעקבות נטרול הגורמים חל <u>היפוך הקשר</u> הסטטיסטי (הפרדוקס של סימפסון). C הוא גורם מתווך שגרם להצמדת גורמים <u>ויצר את הרושם</u> של קשר סטטיסטי בכיוון מסויים בעוד שהקשר האמיתי היה בכיוון ההפוך.

<u>שימו לב!</u>
הכיוון האמיתי של הקשר הסטטיסטי בין A ל- B מתגלה אחרי נטרול הגורמים.

<u>הסיכום יהיה:</u>
הסיכום יכיל את המסקנה (ראה לעיל בסעיף זה) בתוספת הסבר כיצד הקשר הסטטיסטי בין A ל- C והקשר הסטטיסטי בין B ל- C (ראה סעיף 2 א') גרם ליצירת הרושם (המוטעה) שמצאנו לפני הנטרול.

מאד חשוב שבסיכום ייעשה שימוש במילים: **"קשר סיבתי" , "גורם מתווך", "הצמדת גורמים", "נטרול גורמים", "היפוך הקשר"**

<u>לסיכום</u>



<u>על-מנת לזכור</u>!:

<u>מבחינת קשר סיבתי</u> –

אם אחרי הנטרול <u>יש</u> עדיין קשר סטטיסטי (כלומר יש סימן של אי-שיוויון) אז <u>יש</u> קשר סיבתי.
אם אחרי הנטרול <u>אין</u> קשר סטטיסטי (כלומר יש סימן של שיוויון) אז <u>אין</u> קשר סיבתי.

<u>מבחינת C הגורם המתווך</u> -

אם אחרי הנטרול יש <u>אותה</u> משמעות (אי-השיוויון נשמר באותו כיוון) אז C <u>אינו</u> גורם מתווך.
בכל מקרה אחר - C גורם מתווך!

אם יש קשר סיבתי חייב להיות גם קשר סטטיסטי אבל לא להיפך! **(אם יש קשר סטטיסטי לא בהכרח יש קשר סיבתי)**

<u>שיפוט על-פי יציגות</u>

בחיי היום-יום אנו משתמשים במושגים מתורת ההסתברות, למשל: "אין סיכוי ש.." , אני בטוח ב-99% ש...."
במשפטים מסוג זה אנו מבטאים בעצם עד כמה אנחנו בטוחים שמשהו יקרה או לא יקרה, או במילים אחרות – את רמת הביטחון שלנו להתרחשות מאורע בעתיד.

תורת ההסתברות מתאימה לכל מאורע ערך מספרי המבטא את מידת הסבירות (רמת הביטחון) לכך שהמאורע יתרחש.

יש שתי גישות עיקריות למציאת הסתברות:- הגישה הסטטיסטית (המתמטית) והגישה האינטואיטיבית (סובייקטיבית):

<u>הגישה הסטטיסטית (המתמטית):</u>
בגישה הסטטיסטית נחפש את השכיחות היחסית של המאורע וההסתברות תהיה הערך אליו מתקרבת השכיחות היחסית עבור מספר ניסיונות השואף לאינסוף.

לדוגמא: אם נשאל "מה הסיכוי שבזריקת קובייה יתקבל המספר 3?" נענה על פי הגישה המתמטית: כיוון שלכל מספר בקובייה יש סיכוי שווה להתקבל, הרי שאם נטיל את הקובייה מספר רב מאד של פעמים השכיחות היחסית של המספר 3 תתקרב מאד ל- $\frac{1}{6}$.

<u>הגישה האינטואיטיבית (סובייקטיבית):</u>
אנו ניתן סיכוי למאורע להתרחש על-פי מידת האמונה שלנו (רמת הבטחון) בהתרחשותו, כלומר – על פי השיפוט האינטואיטיבי שלנו.
שיפוט אינטואיטיבי הוא בעצם שיפוט סובייקטיבי ולכן לאנשים שונים תהיינה הסתברויות אינטואיטיביות שונות לאותו אירוע.

לדוגמא:
אם נשאל מהו הסיכוי שחיסון חדש ימגר מחלה מסוימת נקבל תשובות שונות מאנשים שונים. רופאים וודאי יתנו סיכוי גבוה יותר לחיסון, אנשים מהשורה יתנו סיכויים שונים בהתאם לאמון שהם רוחשים לרופאים ובהתאם לניסיונם האישי.....

