# שיטות מחקר כמותניות

## פרק 3 - דגימה

### תוכן העניינים

1. דגימה ..... 1

# דגימה:

## מושגי יסוד בדגימה:

### רקע:

השערות המחקר נוגעות תמיד לאוכלוסייה, אך מאילוצים שונים נאלץ לבצע את המחקר על חלק קטן ממנה – מדגם ונקיש מתוצאות המדגם על האוכלוסייה (היקש הסתברותי).
ככל שהמדגם ייצג / ישקף טוב יותר את האוכלוסייה (לגבי משתנים מסוימים ובפרט אלה הנבדקים) כך טעות הדגימה בהיסק מהמדגם לאוכלוסייה תהיה קטנה יותר.

את מה דוגמים?
אנו דוגמים את יחידות הניתוח מתוך אוכלוסיית המחקר.
יחידות הניתוח יכולות להיות, למשל, אנשים, כתבות, פרסומות, תוכניות, טלוויזיה, ארגונים, בתי ספר.

הגדרת האוכלוסייה:
לפני כל דגימה יש להגדיר את אוכלוסיית המחקר.
הגדרת האוכלוסייה צריכה להיות מנוסחת במונחים אופרציונליים, כך שיהיה ברור מי נכלל בה ומי לא.
הגדרת האוכלוסייה תכלול את יחידת הניתוח, היקף האוכלוסייה ואת הזמן ו/או המקום שבו נערך המחקר.

קביעת גודל המדגם:

1. הומוגניות / הטרוגניות האוכלוסייה – כאשר אוכלוסייה הומוגנית במשתנה הנחקר, אין חשיבות לגודל המדגם או לצורת הדגימה. מידת ההטרוגניות (שונות) קובעת באיזו מידה ניתן להסיק מן התוצאה במדגם על המצב באוכלוסייה.

2. טעות הדגימה שהחוקר קבע – קיים קשר הפוך בין טעות הדגימה לגודל המדגם. ככל שהחוקר יקבע טעות דגימה קטנה יותר כך יצטרך לדגום מדגם גדול יותר (אם ידגום את כלל האוכלוסייה לא תהיה טעות דגימה כלל).

**שיטות דגימה:**

שיטות הדגימה נחלקות לשני סוגים:

1. הסתברותית.
2. לא הסתברותית.

**דגימה הסתברותית:**

מתרחשת אם מתקיימים 3 תנאים:

א. לכל הפריטים באוכלוסייה יש הסתברות ידועה להיבחר למדגם.
ב. לשום פריט אין סיכוי ודאי להיבחר.
ג. שום פריט אינו יוצא מכלל אפשרות להיבחר.

שיטות דגימה הסתברותיות מאפשרות הסקה הסתברותית מהמדגם לאוכלוסייה, תוך חישוב טעות הדגימה.

על מנת לבצע דגימה הסתברותית חייבת להיות ברשות החוקר מסגרת דגימה.

מסגרת דגימה כוללת את כל הפריטים שמתוכם נלקח המדגם הלכה למעשה. היא תהיה קטנה מהאוכלוסייה הנחקרת, ובכל אופן לא יכולה להיות גדולה ממנה. ישנו היסק סטטיסטי – מהמדגם למסגרת הדגימה והיסק אינטואיטיבי – ממסגרת הדגימה לאוכלוסייה הנחקרת. דיוקם של ההיסקים תלויים במידת הייצוג של המדגם את מסגרת הדגימה ושל מסגרת הדגימה את האוכלוסייה.

**שיטות דגימה הסתברותיות**

דגימה הסתברותית היא דגימה בה לכל פריט יש הסתברות ידועה להיכלל במדגם. ההסתברויות לכל פריט לא חייבות להיות שוות, אך עליהן להיות ידועות, קטנות מאחד וגדולות מ-0, כדי שההסתברות השונה תוכל להיכלל בחישוב.

א. דגימה מקרית פשוטה.
ב. דגימה שיטתית.
ג. דגימת שכבות.
ד. דגימת אשכולות.

דגימה מקרית פשוטה – היא דגימה בה לכל פריט/אובייקט מתוך האוכלוסייה קיים סיכוי שווה להיבחר למדגם. שיטה זו נעזרת בלוחות מספרים אקראיים או בתוכנת מחשב מתאימה הפולטת מספרים באופן מקרי. שיטה זו מקובלת בעיקר באוכלוסיות קטנות יחסית, כאשר מסגרת הדגימה ממוספרת וקל לאתר את

הפריטים השונים לפי מספריהם.

