

סטטיסטיקה והסתברות

פרק 23 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות פרופורציית ההצלחות במדגם 1
2. חוק המספרים הגדולים 5

התפלגות פרופורציית ההצלחות במדגם:

רקע:

בפרק זה נדון בהתפלגות הדגימה של פרופורציית המדגם.
 Y - מספר ההצלחות במדגם (למשל, מספר המובטלים במדגם).

$$\hat{p} = \frac{Y}{n} - \text{פרופורציית ההצלחות במדגם.}$$

למשל, שיעור המובטלים במדגם - $n = 200$:

מספר המובטלים : $Y = 20$.

$$\hat{p} = \frac{20}{200} = 0.1 : \text{פרופורציית המובטלים במדגם.}$$

נסמן ב- p את שיעור ההצלחה באוכלוסייה וב- q את שיעור הכישלונות באוכלוסייה.
 נבצע מדגם מקרי (הנחה שהתצפיות בלתי תלויות זו בזו) ונתבונן בהתפלגות של פרופורציית המדגם.

התוחלת, השונות וסטיית התקן של פרופורציית המדגם:

$$E(\hat{p}) = p, \quad V(\hat{p}) = \frac{pq}{n}$$

משפט הגבול המרכזי עבור הפרופורציה המדגמית:

$$\text{אם: } np \geq 5 \text{ \& } nq \geq 5, \text{ אזי: } \hat{p} \sim N\left(p, \frac{pq}{n}\right) \quad Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

הערות:

- התנאים לקרוב הנורמאלי הם נזילים, כלומר משתנים ממרצה אחד לשני.
 התנאי שהצגתי כאן הוא הפופולרי ביותר:
 1. $n \cdot p \geq 5$
 2. $n \cdot (1 - p) \geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
 1. $n \cdot p \geq 10$
 2. $n \cdot (1 - p) \geq 10$
- וישנם מרצים המשתמשים בתנאי: $(n \geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור לנורמלית.

- כיוון שפרופורציה אינה חייבת להיות מספר שלם בהכרח לא נהוג לבצע כאן תיקון רציפות.

דוגמה (פתרון בהקלטה):

לפי נתוני משרד החינוך בעיר ירושלים ל-60% מתלמידי התיכון זכאים לתעודת בגרות. נדגמו 200 תלמידי תיכון.

- מה ההסתברות שהשכיחות היחסית ($\hat{p}$) של הזכאים לבגרות במדגם תעלה על 60%?
- מה ההסתברות שפרופורציית הזכאים לבגרות במדגם תעלה על 70%?

שאלות:

- (1) במדינה מסוימת 10% מכלל האוכלוסייה הינם מובטלים. נדגמו באקראי 140 אנשים מהמדינה.
 - א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
 - ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
 - ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?
- (2) נניח כי 30% מהאוכלוסייה תומכים בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
 - א. לפחות 35% יתמכו בהצעת החוק במדגם.
 - ב. לכל היותר 25% יתמכו בהצעת החוק במדגם.
 - ג. יותר מ-27% יתמכו בהצעת החוק במדגם.
- (3) לפי נתוני משרד התקשורת 40% מהאוכלוסייה מחזיקים בטלפון נייד מסוג "סמארטפון". נדגמו 400 אנשים מהאוכלוסייה.
 - א. מה ההסתברות שבמדגם לכל היותר ל-40% יש סמארטפון?
 - ב. מה ההסתברות שבמדגם לרוב יש סמארטפון?
 - ג. מה ההסתברות שפרופורציית בעלי הסמארטפון במדגם תסטה מהפרופורציה באוכלוסייה בלא יותר מ-4%?
 - ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם?
- (4) נתון כי 80% מבתי האב מחוברים לאינטרנט. נדגמו 400 בתי אב אקראיים.
 - א. מה ההסתברות שלפחות 340 מהם מחוברים לאינטרנט?
 - ב. מה ההסתברות שפרופורציית המחוברים לאינטרנט במדגם תסטה מהפרופורציה האמיתית ביותר מ-4%?
 - ג. כמה בתי אב יש לדגום כדי שהסטייה בין הפרופורציה המדגמית לפרופורציה האמיתית לא תעלה על 3% בהסתברות של 90%?
 - ד. מהו העשירון התחתון של התפלגות פרופורציית המדגם?
- (5) נתון שציוני פסיכומטרי מתפלגים נורמלית עם תוחלת 500 וסטיית תקן 100. ב"מועדון ה-700" נכללים נבחנים שמקבלים ציון מעל 700 בפסיכומטרי. מה הסיכוי שבמועד בו נבחנו 2000 נבחנים אקראיים יהיו לפחות 3% המשתייכים למועדון?

(6) נתון ש- $X \sim B(n, p)$, ונגדיר את המשתנה הבא: $\hat{P} = \frac{X}{n}$.

א. הוכיחו ש: $E(\hat{P}) = p$, $V(\hat{P}) = \frac{p(1-p)}{n}$.

ב. מה p המביא את $V(\hat{P})$ להיות במקסימום?

תשובות סופיות:

- (1) א. התוחלת: 0.1, השונות: 0.00064. ב. 0.5. ג. 0.3446.
- (2) א. 0.0618. ב. 0.0618. ג. 0.8238.
- (3) א. 0.5. ב. 0. ג. 0.8968. ד. גדלה.
- (4) א. 0.0062. ב. 0.0456. ג. 0.481. ד. 0.77436.
- (5) 0.0154.
- (6) א. שאלת הוכחה. ב. 0.5.