דוגמא להבדל בתוצאה כאשר משתמשים בשתי הגישות השונות:

ידוע כי אחוז ההצלחה בקורס מסוים עומד על 75%. מה ההסתברות שמיכל, סטודנטית בקורס, תסיים את הקורס בהצלחה?
אם המשיב הוא אדם אקראי, סביר להניח שהוא יאמר שההסתברות להצלחתה של מיכל היא 75%, אך אם המשיבה היא חברתה של מיכל המכירה את יכולותיה הטובות , סביר להניח שתשובתה תהיה גבוהה יותר למשל – 90%.

המשיב האקראי הסתמך על הנתון היחיד שהיה לו (אחוז ההצלחה הידוע) ואילו חברתה של מיכל השתמשה בגישה הסובייקטיבית והוסיפה לנתון המתמטי את מידת האמון שלה ביכולותיה של מיכל.

גורמים פסיכולוגיים משפיעים על החלטותינו כמעט בכל תחום בחיינו
("הפסיכולוגיה של הכלכלה") – שכל לעומת רגש.
למשל, אוהד כדורגל "שרוף" יתערב עם חברו על ניצחון קבוצתו גם אם היא במקום האחרון בטבלה ומשחקת נגד הראשונה בטבלה....

כלומר, כאשר משתמשים בגישה האינטואיטיבית/סובייקטיבית אנו מפעילים "מנגנונים" שונים כמו "תחושות בטן", רגשות, מידע סובייקטיבי - שיכולים לגרום לנו לטעויות בשיפוט.
טעויות מסוג זה נקראות: "כשלים בשיפוט אינטואיטיבי".

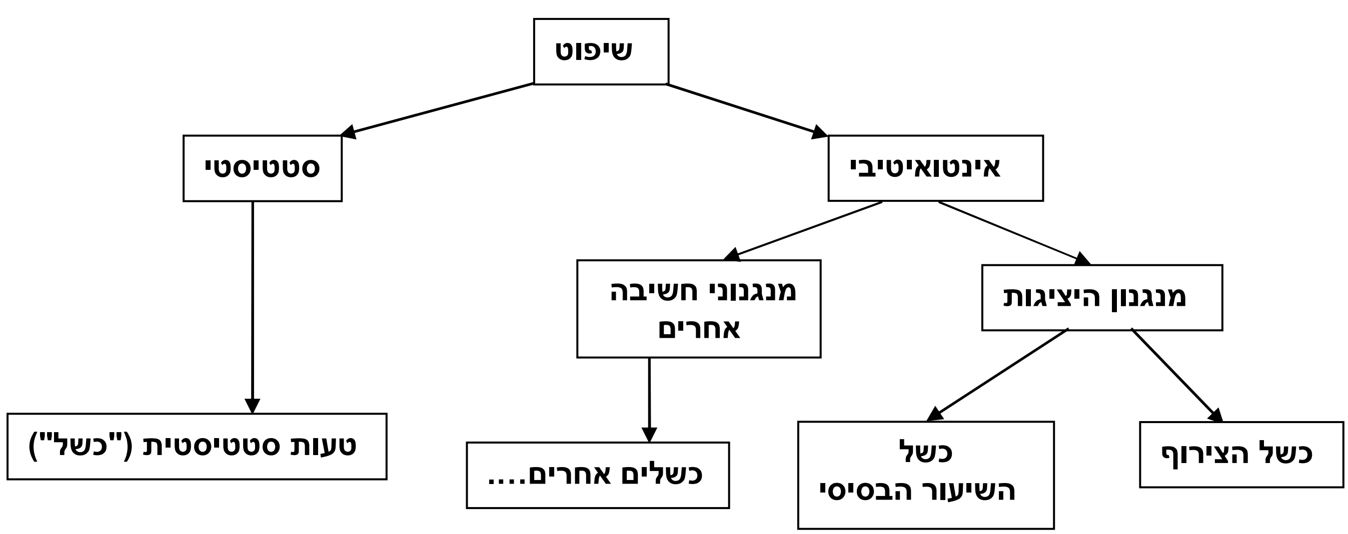
אחד המנגנונים המשפיעים על השיפוט האינטואיטיבי שלנו נקרא: מנגנון היציגות.