דגימה שיטתית – משמשת בעיקר במצבים בהם מסגרת הדגימה או המדגם גדולים מאוד או כאשר מסגרת הדגימה איננה ממוספרת (כמו למשל, מסגרת דגימה המבוססת על ספר טלפונים) ולכן לא ניתן להשתמש בלוח המספרים המקריים. צריך רשימה מסודרת של מסגרת הדגימה. רק הפריט הראשון נדגם באופן מקרי. שאר הפריטים נדגמים בצורה שיטתית לפי יחס הדגימה – היחס בין גודל מסגרת הדגימה לגודל המדגם הרצוי.

הבעיות בשיטת הדגימה השיטתית:

1. כאשר מסגרת הדגימה זוהי רשימה ממוינת / מסודרת לפי סדר מסוים שרלוונטי לתכונה הנמדדת במחקר.
2. כאשר הרשימה שהתקבלה היא רשימה מחזורית ויחס הדגימה חופף ליחס המחזוריות או לכפולות שלו. במקרה זה יידגם כל פעם אותו פריט וכל שאר הפריטים בתוך המחזור יפוספסו.

דגימת שכבות – כאשר קיימת ידיעה באשר לקיומם של משתני רקע (כמו למשל, גיל, מוצא, מגדר, רמת השכלה וכו') שהם רלוונטיים לתכונה הנמדדת במחקר, אזי רצוי קודם כל לחלק את האוכ' לשכבות לפי כל משתנה רקע רלוונטי בנפרד.

כל שכבה היא הומוגנית בתוכה בתכונה הנחקרת אך הטרוגנית ביחס לשכבות אחרות.
חסרון השיטה – לא ניתן להשתמש בה לשיטות מורכבות לניתוח סטטיסטי.

כמה נבדקים צריך לדגום מכל שכבה?
כדי לדעת כמה נבדקים צריך לדגום מכל שכבה, החוקר צריך לדעת את ההתפלגות (באחוזים) של ערכי המשתנה באוכלוסייה (למשל, התפלגות הנשים והגברים באוכ' הנחקרת).
כלומר, צריך לדעת האם לבצע בדיקת שכבות פרופורציונלית או לא פרופורציונלית.

**דגימת שכבות פרופורציונלית** – אם קיימים נתונים לגבי ההתפלגות (באחוזים) של ערכי המשתנה באוכלוסייה – החוקר יחשב את מספר הנבדקים שהוא צריך לדגום מכל שכבה (על פי האחוזים וגודל המדגם הרצוי). הנבדקים ידגמו מכל שכבה בדגימה הסתברותית פשוטה.

**דגימת שכבות לא פרופורציונלית** – כאשר לא קיימים נתונים לגבי ההתפלגות של ערכי המשתנה באוכ' אז פשוט נדגום מכל שכבה מספר נבדקים שווה.
ניתן גם להגדיל / להקטין באופן מלאכותי דגימה של שכבות מסוימות מתוך האוכלוסייה לפי צרכי המחקר.

דגימת אשכולות – משמשת כאשר מסגרת הדגימה גדולה במיוחד ומפוזרת גיאוגרפית או במקרים שאין רשימה אחת מסודרת. כמו למשל במחקרים על בתי ספר, בתי חולים או ארגונים בעלי סניפים מרושתים באזורים שונים.
בשיטה זו מחלקים את האוכלוסייה לתתי-קבוצות (אשכולות) – דוגמים מקרית מספר אשכולות וכוללים במדגם את כל הפריטים המשתייכים לאשכולות שנדגמו.
ישנה מקריות בבחירת האשכולות עצמם.

ישנה גם דגימת אשכולות רב שלבית – הליך דגימה הסתברותי בו מבצעים מספר שלבים של דגימת אשכולות.
הבעיה בדגימת אשכולות בכלל וקל וחומר בדגימת אשכולות רב שלבית היא הגדלת טעות הדגימה בשל הדגימה הרב שלבית-של אשכול אח"כ דגימה בתוך האשכול ואח"כ דגימה של פרטים.
היתרון – חיסכון במשאבים וזמינות של רשימות דגימה מצומצמות יותר, נוחות יותר לשימוש וקלות יותר להשגה.

