רווחי סמך

רווחי סמך – מסגרת כללית

- נתונה פונקציית התפלגות עם פרמטר θ
- $X_1, X_2 \dots, X_n$ הם משתנים מקריים בלתי תלויים מההתפלגות הנתונה – מדגם מייצג
- אנו רוצים למצוא תחום מספרי שנוכל לומר ברמת בטחון/רמת סמך מסויימת כי הפרמטר נמצא בתוכו. רמת הסמך היא מספר בין 0 ל-1. בדרך כלל מעוניינים ברמת סמך של 0.9 לפחות.
- התחום מוגדר על ידי L ו-U שהם בעצם שתי נוסחאות לקצה התחתון ולקצה העליון של התחום שאנו מחפשים
- בסופו של דבר נציב את הערכים $X_1, X_2 \dots, X_n$ בנוסחאות אלה ונקבל את התחום המבוקש.
- בדרך כלל רווח הסמך מתבסס על אמד עבור θ , ואז רווח הסמך מאפשר לנו לומר משהו על שולי הטעות של האמד

דוגמה: בקרת איכות

- תרופה מסויימת נמכרת במארז של 30 טבליות, כאשר כל טבליה אמורה לשקול 30 מ"ג.
- ידוע כי משקל כל טבליה מתפלג התפלגות נורמלית, עם סטיית תקן של 2 מ"ג.
- התוחלת של משקל כל טבליה, $\boldsymbol{\mu}$, אינה ידועה ועשויה להשתנות במהלך תהליך הייצור.
- כדי לוודא כי תהליך הייצור תקין, לוקחים כל חצי שעה מדגם של 10 טבליות ושוקלים אותן.
- הערכים שהתקבלו במדגם הם 32.3 31.7 30.8 37.5 31.7 30.3 31.2 32.8 30.6 34.9 (ממוצע 32.38)
- מהו התחום לגביו נוכל לומר ברמת ביטחון/סמך של 95% כי תוחלת משקל הטבליות נמצאת בתוכו?
- לממוצע המדגם יש תוחלת μ וסטיית תקן השווה ל-0.632 מ"ג (2 חלקי שורש 10).
- אנו מחפשים שני מספרים, U ו-L, כך ש: $P(L < \bar{X}_n < U) = 0.95$

דוגמה: בקרת איכות

$$P(L < \bar{X}_n < U) = 0.95$$

$$P(\bar{X}_n < L) = 0.025 \qquad P(\bar{X}_n > U) = 0.025$$

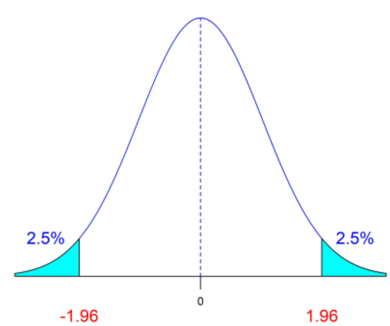
$$P(\bar{X}_n > U) = P\left(\frac{\bar{X}_n - \mu}{0.632} > \frac{U - \mu}{0.632}\right)$$