<u>מנגנון היציגות</u>

לפעמים, כאשר נתונים לנו שני מאורעות, אחד מהם <u>יראה</u> לנו "הגיוני" יותר, סביר יותר שיקרה ולכן נייחס לו הסתברות גבוהה יותר להתרחשות מאשר למאורע השני. במילים אחרות – מאורע אחד <u>ייצג טוב יותר</u> בעינינו את מה שיכול לקרות.
תהליך החשיבה הנ"ל נקרא "מנגנון היציגות".

<u>שימו לב!</u>
המוח שלנו מחפש את כל מה <u>שנראה</u> לנו הגיוני יותר והוא יבחר במשהו ייצוגי יותר על-פי ידע קודם, תחושות בטן, רגשות וכד', מבלי להתחשב תמיד במה שנכון מתמטית!
יוצא, שפעמים רבות אנחנו נוטים "להיתפס לתיאור" בלי להתייחס לכל הנתונים.

יש שני סוגים של כשלים (טעויות) שעלולים לקרות בגלל מנגנון היציגות:
"כשל הצירוף" ו"כשל השיעור הבסיסי".



<u>כשל הצירוף</u>

נתייחס לדוגמא הבאה:
בני הוא עורך דין מצליח וידוע כחובב ריקודים מושבע.
איזה מהמשפטים הבאים נראה מייצג יותר את בני?

א. בני הוא עורך-דין המתמחה בדיני חברות גדולות.
ב. בני הוא עורך דין המתמחה בדיני חברות גדולות ומשתתף בחוג ריקודים פעמיים בשבוע.

רבים מאיתנו יאמרו כי משפט ב' הוא הסביר יותר וזאת מכיוון שלאור האינפורמציה המקדימה שקיבלנו על בני, <u>נראה לנו הגיוני יותר</u> שמשפט ב' <u>מייצג</u> את בני טוב יותר ממשפט א'. אנחנו "נתפסים לתיאור" המקדים של בני – כלומר מתבססים על <u>מנגנון היציגות</u> ולכן טועים בתשובה!!!

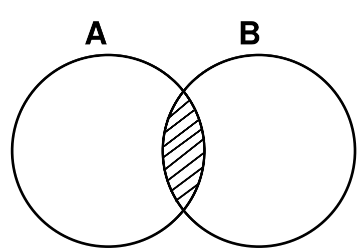
אם נבדוק את הדוגמא בצורה מתמטית, נגלה כי משפט א' הוא הסביר יותר משום <u>שלאפשרות יחידה תהיה תמיד הסתברות גבוהה יותר מאשר לצירוף שתי אפשרויות - זהו "כלל הצירוף"</u>.

תשובה ב' <u>נוגדת</u> בעצם את כלל הצירוף ולכן אינה נכונה.

<u>נסביר</u>:

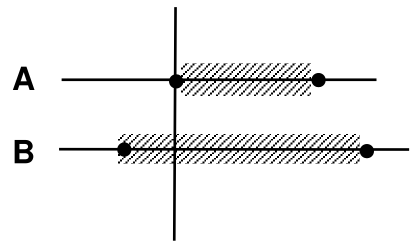
נסמן: A - קבוצת עורכי הדין המתמחים בדיני חברות גדולות.
B - קבוצת המשתתפים בחוג ריקודים פעמיים בשבוע.

<u>נראה את כלל הצירוף בדרך גרפית:</u>

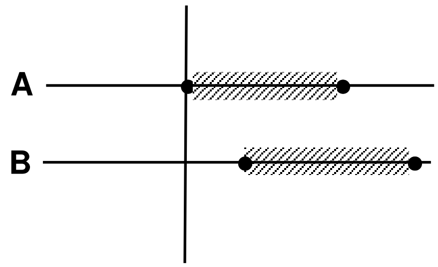


מהשרטוט ברור כי $P(A \cap B) < P(A)$.
אם ניתן הסתברות גבוהה יותר למשפט ב' יהיה פירוש הדבר כי: $P(A \cap B) \geq P(A)$
וזה נוגד את כלל הצירוף.

<u>ובדרך אלגברית</u> (כמו באי-שיוויונים):



$$P(A \cap B) = P(A)$$



$$P(A \cap B) < P(A)$$

מחוקי הפרופורציה וחיתוך (צירוף) מאורעות: $\boxed{P(A \cap B) \le P(A)}$

וגם בדרך זו ניתן לראות כי אם ניתן הסתברות גבוהה יותר למשפט ב' יהיה פירוש הדבר: $P(A \cap B) \ge P(A)$ וזה נוגד את כלל הצירוף.

