מדעי הנתונים להנדסה

פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

1. סולמות מדידה 1

סטטיסטיקה תיאורית – סיווג משתנים וסולמות מדידה:

רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתח אותם.

בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת. באותה קבוצה אנו אוספים נתונים על הישויות באותה קבוצה.

משתנה – תכונה שיכולה לקבל מספר ערכים : דעה פוליטית, מקום מגורים, גובה של אדם וכדומה.

כל ישות בקבוצה שאנו צופים בה ואוספים לגביה נתונים נקראת תצפית.

הנתונים שנאספים בדרך כלל מרוכזים במסד נתונים. במסד הנתונים כל שורה היא תצפית וכל עמודה מייצגת משתנה.

דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכביש החוף.
אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה :

מין	גיל	מצב הפצוע	דופק
גבר	26.6	אנוש	40
גבר	24.5	קשה	38
אישה	32.1	קשה	50
גבר	34.9	בינוני	65
אישה	23.1	קל	89

ענו על השאלות הבאות :
הגדירו את הקבוצה שבדוגמה.
כמה תצפיות בקבוצה?
כמה משתנים בקבוצה?
כמה ערכים יש למשתנה "מין"?

© כל הזכויות שמורות לברק קנדל

את המשתנים במחקר אנו מסווגים ל-"סולמות מדידה" הדבר חשוב לדרך שבה ננתח את הנתונים בהמשך.

מיון משתנים לפי סולמות המדידה:

1. סולם שמי (nominal) – משתנה שלערכיו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות לערכים שלהם לדוגמה: צבע מועדף.

   משתנה דיכוטומי (הינו מסולם שמי) אותם משתנים שיש להם רק שני ערכים אפשריים. למשל: אזרחות ישראלית: יש או אין.

2. סולם סדר (ordinal) – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר מי יותר או מי פחות אבל אין משמעות לגודל.
   משתנה מסולם סדר יכול לקבל ערכים מילוליים או מספריים.
   למשל, דרגה בצבא.

3. סולם כמותי (scale) – משתנה שחייב להיות מספרי, לערכים שלו בנוסף לשם ולסדר בניהם יש משמעות לערך המספרי. משתנה כמותי הוא משתנה שניתן בדרך כלל לספור או למדוד על ידי מכשיר מדידה. למשל, מספר המחשבים בדירה, שטח הדירה במ"ר.

את המשתנה הכמותי אנו מסווגים לשני סוגים:

משתנה בדיד:
משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים כמו: מספר המחשבים בדירה.

משתנה רציף:
משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף וללא קפיצות של ערכים. למשל, שטח הדירה במ"ר.

דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכביש החוף.
אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה:

מין	גיל	מצב הפצוע	דופק
גבר	26.6	אנוש	40
גבר	24.5	קשה	38
אישה	32.1	קשה	50
גבר	34.9	בינוני	65
אישה	23.1	קל	89

סווגו כל משתנה במסד הנתונים: שמי, סדר, כמותי רציף, כמותי בדיד.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) לפניכם טבלה המסכמת נתונים לגבי סקר שנעשה היום:

שם משפחה	מצב משפחתי	מידת דתיות	מספר רכבים
כהן	רווק	חילוני	0
חדד	נשוי	חילוני	1
לביא	גרוש	מסורתי	1
פיינגולד	אלמן	חילוני	2
אבו שוקרא	נשוי	דתי	1
בן חיים	נשוי	מסורתי	0
רוטשילד	רווק	חילוני	0

א. כמה תצפיות בדוגמה זו?

ב. כמה משתנים בדוגמה זו?

ג. כמה ערכים ישנם ל-"מידת דתיות"?

ד. מהם הערכים האפשריים למשתנה "מצב משפחתי"?

2) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן: 50 אנשים נכנסו 0 פעמים לסניף, 20 אנשים נכנסו פעם אחת לסניף, 5 אנשים נכנסו פעמיים לסניף, 5 אנשים נכנסו יותר מפעמיים.

א. הגדירו את הקבוצה בדוגמה זו.

ב. כמה תצפיות בדוגמה זו?

ג. הגדירו את המשתנה בדוגמה זו. מהו סולם המדידה שלו?

3) במחקר רפואי התעניינו לדעת כיצד מינון תרופת "קופקס" משפיע על מספר שעות השינה של אדם. במחקר השתתף אדם אחד בשם דני שבכל יום ניתן לו מינון שונה של התרופה. הטבלה שלהלן מתארת בכל יום את מינון התרופה במ"ג שקיבל האדם וכמו כן את מספר שעות השינה שלו באותו הלילה:

מספר היום	מינון התרופה	מספר שעות שינה
1	12	6
2	14	7
3	16	7.5
4	18	6.5
5	20	8

א. כמה תצפיות נאספו במחקר?

ב. סווגו את סולם המדידה של "מינון התרופה"?

© כל הזכויות שמורות לברק קנדל

4) לפניכם רשימה של משתנים.
ציינו באיזה סולם מדידה מדובר (שמי, סדר, כמותי בדיד, כמותי רציף):

א. גובה אדם בס"מ.

ב. מספר ילדים למשפחה.

ג. מידת חרדה לפני מבחן.

ד. שביעות רצון משירות לקוחות בסקלה מ-1 עד 7 (1 כלל לא מרוצה עד 7 מרוצה מאד).

ה. השכלה.

ו. מספר אוטובוס.

ז. מקום מגורים.

ח. מין (1=גבר ו-2=אישה).

ט. מידת נעליים.

5) לפניכם רשימה של משתנים כמותיים.
ציין ליד כל משתנה אם הוא רציף או בדיד:

א. שכר עובד ב-₪.

ב. ציון בחינת בגרות.

ג. תוצאה בהטלת קובייה.

ד. מהירות ריצה במטר לשנייה בתחרות 100 מטר.

ה. שיעור התמיכה בממשלה בעיר.

6) גברת לוי החליטה לדגום 25 ימים של נסיעה לעבודה, כאשר בדרך לעבודתה יש 3 צמתים מרומזרים.
ב-9 ימים הגיעה גברת לוי לעבודה מבלי לעצור באף צומת.
ב-9 ימים נוספים היא הצליחה לעבור בשני רמזורים ירוקים.
ב-5 ימים נוספים היא הצליח לעבור רק בירוק אחד.
בשאר הימים, היא לא עברה באף רמזור ירוק.
מעוניינים לחקור את מספר הרמזורים האדומים בהם עצרה גברת לוי.

א. מהו המשתנה הנחקר בדוגמה זו?

ב. מהם הערכים האפשריים של משתנה זה?

ג. כמה ערכים אפשריים יש למשתנה?

ד. מהו סולם המדידה של המשתנה?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. $n=7$.
ב. 4.
ג. 3.
ד. רווק, נשוי, גרוש, אלמן.

2) א. לקוחות סניף 543 של בנק "רווה".
ב. $n=80$.
ג. X = מספר הפעמים בשבוע שלקוח נכנס לסניף. כמותי בדיד.

3) א. $n=5$.
ב. כמותי רציף.

4) א. כמותי רציף.
ב. כמותי בדיד.
ג. אין מספיק נתונים.
ד. סדר.
ה. אין מספיק נתונים.
ו. שמי.
ז. שמי.
ח. שמי.
ט. סדר.

5) א. רציף.
ב. בדיד.
ג. בדיד.
ד. רציף.
ה. רציף.

6) א. מספר הרמזורים בהם עוצרת גברת לוי ביום בדרך לעבודה.
ב. 0, 1, 2, 3.
ג. 4.
ד. כמותי בדיד.

© כל הזכויות שמורות לברק קנדל