חוק המספרים הגדולים:

רקע:

חוק המספרים הגדולים מתייחס להשפעת הגדלת גודל המדגם על הסיכוי של פרופורציית המדגם (או ממוצע המדגם) להיות קרובה מהפרופורציה האמיתית (או מהממוצע האימתי).

החוק לגבי פרופורציה:

נניח שמבצעים מדגם מקרי מתוך אוכלוסייה אינסופית בה פרופורציית ההצלחות היא p , ככל שהמדגם גדול יותר: כך הסיכוי שפרופורציית המדגם ($\hat{p}$) תהיה בקרבת הפרופורציה באוכלוסייה (P) גבוה יותר.

ולכן הסיכוי לקבל ערך חריג הרחוק מהפרופורציה של האוכלוסייה קטן יותר. בכתיבה מתמטית רושמים את חוק המספרים הגדולים לגבי הפרופורציה באופן

$$\text{הבא: } \lim_{n \rightarrow \infty} \hat{P}_n = P$$

בספרות המקצועית קוראים לחוק הזה החוק החזק של המספרים הגדולים. את החוק החלש של המספרים הגדולים רושמים באופן הבא: $P(|\hat{P} - P| < \varepsilon) \xrightarrow{n \rightarrow \infty} 1$.

הערה: ככל שהמדגם גדל הסיכוי שפרופורציית המדגם תהיה בדיוק הפרופורציה האמיתית הולך וקטן.

החוק לגבי ממוצע: נניח שמבצעים מדגם מקרי מתוך התפלגות שבה התוחלת μ והשונות סופית. ככל שהמדגם גדול יותר, כך הסיכוי שממוצע המדגם ($\bar{X}$) יהיה בקרבת הממוצע באוכלוסייה (μ) גבוה יותר. לכן, הסיכוי לקבל ערך חריג הרחוק מהממוצע של האוכלוסייה קטן יותר. בכתיבה מתמטית רושמים את חוק המספרים הגדולים לגבי הממוצע באופן הבא: $\lim_{n \rightarrow \infty} \bar{X}_n = \mu$.

בספרות המקצועית קוראים לחוק הזה החוק החזק של המספרים הגדולים. את החוק החלש של המספרים הגדולים רושמים באופן הבא: $P(|\bar{X} - \mu| < \varepsilon) \xrightarrow{n \rightarrow \infty} 1$.

דוגמה (פתרון בהקלטה):

באוכלוסייה מסוימת 20% מהאוכלוסייה מובטלת. איזה סיכוי יותר גבוה? במדגם בגודל 100 פרופורציית המובטלים תסטה מהפרופורציה האמיתית בלא יותר מ-4%. במדגם בגודל 200 פרופורציית המובטלים תסטה מהפרופורציה האמיתית בלא יותר מ-4%. הסבירו.

שאלות:

- (1) באוכלוסייה מסוימת 20% מהאוכלוסייה מובטלת. איזה סיכוי יותר גבוה?
א. אחד מתוך מדגם של חמישה יהיה מובטל.
ב. שניים מתוך עשרה יהיו מובטלים. הסבירו וחשבו.
- (2) באוכלוסייה מסוימת 20% מהאוכלוסייה מובטלת. איזה סיכוי יותר גבוה?
א. לפחות 3 מתוך 10 יהיו מובטלים
ב. לפחות 30 מתוך 100 יהיו מובטלים. הסבירו.
- (3) גובה של אוכלוסייה מסוימת מתפלג נורמלית עם ממוצע 170 ס"מ וסטיית תקן של 10 ס"מ. דוגמים 4 אנשים באקראי.
א. מה ההסתברות שהגובה הממוצע שלהם יהיה מעל 176 ס"מ.
ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם? נמקו.
- (4) ידוע שבהצעת חוק מסוימת תומכים 60% מהציבור. נדגמו באקראי 10 אנשים.
א. מה ההסתברות שבדיוק 60% מהמדגם תומכים בהצעת החוק?
ב. כיצד התשובה הייתה משתנה אם היו דוגמים 20 אנשים?
- (5) שני חוקרים ביצעו מדגם מאותה אוכלוסייה. חוקר א דגם 20 תצפיות והשני דגם 40 תצפיות וכל אחד מהם חישב את ממוצע המדגם: $\bar{X}_{20}$ ו- $\bar{X}_{40}$. ידוע שהתפלגות היא נורמלית ושהתוחלת באוכלוסייה הינה 500. בסעיפים הבאים נמקו אילו הסתברויות מהזוגות המוצגים גדולה יותר או שהן שוות ונמקו.
א. $P(\bar{X}_{20} > 500)$ או $P(\bar{X}_{40} > 500)$
ב. $P(480 < \bar{X}_{20} < 520)$ או $P(480 < \bar{X}_{40} < 520)$
ג. $P(\bar{X}_{20} < 490)$ או $P(\bar{X}_{40} < 490)$
- (6) נתון ש: $X \sim G(P=0.1)$. מבצעים מדגם אקראי בגודל n מההתפלגות הזו ומחשבים את ממוצע המדגם: $\bar{X}_n$. הוכיחו: $\lim_{n \rightarrow \infty} \bar{X}_n = 10$.

תשובות סופיות:

- (1) אחד מתוך מדגם של חמישה יהיה מובטל.
- (2) לפחות 3 מתוך 10 יהיו מובטלים.
- (3) א. 0.1151. ב. קטנה.
- (4) א. 0.2508. ב. קטן.
- (5) א. $P(\bar{X}_{40} > 500) = P(\bar{X}_{20} > 500)$.
- ב. $P(480 < \bar{X}_{40} < 520) > P(480 < \bar{X}_{20} < 520)$.
- ג. $P(\bar{X}_{40} < 490) < P(\bar{X}_{20} < 490)$.
- (6) שאלת הוכחה.