> **<u>לסיכום</u>:**
> **כשל הצירוף קורה כאשר השיפוט שאנו עושים מתבסס על מנגנון היציגות ומעניק לצירוף של שתי אפשרויות הסתברות גבוהה יותר מאשר לאפשרות יחידה– וזה נוגד את כלל הצירוף.**

<u>כשל השיעור הבסיסי</u>

על מנת להבין מהו כשל השיעור הבסיסי עלינו להגדיר מספר מושגים חדשים:

1. **<u>השיעור הבסיסי של A</u>:** $\dfrac{P(A)}{P(\bar{A})}$

השיעור הבסיסי של A מוגדר כיחס בין הפרופורציה של A לבין הפרופורציה של המשלים של A שהוא $\bar{A}$.

2. **<u>דיאגנוסטיות של B</u>:** $\dfrac{P(B/A)}{P(B/\bar{A})}$

דיאגנוסטיות פירושה – יכולת אבחון.

הדיאגנוסטיות של התכונה B מוגדרת כיחס בין פרופורציית בעלי תכונה B מבין קבוצה A לבין פרופורציית בעלי תכונה B מבין קבוצה $\bar{A}$.
במילים אחרות, אנו בודקים עד כמה התכונה B מאבחנת את הקבוצה A.

<u>נסביר:</u>

כאשר השוונו בין שתי הסתברויות אמרנו "אותה תכונה בקבוצות השונות".
כאשר בדקנו קשר סטטיסטי בדקנו האם: $P(B/A) \neq P(B/\bar{A})$
(להזכירכם – התכונה תמיד תופיע מצד שמאל של ההתניה..).
במידה ש: $P(B/A) = P(B/\bar{A})$ פירוש הדבר שאין קשר סטטיסטי.

<u>שימו לב!</u>
ההגדרה של דיאגנוסטיות היא בעצם חילוק של שני האגפים זה בזה, אגף שמאל באגף ימין.

אם מתקיים $1 = \frac{P(B/A)}{P(B/\bar{A})}$ אז: $P(B/A) = P(B/\bar{A})$, כלומר – אין קשר סטטיסטי.

או במילים אחרות – התכונה B אינה מאבחנת כלל את קבוצה A.

ככל שהדיאגנוסטיות גדולה מ-1 אז פרופורציית תכונה B בקבוצה A גדולה מפרופורציית תכונה B בקבוצה $\bar{A}$:

$$1 > \frac{P(B/A)}{P(B/\bar{A})} \quad \text{אז:} \quad P(B/A) > P(B/\bar{A})$$

ולהיפך: ככל שהדיאגנוסטיות קטנה מ-1 אז פרופורציית תכונה B בקבוצה A קטנה מפרופורציית תכונה B בקבוצה $\bar{A}$:

$$1 < \frac{P(B/A)}{P(B/\bar{A})} \quad \text{אז:} \quad P(B/A) < P(B/\bar{A})$$

3. <u>היחס המעודכן:</u> $R = \frac{P(A/B)}{P(\bar{A}/B)}$

מוגדר כיחס בין פרופורציית תכונה A מתוך קבוצה B לבין פרופורציית חסרי תכונה A (המשלים של A שהוא $\bar{A}$) מתוך קבוצה B.

<u>שימו לב!</u>
בעצם R הוא יחס של שני חיתוכים:

$$R = \frac{P(A/B)}{P(\bar{A}/B)} = \frac{\frac{P(A \cap B)}{P(B)}}{\frac{P(\bar{A} \cap B)}{P(B)}} = \frac{P(A \cap B)}{P(\bar{A} \cap B)}$$

ולכן כאשר הטבלה מוכנה, ניתן למצוא את R בצורה מיידית.

כזכור, בנוסחת התניה מתקיים:

$$P(B/A)=\frac{P(A\cap B)}{P(A)} \qquad , \qquad P(A/B)=\frac{P(A\cap B)}{P(B)}$$

מכאן: $P(A/B)\cdot P(B)=P(B/A)\cdot P(A)$ (נוסחת בייס)

ולכן: $P(\bar{A}/B)\cdot P(B)=P(B/\bar{A})\cdot P(\bar{A})$ (החלפנו את A במשלים שלו $\bar{A}$).

