סטטיסטיקה היסקית - הסקה לגבי כלל האוכלוסייה בהינתן ידע על מדגם (חלק מהאוכלוסיה)

סימונים:

	מדגם חלקי	כלל האוכלוסיה
ממוצע	$\bar{x}$	μ
שונות	s^2	σ^2

רווח סמך עבור הממוצע האמיתי של כלל האוכלוסיה (μ)

$$\bar{x} - \epsilon \leq \mu \leq \bar{x} + \epsilon$$

אנחנו רוצים לפגוע בסיכוי מסויים: $P(\bar{x} - \epsilon \leq \mu \leq \bar{x} + \epsilon) = 1 - \alpha$

סימונים

α - רמת מובהקות - סיכוי שהממוצע האמיתי של כלל האוכלוסיה הוא מחוץ לטווח

$1 - \alpha$ - רמת בטחון / סמך - סיכוי שהממוצע האמיתי של כלל האוכלוסיה אכן בתוך הטווח

$L = 2\epsilon$ - אורך רווח הסמך

$\frac{L}{2} = \epsilon$ - מחצית אורך רווח סמך - הסטייה המקסימלית (ברמת הביטחון) בין בממוצע של המדגם לבין הממוצע האמיתי של האוכלוסיה

n - גודל המדגם

מקרה1: רווח סמך עבור הממוצע האמיתי של כלל האוכלוסיה (μ), כאשר נתונה השונות האמיתית בכלל האוכלוסיה (σ^2)

$\frac{L}{2} = \epsilon = Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ כלומר $\bar{x} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

- חישוב גודל מדגם מינימאלי (n) להגבלת אורך רווח הסמך: $L = 2Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \iff n = \left(\frac{2\sigma Z_{1-\frac{\alpha}{2}}}{L}\right)^2$

דוגמה:

חוקר רוצה להעריך ברמת בטחון של 95% את ממוצע השכר בישראל. ידוע שסטיית התקן באוכלוסייה היא 2000 ₪. מהו גודל המדגם הנדרש כך ש א) אורך רווח השמך לא יעלה על 400 שח? ב) הסטייה בין המדגם לאוכלוסייה לא תעלה על 400 שח?

פתרון:

א) $\alpha = 0.05$, נציב בנוסחא שפיתחנו כאשר σ=2000. $Z_{1-\frac{0.05}{2}} = 1.96$. מכאן $n \geq \left(\frac{2\cdot1.96\cdot2000}{400}\right)^2$ ש- $n \geq 384.16 \approx 385$

ב) $n \geq \left(\frac{2\cdot1.96\cdot2000}{800}\right)^2$ מכאן ש- $n \geq 96.04 \approx 97$

שאלה 1

מכונה מייצרת ברגים שאורכם מתפלג נורמלית עם תוחלת 4 ס"מ וסטיית תקן 0.2 ס"מ.
עקב תקלה במכונה הועלה חשד כי המכונה אינה מייצרת ברגים באורך הנדרש.
לשם בדיקה נלקח מדגם מקרי של 25 ברגים אשר יוצרו במכונה לאחר גילוי התקלה ונמצא כי האורך הממוצע שלהם הוא 3.9 ס"מ . נניח כי השונות נותרה ללא שינוי.

א. מהו רווח סמך (רו"ס) לתוחלת אורך הברגים לאחר התקלה ברמת מובהקות (ר"מ) 2% ?
ב. רוצים למצוא רווח סמך לתוחלת, ברמת בטחון של 95%, ע"י לקיחת מדגם של מספר מסויים של ברגים. כמה ברגים לפחות צריך לקחת על מנת שאורכו של רווח הסמך יהיה לכל היותר 0.1 ס"מ ?
ג. בנו רו"ס בגודל 0.118 ס"מ בעזרת מדגם בגודל 49 ברגים. מה רמת המובהקות α ?

פתרון:

א). נסמן ב- X את אורך הברגים לאחר התקלה , $X \sim N(\mu, 0.2^2)$. הפרמטר μ – לא ידוע.

כמו כן נתון : $n = 25$ ו- $\bar{X} = 3.9$

ונתונה רמת מובהקות של 2% $\Leftarrow$ $\alpha = 0.02$ $\Leftarrow$ $Z_{1-\frac{\alpha}{2}} = 2.326$

רו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$$\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

נציב את הנתונים ונקבל: $3.9 - 2.236 \cdot \frac{0.2}{5} \le \mu \le 3.9 + 2.236 \cdot \frac{0.2}{5} \Rightarrow \boxed{3.81 \le \mu \le 3.99}$

ב). נשתמש בנוסחה לרו"ס ברמת מובהקות של α לתוחלת כאשר השונות ידועה:

$$\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

מכאן שגודל רווח הסמך הוא: $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

במקרה שלנו, רמת ביטחון של 95% $\Leftarrow$ $\alpha = 0.05$ $\Leftarrow$ $Z_{1-\frac{\alpha}{2}} = 1.96$.

נרצה שאורך רווח הסמך יהיה לכל היותר 0.1 לכן:

$$2 \cdot 1.96 \cdot \frac{0.2}{\sqrt{n}} \le 0.1 \Rightarrow \sqrt{n} \ge 7.84 \Rightarrow n \approx 61.47 .$$

ז"א שיש צורך במדגם בגודל 62 ברגים לפחות.

