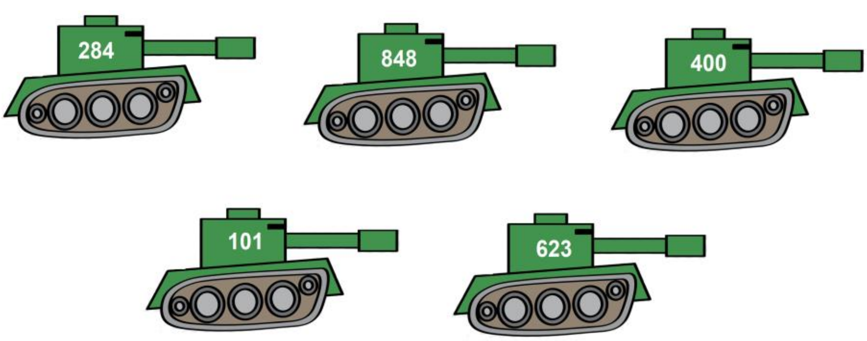
מבוא לסטטיסטיקה - אמידה

- מרגל ראה חמישה טנקים. כמה טנקים יש בסך הכל?



פתרון אפשרי

- X הוא משתנה מקרי המציין את מספרו של טנק מקרי

$$E(X) = \frac{1+2+\cdots+N}{N} = \frac{N+1}{2}$$

$$\bar{X}_n = \frac{284+848+400+101+623}{5} = 451.2$$

$$451.2 \approx \frac{N+1}{2}$$

$$N \approx 2 \cdot 451.2 - 1 \approx 901$$

פתרון אחר

- נסתכל על ההפרשים בין המספרים:

$$\underbrace{1 \quad 101}_{100} \quad \underbrace{\quad 284}_{184} \quad \underbrace{\quad 400}_{116} \quad \underbrace{\quad 623}_{223} \quad \underbrace{\quad 848}_{225}$$

- נחשב את ממוצע ההפרשים ונוסיף אותו למספר הגדול ביותר

$$\frac{(848-623)+(623-400)+\cdots.+(101-1)}{5} = \frac{848-1}{5} = 169.4$$

$$848 + 169.4 = 1017.4$$

איזה אמדן יותר טוב: 901 או 1017?

דוגמה: כדורים בכד

- בכד יש 90 כדורים. מספר הכדורים האדומים בכד הוא 45 או 55 – אין אפשרות אחרת.
- מוציאים מהכד 30 כדורים עם החזרה. מתוך ה-30, 22 היו אדומים. אז כמה כדורים אדומים יש בכד?
- מספר הכדורים האדומים שנשלפו הוא משתנה מקרי בינומי X עם n=30

- אם בכד יש 55 כדורים אדומים אז p=0.611 ואז P(X=22)=0.060
- אם בכד יש 45 כדורים אדומים אז p=0.5 ואז P(X=22)=0.0054

עקרון הנראות המקסימלית: אם ראינו משהו, סימן שהסיכוי שזה יקרה הוא גבוה

לפי עקרון הנראות המקסימלית, האמדן שלנו למספר הכדורים האדומים בכד הוא 55

דוגמה: סקר בחירות

- אדם מסויים מעוניין לרוץ לראשות הממשלה. הוא עורך סקר של מדגם מייצג של 8 אנשים. 5 מתוכם תומכים במועמדותו. מה האמדן לשיעור התמיכה באוכלוסיית הבוחרים?
- נסמן ב-X את מספר התומכים במדגם, וב-p את שיעור התמיכה באוכלוסייה. X הוא משתנה מקרי בינומי עם n=8 והסתברות p להצלחה. p הוא הפרמטר שאנחנו רוצים לאמוד.
- פתרון אפשרי: שיעור התומכים במדגם = שיעור התומכים באוכלוסייה.
- לכן האמדן לשיעור התומכים האוכלוסייה הוא 62.5%

דוגמה: סקר בחירות

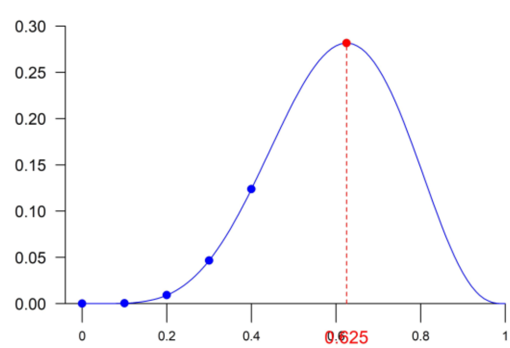
- פתרון אחר: אם היו 5 תומכים במדגם, אז ההסתברות לכך היא גבוהה. מספר התומכים במדם הוא משתנה בינומי עם n=8 והסתברות להצלחה p. ולכן

$$P(X=5)=\binom{8}{5}\cdot p^5\cdot(1-p)^3=56\cdot p^5\cdot(1-p)^3$$

- עבור איזה ערך של p ההסתברות הזו מקסימלית?