נחלק את שתי המשוואות זו בזו ונקבל:

$$\frac{P(A/B)}{P(\bar{A}/B)}=\frac{P(B/A)}{P(B/\bar{A})}\cdot\frac{P(A)}{P(\bar{A})}$$

כלומר:

$$R=\frac{P(A/B)}{P(\bar{A}/B)}=\frac{P(B/A)}{P(B/\bar{A})}\cdot\frac{P(A)}{P(\bar{A})}$$

השיעור הבסיסי של A

דיאגנוסטיות של B

היחס המעודכן R

כעת נפתח נוסחה שתאפשר לנו למצוא את היחס $P(A/B)$ כשנתון לנו היחס המעודכן R:

$$R=\frac{P(A/B)}{P(\bar{A}/B)}=\frac{P(A/B)}{1-P(A/B)}$$

$$R[1-P(A/B)]=P(A/B)$$

$$R=P(A/B)\cdot[1+R]$$

$$\boxed{P(A/B)=\frac{R}{1+R}}$$

<u>הערות:</u>

א. היות ולא פיתחנו נוסחא חדשה, אלא נעזרנו בכל הנוסחאות שלמדנו קודם, ניתן לפתור את רוב התרגילים בעזרת טבלה דו-מימדית ובלי R.
במקרה כזה ייתכן (אבל לא תמיד) שעל מנת לבנות את הטבלה הדו-מימדית נאלץ להשתמש בנעלמים ותהיה לנו עבודה יותר "שחורה".
ולפיכך, כשלא מבקשים מכם במפורש לפתור באמצעות טבלה, עדיף לפתור באמצעות R.

ב. בדרך-כלל נדרש לחשב את $P(A/B)$. (ראה שאלות אפשריות בהמשך).

ג. היתרון בנוסחא $\frac{P(A/B)}{P(\bar{A}/B)} = \frac{P(B/A)}{P(B/\bar{A})} \cdot \frac{P(A)}{P(\bar{A})}$ **שהפרופורציה $P(B)$ אינה מופיעה בה ולכן גם אם איננה ידועה אפשר לחשב את $P(A/B)$.**

<u>דוגמא מספרית להכרת המושגים:</u>

A ו- B הן קבוצות חלקיות של קבוצה כוללת S.

נתון: $P(A) = \frac{3}{5}$

$P(B/A) = \frac{2}{3}$

$P(B/\bar{A}) = \frac{3}{4}$

א. חשב את השיעור הבסיסי של A: -

$P(A) = \frac{3}{5}$ **(נתון)**

$$P(\bar{A}) = 1 - P(A) = 1 - \frac{3}{5} = \frac{2}{5}$$

השיעור הבסיסי של A $\frac{P(A)}{P(\bar{A})} = \frac{3/5}{2/5} = \frac{3}{5} \cdot \frac{5}{2} = \frac{3}{2}$

ב. חשב את הדיאגנוסטיות של B: -

$$\frac{P(B/A)}{P(B/\bar{A})} = \frac{2/3}{3/4} = \frac{2}{3} \cdot \frac{4}{3} = \frac{8}{9}$$

ג. חשב את היחס המעודכן R: -

$$R = \frac{P(B/A)}{P(B/\bar{A})} \cdot \frac{P(A)}{P(\bar{A})} = \frac{8}{9} \cdot \frac{3}{2} = \frac{4}{3}$$

ד. חשב את $P(A/B)$: -

$$P(A/B) = \frac{R}{1+R} = \frac{4/3}{1+4/3} = \frac{4}{3+4} = \frac{4}{7}$$

ה. חשב את $P(B)$ בעזרת נוסחת בייס: -

$$P(A/B) \cdot P(B) = P(B/A) \cdot P(A)$$

$$\frac{4}{7} \cdot P(B) = \frac{2}{3} \cdot \frac{3}{5}$$

$$\frac{4}{7} \cdot P(B) = \frac{2}{5}$$

$$P(B) = \frac{2}{5} \cdot \frac{7}{4} = \frac{7}{10}$$

ו. בנה טבלה דו-מימדית והשווה בין התשובות: -

$$P(B/A) = \frac{2}{3} = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{3/5}$$

