

## סטטיסטיקה לכריית מידע (תשפ"ו)

### פרויקט - משימה 1

#### בניית בסיס נתונים

1. בשלב הראשוני יש למצוא (ולמתקשים - אולי לפחות להציע) "אוכלוסייה" שתיחקר סטטיסטית. בת לפחות כ- 36 אובייקטים (ואולי אם אפשר, אז עדיף 101 או יותר). בכפולות של לפחות 4 (יוסבר בשיעור), ממקורות שונים.

בסופו של דבר האוכלוסייה תיבחן לפי משתנים שונים, ובהתאם לאופייה.

דוגמאות ל"אוכלוסיות":

שקית - קרקרים, בייגלה, ביסלי, במבה, צ'יטוס וכו'.  
ק"ג (או יותר) - ענבים, שזיפים, משמשים (וישנה אפשרות לבחון רק את החרצנים שלהם), תפוחים וכו'.  
ק"ג (או יותר) - אגוזי מלך (קלופים או שלמים), משמש/תאנה מיובשת.  
כדורי משחק, חלקי משחק - אחידים ושונים בו זמנית..  
שקית - סוכריות גומי, סוכריות קשות.

הנכם/ מוזמנים להציע - "אוכלוסיות" מחקר דומות..  
אנא, התייעצו והציעו לי את המקור לבניית בסיס הנתונים, וקבלו את אישורי.

2. האוכלוסייה תיבחן ותימדד לפי פרמטרים ומשתנים שונים, וכדוגמת: משקל, גובה, אורך, רוחב, צבע (או צבע העטיפה), סימטריות, יופי, ואולי (וכדוגמה) בסיווג משתנה נומינלי - מושלם/כמעט מושלם, טוב, כמעט טוב, לא טוב - ועם פונקציית התאמה וקידוד מתאימה.

3. הכנסת המידע והתצפיות לקובץ אקסל.

#### כמה ניתוחים סטטיסטיים ראשוניים

4. ביצוע של ניתוחים סטטיסטיים ראשוניים (והיכן שרלבנטי) - ממוצע, שונות, סטיית תקן, חציון, שכית, טווח, אמצע הטווח, אחוז השגיאות, פרופורציות רלבנטיות. ממוצע הסטיות המוחלטות.

5. יצירת מחלקות ונתונים מקובצים בלוח שכיחויות מתאים, גבולות מדומים ואמתיים. שכיחות מצטברת, שכיחות יחסית, שכיחות יחסית מצטברת, רוחב המחלקה, חציון והטווח הבין רבעוני (מאונים מתאימים), המאון ה- 0.1, והמאון ה- 0.9.

6. הצגה גרפית: גרפי - מקלות/עמודות, היסטוגרמות, פאי ("עוגה"), מצולע שכיחויות.

7. ציון תקן, ל- 3 אובייקטים מסוימים ולמשתנה מסוים (ולבחירתכם).

הפרויקטים ניתנים לעשייה לבד, או בזוג (עדיף).