## ההבדלים בין דגימת שכבות לדגימת אשכולות:

לשתי השיטות שני שלבים: בשלב הראשון – מחלקים את האוכלוסייה לקבוצות (שכבות או אשכולות) ובשלב השני – דוגמים מכל קבוצה (או את הקבוצות עצמן – במקרה של דגימת אשכולות).

לשתי השיטות מרכיב של בחירה מקרית. בדגימת שכבות הבחירה המקרית היא מתוך השכבות ואילו בדגימת אשכולות הבחירה המקרית היא בבחירת האשכולות עצמן.

| | דגימה מקרית | טעות הדגימה | כיצד נקטין את טעות הדגימה |
|---|---|---|---|
| **שכבות** | מתוך כל שכבה | שונות בתוך השכבות | הומוגניות בתוך כל שכבה<br>הטרוגניות בין השכבות |
| **אשכולות** | של אשכולות | שונות בין אשכולות | הומוגניות בין האשכולות<br>הטרוגניות בתוך כל אשכול |

המשמעות: כאשר חוקרים משתנה שיש סבירות גבוהה לשונות גדולה יותר בתוך כל אחד מהערכים של המשתנה מאשר בין ערכי המשתנה, נבחר בדגימת אשכולות.
לעומת זאת כאשר חוקרים משתנה שקיימת סבירות גבוהה לשונות גדולה יותר בין הערכים השונים מאשר בתוך כל אחד מהערכים, נבחר בדגימת שכבות.

שיטות דגימה רב שלביות

ניתן לשלב את שיטות הדגימה השונות.

למשל:
לצורך סקר על איכות חיים בתל אביב חוקר מחלק שכונות לשכבות על-בסיס סוציו-אקונומי, ואז בכל שכבה מתייחס לשכונות כאשכולות שמתוכם הוא דוגם בצורה מקרית או שיטתית את האנשים שישתתפו בסקר.

**דגימה לא הסתברותית:**

לפחות אחד מ-3 התנאים הנ"ל אינו מתקיים (1> P ידועה>0).
או שהחוקר מבחין בבעיות העלולות להתעורר בכל אחד משלבי איסוף הנתונים:

אין מסגרת דגימה.

קושי באיתור האנשים (אוכלוסיות סגורות, פעילות לא חוקית).

קושי בשיתוף הפעולה של הנחקרים (נשים מוכות, למשל).

דגימה לא הסתברותית איננה תקפה במחקר מדעי שכן לא ניתן להסיק הסתברותית מהמדגם לאוכלוסייה (לא ניתן לחשב את טעות הדגימה).
אולם ישנם מקרים בהם דגימה הסתברותית אינה אפשרית ולכן קיימות גם שיטות דגימה לא הסתברותיות.

שיטות דגימה לא הסתברותיות:

מדגם נוחות – בחירת הנחקרים הקלים ביותר להשגה על ידי עורך המחקר.
מדגם כזה בד"כ לא מייצג ומקובל במחקרי גישוש.

מדגם כדור השלג – המדגם נבחר בעזרת קבוצה קטנה של מודיעים שמהם מקבלים שמות וכתובות של אנשים נוספים השייכים לקבוצה הנחקרת. משמש כאשר יש קושי להגיע לקבוצה הנחקרת.

מדגם מתנדבים – המדגם מורכב ממתנדבים למחקר. משמש כאשר יש בעיות בנוגע לשיתוף פעולה של המשתתפים (כאשר המחקר מצריך זמן רב למשל).

שיטת המכסה – החוקר בוחר את מי יראיין כשהמגבלה היחידה המחייבת אותו היא מספר המרואיינים מאותה אוכלוסייה. משמש כאשר אין רישום מרכזי של האוכלוסייה כולה או כדי לחסוך בזמן ובמשאבים.

## תרגול:

**1)** חוקר רצה לבחון את השימוש בשירותי בנק אוטומטיים לעומת פנייה לפקיד באשנב, בקרב בנים ובנות ובקרב קבוצות גיל שונות.
הנח כי רשימת מסגרת הדגימה שבידיו (לקוחות רשומים של בנק גדול בארץ) כוללת 10,000 איש, המתפלגים מבחינת קבוצות הגיל כדלהלן:
עד 18: 20%.
18-65: 40%.
מעל 65: 40%.
בנוסף, אחוז הנשים במסגרת הדגימה: 40%. כמו כן אין קשר בין קבוצת הגיל למין הנבדק. החוקר מעוניין לדגום 200 איש מתוך מסגרת הדגימה הנתונה.