$$P(B \cap A) = \frac{2}{3} \cdot \frac{3}{5} = \frac{2}{5}$$

$$P(B/\bar{A}) = \frac{3}{4} = \frac{P(B \cap \bar{A})}{P(\bar{A})} = \frac{P(B \cap \bar{A})}{2/5}$$

$$P(B \cap \bar{A}) = \frac{3}{4} \cdot \frac{2}{5} = \frac{3}{10}$$

	$\bar{A}$	A	
$7/10$	$3/10$	$2/5$	B
$3/10$	$1/10$	$1/5$	$\bar{B}$
1	$2/5$	$3/5$	

בהתאם לנתוני השאלה והחישובים הנ"ל נבנה את הטבלה (נתוני השאלה והחישובים מסומנים בטבלה ב"עיגול").

לפי תוצאות הטבלה: -

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{2/5}{7/10}=\frac{4}{7}$$

התוצאה יצאה שווה לתוצאה שחושבה ללא הטבלה (כצפוי).

<u>נחזור לשיפוט על-פי יציגות ולכשל הנקרא "כשל השיעור הבסיסי".</u>

נביא דוגמא להמחשה בלבד:

בבית ספר תיכון 20% מהתלמידים לומדים מתמטיקה ברמה של 5 יח"ל והשאר ברמות אחרות.
בבדיקת תוצאות בחינת הבגרות האחרונה במתמטיקה הסתבר כי 88% מתלמידי
5 יח"ל הצליחו בבחינה בעוד שרק 40% מתלמידי הרמות האחרות הצליחו בבחינה.
כתב של עיתון ביה"ס בחר לראיין אחד מהתלמידים שהצליח בבחינה. מה ההסתברות שהתלמיד שרואיין לומד ברמה של 5 יח"ל?

במבט ראשון, נראה כי מכיוון שאומרים לנו שהתלמיד המרואיין הצליח בבחינה, יש סיכוי של 88% שהוא לומד ברמה של 5 יח"ל. <u>רובנו</u> נתייחס לתכונה "הצליח במבחן" כמתאימה יותר, ומייצגת יותר תלמיד ברמה של 5 יח"ל <u>ונטעה</u> בתשובה. אנו משתמשים כאן <u>במנגנון היציגות</u> <u>ומתעלמים מנתון סטטיסטי חשוב</u> שהוא: רק 20% מכלל התלמידים לומדים ברמה של 5 יח"ל.
כלומר, רק חלק קטן יחסית מהתלמידים לומדים ברמה של 5 יח"ל וזה מוריד את הסיכוי שהתלמיד המרואיין יהיה ברמה זו.
הנתון ממנו התעלמנו הוא <u>השיעור הבסיסי</u> – היחס בין מספר תלמידי 5 יח"ל (קבוצה A) לבין מספר התלמידים ברמות האחרות (קבוצה $\bar{A}$).
ההתעלמות מהשיעור הבסיסי הובילה אותנו לטעות בהערכת ההסתברות ולכן קוראים לטעות מעין זו <u>"כשל השיעור הבסיסי".</u>

נפתור את אותה שאלה בדרך סטטיסטית נכונה, תוך התייחסות לשיעור הבסיסי.

<u>נסמן</u>: A - קבוצת התלמידים הלומדים ברמה של 5 יח"ל.
B - קבוצת התלמידים שהצליחו בבגרות במתמטיקה.

<u>נסמן גם את המשלימים שלהם</u>:
$\bar{A}$ - קבוצת התלמידים הלומדים ברמות אחרות.
$\bar{B}$ - קבוצת התלמידים שלא הצליחו בבגרות במתמטיקה.

<u>נתון</u>:

$$P(A) = 0.20$$
$$P(B/A) = 0.88$$
$$P(B/\bar{A}) = 0.40$$

בשאלה אנו נדרשים בעצם למצוא את $P(A/B)$ (ידוע שהתלמיד הצליח בבגרות).

