

# הגרסה הלימודית המאוירת

## סטטיסטיקה בימי הקורונה

ימי קורונה. החשש מן האיום הלא מוכר שינה את פניו. ערפל אי-הוודאות ששרר בתחילה הופך, ככל שהזמן חולף והנתונים מצטברים, לסבך מבלבל של טענות מורכבות.

כל מי שמניח ידיו על טורי מספרים ומסוגל להפעיל תוכנה מתאימה יכול ליצור כמה גרפים משכנעים, לחזות את העתיד להתרחש ולהציע נתיבים אל החזרה הנכספת לשגרה. והציבור, שלמד במהירות ובדרך הקשה מהו גידול אקספוננציאלי (מעריכי), צריך עכשיו להתמודד עם תרחישים ותחזיות סותרות: האם הסגר הכרחי, או אולי מיותר? מה מצבנו בהשוואה בינלאומית? מה אפשר ללמוד מדרום קוריאה ומשוודיה? ומתי כל זה ייגמר?

במקום להציע תשובות לשאלות האלה, נציע כמה מושכלות יסוד בסטטיסטיקה שצריכות להילקח בחשבון בעת ניתוח הנתונים והצגתם. כללים אלו יסייעו לקורא להבחין בין תחזיות המבוססות על ניתוח רציני לבין להטוטי מספרים בעלמא.

1. **דע מה אתה סופר.** הנתון הנפוץ ביותר בדיווחים החדשניים ובניתוחים המבוססים עליהם (כולל של חוקרים בכירים) הוא "מספר החולים" במדינה פלונית. אלא שבשל מגבלות הבדיקה, המספר הזה, שכל כך הרבה תלוי בו, אינו ידוע! אנו יודעים את **מספר החולים המאומתים**, אלו שנבדקו או שהגיעו לבתי החולים. יתכן שמספר החולים בפועל עולה על המספר הזה פי כמה וכמה. למשל, אחוז המתים באיטליה מכלל החולים המאומתים הוא 13% בערך. אף אחד אינו סבור שאחוז התמותה מקורונה גבוה כל כך, גם לא באיטליה; הסיבה היא שחלק מהחולים מעדיפים לא להיבדק, או שאין מספיק בדיקות עבורם.

חתולים מאומתים



חתולים



2. **נתונים יחסיים ולא מוחלטים.** התקשורת רגילה להציג את האירועים דרך הדרמה האישית וסיפורו המעניין של הפרט. אבל כדי להבין תופעה מתפתחת, יש לטפל בה בכלים יחסיים. תוספת של 500 חולים היא "דרמטית" כשמספר החולים נאמד במאות, ו"מינורית" כשהוא נאמד בעשרות אלפים. גידול – נכון להציג באחוזים, ולא במספרים מוחלטים.



3. **להשוות בין שווים.** מספר החולים המאומתים באשקלון שווה לזה של אלעד, אבל מצבה של אלעד חמור יותר משום שמספר התושבים באשקלון גדול פי שלושה. בעובדה שירושלים מובילה במספר החולים המאומתים אין שום רבותא: יש בה כפליים אנשים מבכל עיר אחרת. טבלאות השוואתיות צריכות להציג את מספר החולים ל-100,000 איש. טבלה כזו תראה, למשל, שבלגיה היא המובילה במספר המתים היחסי, ומצבה לכן חמור משל איטליה. האם שמעתם עליה לאחרונה?



4. **לחשב יחסים רלוונטיים.** אין כמעט שום קשר בין מספר החולים שהתגלו היום למספר הבדיקות שנערכו היום. זאת משום שתוצאות הבדיקה מגיעות לאחר 3-4 ימים, ולפעמים יותר. הנתון המשמעותי הוא אחוז החולים שהתגלו מכלל הבדיקות שנערכו ביום מסויים, בלי קשר לשאלה באיזה יום התקבלה תשובת המעבדה.