ג). גודל רווח הסמך הוא: $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

במקרה שלנו צריך למצוא $\alpha = ?$

$$2Z_{1-\frac{\alpha}{2}} \cdot \frac{0.2}{\sqrt{49}} = 0.118 \quad \Rightarrow Z_{1-\frac{\alpha}{2}} = 2.06 \Rightarrow 1 - \frac{\alpha}{2} = 0.98 \Rightarrow \alpha = 0.04$$

תרגיל:

רופא רוצה לאמוד את ממוצע לחץ השם של אוכלוסיית מבוגרים. לשם כך לשח מדגם מקרי של 225 מבוגרים ונמצא ממוצע של 148. ידוע כי סטטית התקן באוכלוסייה היא 20.

א) חשבו רווח סמך ברמת בטחון של 95%

ב) מהו אורך רווח הסמך?

ג) מה הסטייה בין ממוצע המדגם לתוחלת האוכלוסייה?

ד) אם ניקח מדגם גדול פי ארבע, מה יקרה לאורך הרווח סמך?

ה) חשבו את סעיף א עבור רמת בטחון של 99%

פתרון:

א) $\alpha = 0.05$, נציב בנוסחא לרווח סמך כאשר σ=20. n=225, $\bar{x} = 148$

$\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ נציב בהתאם:

$148 - Z_{1-\frac{0.05}{2}} \cdot \frac{20}{\sqrt{225}} \leq \mu \leq 148 + Z_{1-\frac{0.05}{2}} \cdot \frac{20}{\sqrt{225}}$ נבצע חישובים ונסתכל בטבלת ההתפלגות הנורמאלית כדי למצוא את Z. $Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$, שוב נציב. ונקבל כי:

$145.387 \leq \mu \leq 150.613$ ברמת בטחון של 95%.

ב) $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 2 \cdot 1.96 \cdot \frac{20}{\sqrt{225}} = 5.226$

ג) $\frac{L}{2} = \epsilon = \frac{5.226}{2} = 2.613$

ד) $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ ע"פ הנוסחא נקבל כי L יקטן פי 2.

ה) $= 0.01$, נציב בנוסחא לרווח סמך כאשר σ=20. n=225, $\bar{x} = 148$

$\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ נציב בהתאם:

$148 - Z_{1-\frac{0.01}{2}} \cdot \frac{20}{\sqrt{225}} \leq \mu \leq 148 + Z_{1-\frac{0.01}{2}} \cdot \frac{20}{\sqrt{225}}$ נבצע חישובים ונסתכל בטבלת ההתפלגות הנורמאלית כדי למצוא את Z. $Z_{1-\frac{0.01}{2}} = Z_{0.995} = 2.575$, שוב נציב. ונקבל כי:

$144.57 \leq \mu \leq 151.43$ ברמת בטחון של 99%.

מקרה2: רווח סמך עבור הממוצע האמיתי של כלל האוכלוסיה (μ), כאשר השונות האמיתית בכלל האוכלוסיה לא ידועה

$$S_{n-1} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_i x_i^2 - n\bar{x}^2}{n-1}}$$

ואז

$$\frac{L}{2} = \epsilon = t_{\left(n-1,\ 1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}}$$

כלומר

$$\bar{x} - t_{\left(n-1,\ 1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\left(n-1,\ 1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}}$$

שאלה 2

חוקר מציע סוג של דשן שעשוי להגדיל את משקלם של תפוחים. נלקח מדגם מקרי של 9 תפוחים והתקבלו התוצאות הבאות (בגרמים): $\sum x = 500,\ \sum x^2 = 28166$

א. אמוד ברווח סמך את משקל התפוחים, ברמת סמך של 90%, על סמך תוצאות המדגם הנ"ל.

ב. לאחר האמידה התברר שחלה תקלה שיטתית במדידות ויש להוסיף 8 גרם לכל ערך שנמדד.
הסבר בקצרה, האם וכיצד ישפיע תיקון הטעות על גודל ומיקום רווח הסמך.

פתרון:

א). נחשב את ממוצע המדגם וסטיית התקן המדגמית:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{500}{9} = 55.556, \qquad S = \sqrt{\frac{\sum x_i^2 - n\,\bar{x}^2}{n-1}} = \sqrt{\frac{28166 - 9 \cdot 55.556^2}{9-1}} = 6.966$$

הערך המתאים לנתוני השאלה – מטבלת t:

$n = 9,\ \ 1-\alpha = 0.9$ <= נמצא בטבלה: $t_{8,0.95} = 1.86$

נציב בנוסחת רו"ס:

$$55.55 - 1.86 \cdot \frac{6.966}{3} \le \mu \le 55.55 + 1.86 \cdot \frac{6.966}{3}$$

$$\boxed{51.231 \le \mu \le 59.868}$$

ב). השפעת תיקון הטעות:

הוספה של קבוע לא משפיעה על השונות (וסטיית התקן), לכן גודל רווח הסמך לא ישתנה.

הוספה של קבוע מעלה את הממוצע בקבוע, כלומר הממוצע יגדל ב-8. לכן מיקום רו"ס יוזז ב-8 יח' ימינה.