הנוסחה אומרת:

$$\boxed{P(A/B) = \frac{R}{1+R}}$$

אנו יודעים כי:

$$R = \frac{P(A/B)}{P(\bar{A}/B)} = \frac{P(B/A)}{P(B/\bar{A})} \cdot \frac{P(A)}{P(\bar{A})}$$

השיעור הבסיסי של A

דיאגנוסטיות של B

היחס המעודכן R

נציב את הנתונים:

$$R = \frac{0.88}{0.40} \cdot \frac{0.20}{0.80} = 2.20 \cdot 0.25 = 0.55$$

ומכאן:

$$P(A/B) = \frac{0.55}{1+0.55} \cong 0.35$$

כלומר, התשובה הנכונה היא שההסתברות שהתלמיד המרואיין לומד ברמה של 5 יח"ל היא 35% בערך ולא 88%.

<u>לסיכום:</u>
כשל השיעור הבסיסי קורה כאשר השיפוט שאנו עושים מתבסס על מנגנון היציגות ולכן אנו נוטים לייחס הסתברות לתכונה המתוארת תוך התעלמות מהשיעור הבסיסי של אותה תכונה. ההתעלמות מהשיעור הבסיסי גורמת לטעות בתשובה.
התעלמות מהשיעור הבסיסי מאפיינת "נשאל נאיבי" – כלומר כל אדם שאינו מצוי ברזי החשיבה ההסתברותית...

שאלות אפשריות בשיפוט על-פי יציגות

1. נותנים לנו אינפורמציה מקדימה על אדם ואתם נשאלים איזה משפט מייצג יותר את האדם. בדרך כלל תקבלו משפט עם אפשרות אחת ומשפט הכולל שתי אפשרויות. זוהי שאלה שמאפיינת שאלה הנוגעת ל"כשל הצירוף".

   אם תתבקשו להסביר את הטעות, אחד הניסוחים האפשריים של ההסבר יכול להיות:

   בשיפוט על-פי יציגות נוטים לתת הסתברות גבוהה יותר למשפט הכולל שתי אפשרויות מאשר למשפט הכולל אפשרות אחת וזה בגלל ששתי האפשרויות נתפסות כמייצגות יותר. זה נוגד את כלל הצירוף.

2. מבקשים למצוא מנתוני השאלה בעזרת חישוב את השיעור הבסיסי, הדיאגנוסטיות והיחס המעודכן ודרכם את $P(A/B)$.

   אחד הסעיפים של שאלה מסוג זה יכול לבקש תשובה לשאלה:
   כיצד היו עונים על השאלה "נשאלים נאיביים", "אנשים שלא למדו חשיבה הסתברותית" או "רוב האנשים" וכיצד קוראים לתופעה זו ו/או מאיזה נתון מתעלמים אנשים אלה?
   זוהי שאלה הנוגעת ל"כשל השיעור הבסיסי".

   אחד הניסוחים האפשריים של תשובה לשאלה מסוג זה יכול להיות:

   "נשאלים נאיביים / אנשים שלא למדו.../רוב האנשים.. מתעלמים מהשיעור הבסיסי (כי הם שופטים לפי מנגנון היציגות) ולכן מקבלים תשובה מוטעית. לתופעה זו קוראים "כשל השיעור הבסיסי"....

כאשר בשאלה השיעור הבסיסי <u>אינו נתון כלל</u>, ואתם נשאלים "מה הטעות האפשרית?" ניסוח התשובה יכול להיות: "מכיוון שהשיעור הבסיסי אינו נתון סביר להניח שמדובר בטעות מסוג "כשל השיעור הבסיסי".

> **<u>שימו לב!</u>**
> מאד חשוב להשתמש בתשובה במילים הרלוונטיות - "השיעור הבסיסי" ,
> כאשר מתעלמים מהשיעור הבסיסי מתקבלת תשובה מוטעית וזהו "כשל השיעור הבסיסי".

3. שאלות "הפוכות":
   כאשר נתונים $P(A/B)$ והדיאגנוסטיות ויש למצוא את $P(A)$.
   הדרך לפתרון תהיה:
   ידוע ש:- $$P(A/B)=\frac{R}{1+R}$$
   מנוסחה זו ניתן למצוא את R ולאחר שמצאנו את R ומהנתון של הדיאגנוסטיות ניתן למצוא את השיעור הבסיסי שהוא: $$\frac{P(A)}{P(\bar{A})}=\frac{P(A)}{1-P(A)}$$
   מכאן נמצא את $P(A)$ (ואת $P(\bar{A})$, אם יש צורך בכך).

4. שאלות בהן נתונים הדיאגנוסטיות והשיעור הבסיסי ויש למצוא את $P(A/B)$.