יחסים לא רלוונטיים

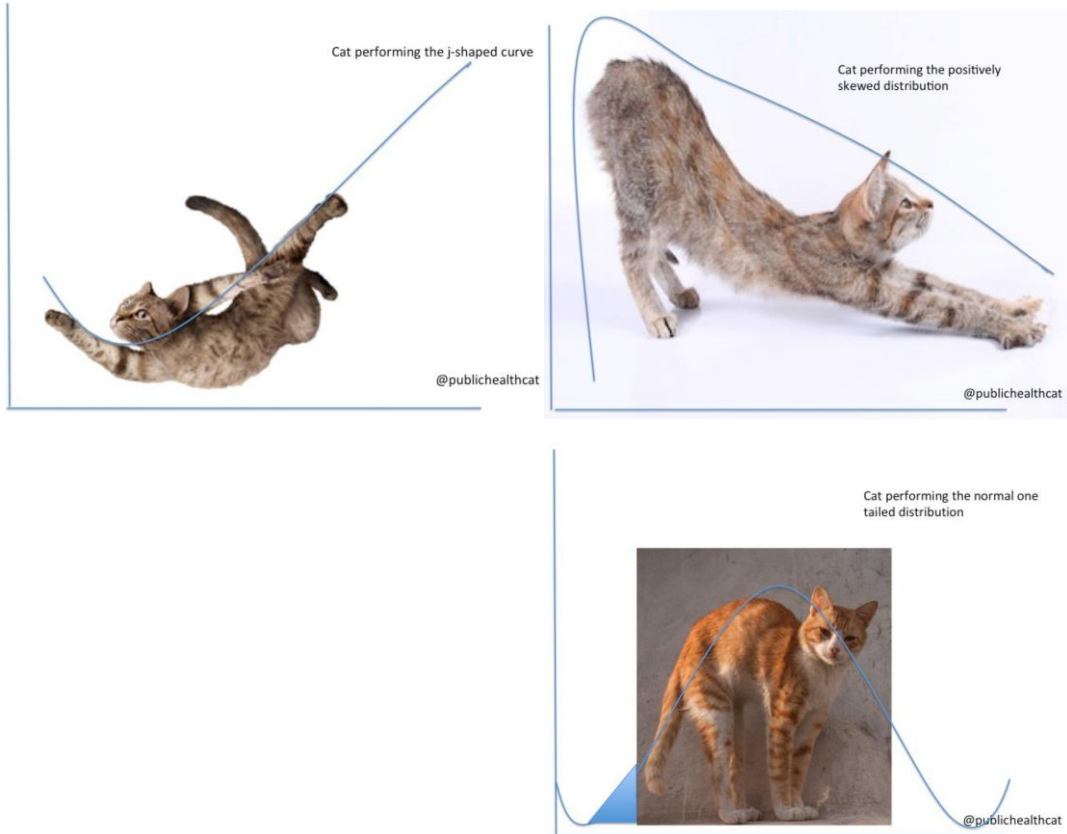


יחסים רלוונטיים



5. **גידול מציגים על ציר לוגריתמי.** אחת השאלות הקריטיות היא האם מספר החולים או המתים גדל בקצב מעריכי (אקספוננציאלי), כלומר, האם מספר זה מוכפל מדי יום בגורם קבוע. אם כן, הדרך הטובה ביותר להציג את הנתונים היא על ציר המתאר את הלוגריתם של מספר החולים במקום את המספר עצמו. בעוד שבגרף רגיל עליה ביחידה אחת פירושה תוספת קבועה, בגרף לוגריתמי עליה ביחידה אחת משמעה הכפלה בגורם קבוע. כך, גידול מעריכי של הנתונים יבוא לידי ביטוי בקו ישר, ואילו עליה מתונה יותר תלך ותשתפע בעדינות כלפי האופק.