דוגמה: סקר בחירות

$$f(p)=56\cdot p^5\cdot(1-p)^3$$



לפי עקרון הנראות המקסימלית האמדן ל-p הוא 0.625

מסגרת כללית

- נתונה פונקציית התפלגות עם פרמטר θ
- $X_1, X_2 \dots, X_n$ הם משתנים מקריים בלתי תלויים מההתפלגות הנתונה – מדגם מייצג
- $\hat{\theta}_n$ הוא **אמד** ל-θ – נוסחה בה מופיעים $X_1, X_2 \dots, X_n$
- לאחר שמציבים את בנוסחה את ערכי $X_1, X_2 \dots, X_n$ שהתקבלו במדגם אנו מקבלים **אמדן** ל-θ

דוגמת הטנקים

- התפלגות מספרי הטנקים היא התפלגות אחידה עם פרמטר N. כלומר: אם X הוא ההסתברות כי המספר של הטנק שצפינו בו אז הוא k היא $P(X = k) = \frac{1}{N}$
- צפינו במדגם $X_1, X_2 \dots, X_n$, כאשר *n=5*. הערכים שנצפו היו : 284, 848, 400, 101 ו-623
- הוצעו שני אמדים אפשריים:
  - $\hat{N} = 2 \cdot \bar{X}_n + 1$
  - $\hat{N} = M + \frac{M-1}{n}$ כאשר $M = max(X_1, X_2 \dots, X_n)$
- על ידי הצבת ערכי המדגם באמדים קיבלנו שני אמדנים אפשריים: 901 ו-1017

דוגמת הסקר

- התפלגות מספר התומכים היא התפלגות בינומית עם פרמטרים $n=8$ ו-p
- צפינו למעשה במדגם בגודל **1!** הערך בו צפינו היה $X=5$.
- האמד שהוצע היה $\hat{p}=\frac{X}{n}$, והגענו אליו על ידי שני שיקולים שונים
- האמדן שלנו ל-p הוא לכן 0.625.

הערכת טיב האמד – תוחלת הטעות הריבועית - MSE

$$MSE(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right]$$
$$= E\left[\hat{\theta}^2 - 2\cdot\theta\cdot\hat{\theta}+\theta^2\right]$$
$$= E(\hat{\theta}^2) - 2\cdot\theta\cdot E(\hat{\theta})+\theta^2$$
$$= E(\hat{\theta}^2) - E(\hat{\theta})^2 + E(\hat{\theta})^2 - 2\cdot\theta\cdot E(\hat{\theta})+\theta^2$$
$$= V(\hat{\theta}) + \left[E(\hat{\theta})-\theta\right]^2$$

תוחלת הטעות הריבועית של אמד שווה לשונות האמד ועוד ריבוע ההטיה

אמד חסר הטיה הוא אמד שעבורו ההטיה שווה לאפס, כלומר $E(\hat{\theta})=\theta$

דוגמה: טנקים

- טנקים (1)

$$\mu = 1\cdot\frac{1}{N} + 2\cdot\frac{1}{N} + \cdots + N\cdot\frac{1}{N} = \frac{N+1}{2}$$

$$E(\bar{X}_n) = \mu = \frac{N+1}{2}$$

$$E(\widehat{N}) = E(2\cdot\bar{X}_n - 1) = 2\cdot\frac{N+1}{2} - 1 = N$$

- טנקים (2)

$$E[max(X_1, X_2 \ldots, X_n)] = ???$$

דוגמה: סקר

$$\hat{p} = \frac{X}{n}$$

$$E(\hat{p}) = \frac{E(X)}{n} = \frac{n\cdot p}{n} = p$$

$$MSE(\hat{p}) = V(\hat{p}) + 0 \quad = V\left(\frac{X}{n}\right) \quad = \frac{1}{n^2}\cdot V(X) \quad = \frac{n\cdot p\cdot(1-p)}{n^2} \quad = \frac{p\cdot(1-p)}{n}$$

אמידת השונות

- תזכורת מסטטיסטיקה תיאורית
  - הגדרנו את השונות כסכום ריבועי הסטיות מהממוצע
  - נוסחה מקוצרת לחישוב השונות: ממוצע הריבועים פחות הממוצע בריבוע
- תזכורת משתנים מקריים $V(X) = E[(X-\mu)^2] = E[X^2] - \mu^2$
- אמידת מומנטים:

$$\widehat{E(X)} = \hat{\mu} = \bar{X}_n \qquad \widehat{E[X^2]} = \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}$$

לכן, למעשה, השונות כפי שהוגדרה בפרק הסטטיסטיקה התיאורית היא למשה אמד מומנטים לשונות של משתנה מקרי

אמד חסר הטיה לשונות

ראינו כי אמד מומנטים לשונות הוא $\hat{\sigma}^2 = \frac{(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \cdots + (X_n-\bar{X})^2}{n} = \frac{\sum(X_i-\bar{X})^2}{n}$

ניתן להוכיח כי $E(\hat{\sigma}^2) = \frac{n-1}{n}\cdot\sigma^2$ כלומר אמד המומנטים אינו אמד חסר הטיה

אבל אם נגדיר $S^2 = \frac{n}{n-1}\cdot\hat{\sigma}^2$ נראה כי זהו אמד חסר הטיה לשונות:

$$E(s^2) = E\left(\frac{n}{n-1}\cdot\hat{\sigma}^2\right) = \frac{n}{n-1}\cdot E(\hat{\sigma}^2) = \frac{n}{n-1}\cdot\frac{n-1}{n}\cdot\sigma^2 = \sigma^2$$

מצד שני:

$$S^2 = \frac{n}{n-1}\cdot\hat{\sigma}^2 = \frac{n}{n-1}\cdot\frac{\sum(X_i-\bar{X})^2}{n} = \frac{\sum(X_i-\bar{X})^2}{n-1}$$

סיכום

- $\hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{n}$ הוא אמד מומנטים לשונות
- $s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$ הוא אמד חסר הטיה לשונות
- לכן $\hat{\sigma}$ ו- S הם אמדים לסטיית התקן
- עם זאת $\hat{\sigma}$ ו- S אינם אמדים חסרי הטיה לסטיית התקן