שימו לב! את הנתון של הדיאגנוסטיות אנו מסיקים מדרך הניסוח למשל:
"פרופורציית "התכונה" מבין "הקבוצה" גדולה פי x , או קטנה פי y מפרופורציית "התכונה" ב"קבוצה המשלימה".

הדיאגנוסטיות תהיה בעצם היחס הנתון : x או $\frac{1}{y}$ בהתאמה.

מהנתונים ניתן למצוא את R ולאחר מכן את $P(A/B)$.

5. שאלות ניבוי:

בשאלה מובאת אינפורמציה לגבי יכולתם של אדם, חיה, מכונה, מבחן וכד' לצפות מראש (לנבא) התרחשות של אירוע מסויים.

דוגמאות:
מבחני כניסה שמנבאים הצלחה בלימודים / קורס וכד'...
כלבים מאומנים שמסתובבים בשדות תעופה ושמנבאים מי נושא סמים...
גרפולוגים שמנבאים אמינות של מועמדים... / הצלחה בתפקיד...
רופאים שמנבאים הצלחה בניתוחים...
בדיקות שמנבאות סיכוי לחלות במחלה כלשהי...

שימו לב! לצורך ההסבר על הסוגים השונים של שאלות הניבוי נשתמש בדוגמא של "נביא" שמנבא סיום קורס (באותה מידה יכולנו לבחור בכלב שמנבא למי יש סמים וכו').

סוגי שאלות ניבוי:

א. יש שאלות בהן נתונים אחוזי ההצלחה והכשלון של ה"נביא" לנבא אירוע מסויים, למשל:
נתון כי 95% מהנבחנים שסיימו את הקורס הנביא אכן אמר שהם יסיימו. (כלומר הוא מצליח בניבוי ב-95% מהמקרים).
כמו כן נתון כי 10% מהמועמדים שלא סיימו את הקורס הנביא אמר עליהם שהם כן יסיימו אותו. (כלומר הוא נכשל בניבוי ב- 10% מהמקרים).

נגדיר:
A - קבוצת מסיימי הקורס.
B - הקבוצה ש"הנביא" אמר שיסיימו את הקורס.

ולכן הנתונים הם:

$$P(B/A) = 0.95$$
$$P(B/\overline{A}) = 0.10$$

ב. יש שאלות שיגידו לנו ש"הנביא" צודק ב- 85% מהמקרים - זה אומר:
85% מהמועמדים שסיימו את הקורס הנביא אכן אמר שהם יסיימו ובאותה מידה -
85% מהמועמדים שלא סיימו את הקורס הנביא אכן אמר שלא יסיימו את הקורס.

ולכן הנתונים הם:

$$P(B/A)=0.85$$
$$P(\overline{B}/\overline{A})=0.85$$

בשאלות מסוג א' ו- ב' מופיע לעיתים קרובות סעיף שמניסוחו ניתן להבין כי המועמד הוא מהקבוצה שהנביא ניבא שיסיים את הקורס (B) ושואלים מהי ההסתברות שהמועמד <u>אכן</u> סיים את הקורס (A).

בשאלה כזו בעצם אתם מתבקשים למצוא את $P(A/B)$:

P(קבוצת מסיימי הקורס A / הקבוצה שהנביא ניבא שיסיימו את הקורס B)

<u>הערה כללית</u>

לפעמים מדרך ניסוח השאלה ניתן להסיק באיזה סוג שאלה מדובר:

כאשר בשאלה מצויין "אין צורך בחישובים" ניתן להסיק שהכוונה היא ל"נשאל נאיבי" שלא למד הסתברות והתשובה נמצאת בעצם בנתונים.

כמו כן יש רמז בשאלה על טעות אפשרית, דהיינו אחד מה"כשלים" האפשריים.

אנו יודעים שנשאל נאיבי יכול להכשל או ב"כשל הצירוף" או ב"כשל השיעור הבסיסי".

אם בשאלה מובאים תיאורים ומשווים בין סיכוי של אפשרות אחת לעומת צירוף אפשרויות – מדובר ב"כשל הצירוף".

אם בשאלה מובאות פרופורציות של קבוצות שונות סביר להניח שמדובר בשאלה הקשורה ב"כשל השיעור הבסיסי".

כעת אתם יכולים וצריכים לדרוש מעצמכם יותר

בהצלחה!!!