**איור: חתולים באופן כללי הם דרך נהדרת להצגת עקומות סטטיסטיות שונות.**

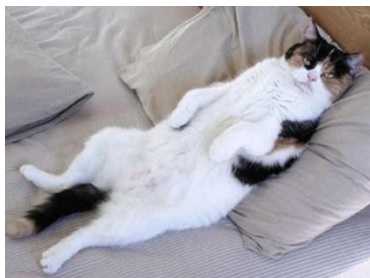


**איור: חתולים מדגימים את עקומת התפשטות הקורונה במדינות שונות**

חתול מדגים את עקומת התפשטות הקורונה בדרום קוריאה

חתול מדגים את עקומת התפשטות הקורונה בצרפת

חתול מדגים את עקומת התפשטות הקורונה באיטליה



6. **דגימה מוטית היא רעה חולה.** הדבר היחיד הגרוע יותר מהיעדר מידע הוא מידע מוטעה. כוחה של סטטיסטיקה הוא להסיק מסקנות על האוכלוסיה מתוך נתונים על קבוצה קטנה (הקרויה "מדגם"). ההיסק נכשל לחלוטין אם הקבוצה שנדגמה היא בעלת תכונות מיוחדות, שאינן אופייניות לכלל האוכלוסיה. ב-2 לאפריל העריך מנכ"ל קופ"ח מכבי באזני ועדת הכנסת כי "38% מתושבי בני ברק חולים, שזה 75 אלף איש". ההערכה המופרכת הזו, שהתבססה על אחוז החולים מבין בעלי סימפטומים שבאו להבדק (!), קנתה לה שביטה וצוטטה אחר כך על ידי חברי כנסת ומקבלי החלטות. אחוז החולים המאומתים בבני ברק גבוה, אבל שבועיים אחרי ההערכה המבהילה הזו אחוז החולים המאומתים הוא כ-1%.

**איור לימודי: אי אפשר להסיק שום דבר על צבע החתולים באוכלוסיה כשדוגמים משפחה אחת של חתולים ג'ינג'ים**

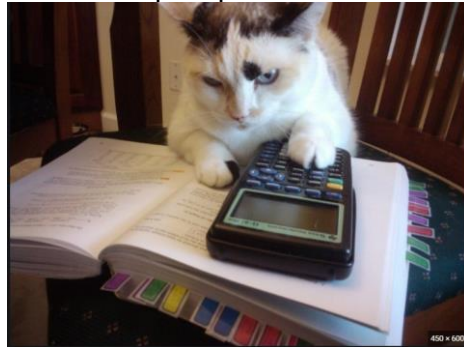


7. **מגמה ולא פרטים.** מספר החולים החדשים מדי יום תלוי בגורמים מקומיים ואקראיים, ולכן נכון להתבונן במגמת הגידול הממוצע לאורך כמה ימים. בדומה לזה, מספר החולים במצב קשה מאפשרת להעריך את העומס על מערכת הבריאות, אבל היא אינה מודדת את התפשטות המחלה אם לא לוקחים בחשבון את מספר המתים. לדוגמא, ב-15.4 היתה ירידה קלה במספר החולים במצב קשה, שלא הצביעה על שיפור במצב – היא נבעה מכך ש-10 מחולים אלה נפטרו.