א. החוקר מאמין כי הן גיל והן מגדר הם גורמים מרכזיים בקביעת השימוש בשירותי בנק אוטומטיים.
כיצד ידגום את הנבדקים בהתחשב במידע זה?
נמק תוך הצגת הרכב המדגם המתאים.

ב. החוקר מאמין כי רק הגיל הוא גורם מרכזי בקביעת השימוש בשירותי בנק אוטומטיים. כיצד יבצע את הדגימה בהתחשבות במידע זה?
נמק תוך הצגת הרכב המדגם המתאים.

ג. מה מקור טעות הדגימה בכל אחת מהדגימות בסעיפים א' ו-ב'.

**2)** חוקר רוצה ללמוד מהם חמשת משחקי המחשב האהודים ביותר על ידי ילדים ישראלים מכיתות א', באמצעות שאלון. הוא הצליח לקבל ממשרד החינוך רשימה עם שמותיהם של 25,000 תלמידי כיתות א' מ-450 בתי ספר יסודיים ברחבי הארץ.

א. באיזו שיטת דגימה כדאי לו לבחור? נמק והסבר.

ב. מה מקור טעות הדגימה בשיטה שנבחרה?

**3)** לגבי כל אחד מהתיאורים הבאים רשמו מהי שיטת הדגימה המתאימה ביותר, מהן יתרונותיה והאם צריך ידע מוקדם לשיטת דגימה זו:

א. חוקר רוצה לבדוק באילו סוגי עבודות נוטים להשתלב נערים ונערות שנשרו ממסגרות לימודיות. הוא מחליט לבחור מדגם של 400 מבין 6,000 בני הנוער שנשרו בשנה הקודמת ממסגרות לימודיות. החוקר מאמין שהגיל בו נושרים מביה"ס הוא הגורם המרכזי בבחירת סוג העבודה. ולכן מעוניין להבטיח שבמדגם יהיה ייצוג הולם לבני הגילאים השונים. החוקר יודע את אחוז כל גיל ביחס לאוכלוסייה הכללית.

ב. חוקר רוצה לבצע סקר על הרגלי הקריאה של האוכלוסייה. ברשות החוקר אין רישום של כל מנויי הספריות העירוניות בארץ, אך יש רישום של כל הספריות.

ג. במחקר על עיירות פיתוח החוקרים נדרשו לכלול 70% תושבי עיירות פיתוח, ו-30% תושבי ערים אחרות.

ד. חוקר מעוניין לבחון מניעים של אנשים הנוהגים לגנוב מחנויות.

**4)** חוקר רצה לבדוק את שביעות הרצון של אנשים המשתמשים בתחבורה הציבורית. הוא הגיע לתחנת אוטובוס ותחנת רכבת, דגם קרון אחד ברכבת ואוטובוס אחד בתחנת אוטובוסים. בחלק הקדמי בחר מקרית 3 אנשים, ובחלק האחורי בחר מקרית 3 אנשים.

א. מהי יחידת הדגימה / ניתוח במחקר זה?

ב. באיזה שיטה/ות דגימה השתמש?

## תשובות סופיות:

**1)** א. דגימת שכבות פרופורציונאלית לפי משתנים: גיל ומגדר.
ב. דגימת שכבות פרופורציונאלית לפי משתנה: גיל.
ג. סעיף א': בתוך קבוצת גיל ומגדר, סעיף ב': בתוך קבוצת גיל.

**2)** א. דגימת אשכולות.
ב. בין האשכולות.

**3)** א. שכבות פרופורציונאלית, יתרונות: ייצוג למשתנה רקע רלוונטי, נדרש: כן.
ב. אשכולות, יתרונות: ביצוע דגימה הסתברותית, נדרש: כן.
ג. שכבות לא פרופורציונאלית, יתרונות: ייצוג לשכבת מיעוט, נדרש: מסגרת דגימה.
ד. כדור שלג, יתרונות: חקירת נושא רגיש עם קושי באיתור אוכלוסייה רלוונטית, נדרש: "מודיעים".

**4)** א. אנשים המשתמשים בתחבורה ציבורית.
ב. דגימת שכבות (סוג תחבורה), דגימת אשכולות (קרון ואוטובוס), דגימת שכבות (מקום ישיבה), מקרית פשוטה (נוסעים).