# מבוא לסטטיסטיקה ועיבוד נתונים ב 71724

## פרק 6 - רווח סמך לפרופורציה

### תוכן העניינים

1. רווח הסמך לפרופורציה ..... 1
2. קביעת גודל מדגם ..... 4

# רווח הסמך לפרופורציה:

## רקע:

המטרה היא לאמוד את $P$ – פרופורציה באוכלוסייה.

**האומד הנקודתי:**

$\hat{p}=\frac{y}{n}$ ( $Y$ - מספר ההצלחות שבמדגם).

רווח הסמך ל- $p$ : $\hat{p}\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ .

**תנאי לבניית רווח הסמך:**

מדגם של לפחות 30 תצפיות (לעיתים נותנים תנאי של מספר הצלחות ומספר כשלונות לפחות 5 או לפחות 10).

**האומד לטעות התקן:** $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

**מתקיים ש:** $\hat{P}=\frac{A+B}{2}$ , $L=2\varepsilon$ .

**דוגמה (פתרון בהקלטה):**

במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים, מתוכם התקבל ש-24 היו מובטלים.

א. בנו רווח סמך לאחוז המובטלים באוכלוסייה ברמת סמך של 95%.

ב. מהו האומד לטעות התקן?


© כל הזכויות שמורות לברק קנדל

## שאלות:

**1)** נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.

א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.

ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?

ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95% למספר הדירות בחיפה עם ממ"ד בפועל.

**2)** במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.

א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי הייטק).

ב. כיצד רווח הסמך של סעיף א' היה משתנה אם היינו מקטינים את רמת הסמך?

ג. כיצד רווח הסמך היה משתנה אם הינו מגדילים את גודל המדגם?

**3)** במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים: $0.08 < p < 0.18$.

א. כמה נהגים במדגם היו נהגים חדשים?

ב. מהי רמת הסמך של רווח הסמך שנבנה?

**4)** במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו. 510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של $\pm 3\%$ מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?

**5)** במדגם של 300 נשים בגילאי 35-40 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.

א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.

ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?

**6)** ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95%. אורכו הינו 8.3156%. מהו גודל המדגם שנלקח?


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות:

1) א. $18.1\% < p < 29.9\%$.
ב. בביטחון של 95% שגיאת האמידה היא לכל היותר 0.059.
ג. $14,480 < \mu < 23,920$.

2) א. $0.545 \le p \le 0.655$. ב. האורך שלו היה קטן. ג. לא ניתן לדעת.

3) א. 52. ב. 0.997.

4) 0.925.

5) א. $30.9\% > p > 22.5\%$. ב. $60.72\% > p > 45.91\%$.

6) 200.


© כל הזכויות שמורות לברק קנדל

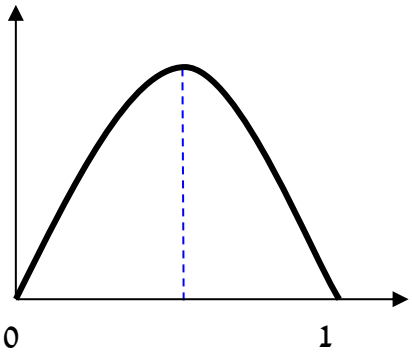
# קביעת גודל מדגם:

## רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת:
החוקר קובע מראש את רמת הסמך הרצויה: $1-\alpha$.
החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: $\varepsilon$
(או את אורך רווח הסמך).

$L=2\varepsilon$ - אורך רווח הסמך.
$\varepsilon$ - טעות אמידה מרבית: המרחק המקסימאלי (הסטייה) בין הפרמטר ( $p$ ) לאומד ( $\hat{p}$ ).

$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$




ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

נקבל ש: $n \geq \left( \frac{2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2$.

הבעיה שאין אנו יודעים את $\hat{p}$.
נתבונן בביטוי: $\hat{p}(1-\hat{p})$.

כיוון שאין לנו ידע מוקדם על $\hat{p}$ נציב את המקרה השמרני ביותר שממקסם את הביטוי עבור: $\hat{p}=0.5$.

$$n \geq \left( \frac{2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{0.5 \cdot 0.5}}{L} \right) \Rightarrow n \geq \left( \frac{z_{1-\frac{\alpha}{2}}}{L} \right).$$

אך אם תהיה לנו אינפורמציה מוקדמת על הפרופורציה נציב את הערך הקרוב ביותר ל-0.5 האפשרי.

**דוגמה (פתרון בהקלטה):**

מעוניינים לאמוד את שיעור האבטלה במשק. האמידה צריכה להתבצע ברמת סמך של 90% ועם שגיאת אמידה שלא תעלה על 4%.

א. מהו גודל המדגם המינימאלי שיש לקחת?
ב. חזור לסעיף א' אם ידוע שהאבטלה לא אמורה לעלות על 20%.


© כל הזכויות שמורות לברק קנדל

## שאלות:

**1)** הממשלה אומדת מדי חודש את אחוז התמיכה בה.
מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?

**2)** משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.

א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?

ב. חזרו על סעיף א' אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.

**3)** ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים.
המטרה שבביטחון של 95% הסטייה המרבית בין האומד לרייטינג האמתי לא תעלה על 4%.

א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?

ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?

**4)** ענו על הסעיפים הבאים:

א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?

ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.

ג. על סמך סעיף ב', האם תקבלו את הטענה שמיעוט האוכלוסייה תומך הממשלה?

**5)** משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.

א. כמה מחוסנים יש לדגום?

ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבת בסעיף הקודם וקיבל ש-15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.

ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98%? מדוע הוא קטן מ-3%?


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות:

1) 1068.

2) א. 423. ב. 271.

3) א. 601. ב. 108,000₪.

4) א. 335. ב. $0.367 < p < 0.473$.

ג. בביטחון של 0.95 ניתן להגיד שמיעוט באוכלוסייה תומך בממשלה.

5) א. 1509. ב. $0.15 \pm 0.02$ ג. ראה סרטון.


© כל הזכויות שמורות לברק קנדל