חתול עוסק בטפל

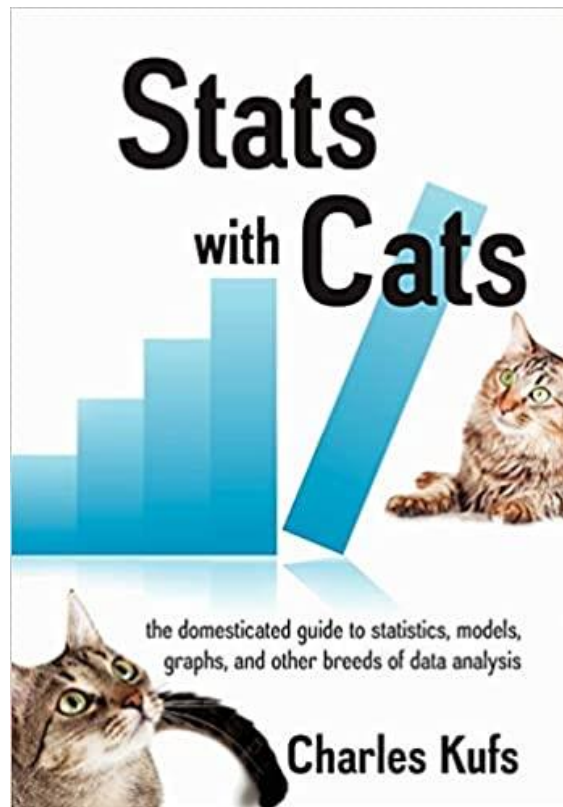


חתול עוסק בעיקר

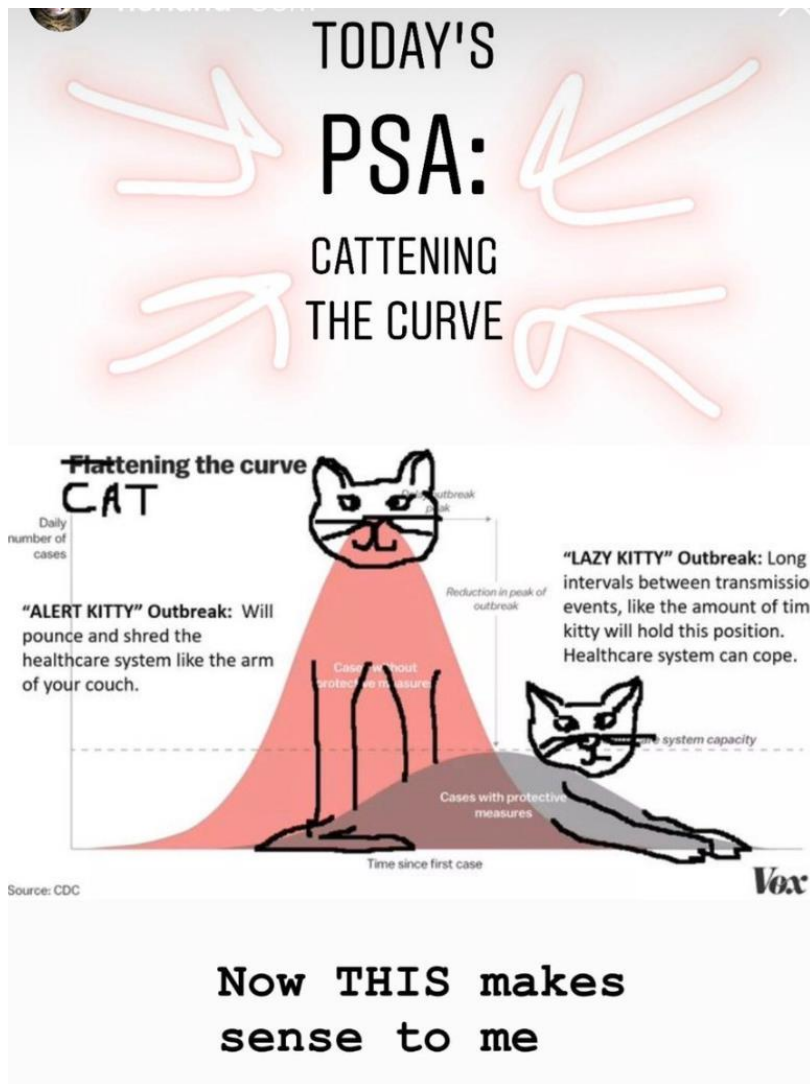


8. **לסטטיסטיקה יש כח.** אחרי כל האזהרות שמנינו, חשוב לזכור שכאשר משתמשים בה נכון, הסטטיסטיקה היא כלי חיזוי חזק מאין כמוהו. דגימה של 500 איש מאפשרת לסקרי הבחירות לתת הערכה טובה מאד לחלוקת המנדטים (טובה במונחים אפידמיולוגיים, שאינם מוטרדים מחודו של קול). באותו אופן, אפשר להשיג תמונת מצב טובה של התפתחות הקורונה על ידי בדיקות מדגמיות לא רבות, וזאת, כמובן, בתנאי שמצליחים לבדוק מדגם מייצג של האוכלוסייה – מתוך ערים שונות, בגילאים שונים וכדומה. המידע על האוכלוסיה הכללית, שטרם נבדקה, נחוץ לנו ביותר, ודגימה אקראית היא אחת הדרכים הטובות ביותר להשיג אותו. יש הטוענים שמספר נשאי הקורונה בישראל מגיע לכדי מאות אלפים. בדיקה מדגמית תוכל לאשר או להכחיש אפשרות כזו ברמת ודאות גבוהה. מצער מאד שיוזמות לבצע בדיקה כזו (בסופמרקטים, או בערים שונות) טורפדו מסיבות פוליטיות.

**איור: לסטטיסטיקה יש כוח, קל וחומר לסטטיסטיקה עם חתולים**







בשימוש נכון [ובמקרה שלנו תוך התחשבות בקהל היעד], סטטיסטיקה היא כלי עבודה אפקטיבי. עלינו להשתמש בו בתבונה.

פרופ' עוזי וישנה ופרופ' נתן קלר הם חברי סגל במחלקה למתמטיקה באוניברסיטת בר אילן

עוזי וישנה [vishne@math.biu.ac.il](mailto:vishne@math.biu.ac.il)

נתן קלר [nkeller@math.biu.ac.il](mailto:nkeller@math.biu.ac.il)