$$\frac{U - \mu}{0.632} = 1.96$$

$$U = \mu + 1.96 \cdot 0.632$$

$$U = \bar{X}_n + 1.96 \cdot 0.632 = 32.38 + 1.96 \cdot 0.632 = 33.6$$

$$L = 32.38 - 1.96 \cdot 0.632 = 31.1$$

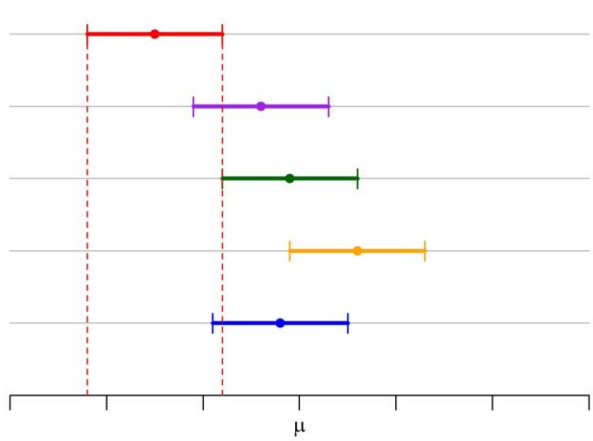


דוגמה: בקרת איכות

- סיכום: מצאנו כי (31.1 ,33.6) הוא רווח סמך ל-μ וזאת ברמת סמך של 95%
- באופן כללי, רווח הסמך הוא $$CI = \bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$
-
- פרשנות: אנחנו יכולים לומר ברמת ביטחון של 95% כי תוחלת משקל הטבליות נמצאת בתחום שבין 31.1 ל-33.6
- מכאן אנחנו גם יכולים להסיק כי ברמת ביטחון של 95% כי תוחלת משקל הטבליות אינה 30 מ"ג
- **פרשנות לא נכונה: ~~ההסתברות כי התוחלת נמצאת בין 31.3 ל-33.6 היא 95%~~.**
  μ, L ו-U אינם משתנים מקריים!

פרשנות רווח הסמך

- **אם נערוך מספר רב של מדגמים, אז בערך ב-95% מהם רווח הסמך "יתפוס" את התוחלת.**



ערכי z

הערך z_α הוא המספר עבורו מתקיים $P(Z < z_\alpha) = 1 - \alpha$ או באופן שקול $P(Z > z_\alpha) = \alpha$

נהוג לסמן את רמת הסמך ב- $1-\alpha$. ערכים מקובלים של רמת הסמך הם 90%, 95% או 99%.

הערכים המתאימים של α הם 5%, 2.5% ו-0.5%

בכל שרמת הסמך עולה – כך רווח הסמך יותר ארוך!

בקרת איכות – המשך הדוגמה

- תזכורת: מפעל תרופות מייצר טבליות, בשכל טבליה אמורה לשקול 100 מ"ג.
- התוחלת של משקל כל טבליה, μ, אינה ידועה ועשויה להשתנות במהלך תהליך הייצור.
- אבל גם סטיית התקן אינה ידועה. מה עושים?

הפתרון הוא להחליף את ערך סטיית התקן σ באמד s

- אם ההתפלגות היא נורמלית והמדגם גדול - משתמשים בערך z
- אם ההתפלגות היא נורמלית והמדגם קטן – במקום ערך z משתמשים בערך t
- אם ההתפלגות לא נורמלית אך המדגם מספיק גדול - משתמשים בערך z
- אם ההתפלגות לא נורמלית והמדגם קטן – צריך לשבור את הראש

בקרת איכות – המשך הדוגמה

אם סטיית התקן אינה ידוע נאמוד אותה על ידי s, ואז רווח הסמך יהיה $CI = \bar{X}_n \pm t_{n-1,\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

$$\bar{X}_n = \frac{34.9 + 30.6 + \cdots + 32.3}{10} = 32.38$$

$$s^2 = \frac{(34.9 - 32.38)^2 + (30.6 - 32.38)^2 + \cdots + (32.3 - 32.38)^2}{9} = 5.01$$

$$s = \sqrt{5.01} = 2.237 \qquad t_{n-1,\frac{\alpha}{2}} = t_{9,0.025} = 2.228$$

$$CI = 32.38 \pm 2.228 \cdot \frac{2.237}{\sqrt{10}} = (30.71, 34.05)$$

בקרת איכות – סיכום הדוגמה

כאשר סטיית התקן הייתה ידועה השתמשנו ברווח הסמך $CI = \bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

רווח הסמך שקיבלנו עבור התוחלת היה (31.1 ,33.6). אורך רווח הסמך הוא 2.5 מ"ג

כאשר סטיית התקן לא הייתה ידועה אמדנו אותה על ידי s והשתמשנו ברווח הסמך $CI = \bar{X}_n \pm t_{n-1,\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

רווח הסמך שקיבלנו עבור התוחלת היה (30.7 ,34.0). אורך רווח הסמך הוא 3.3 מ"ג

רווח הסמך השני הוא ארוך יותר כי הוא השתמש בפחות אינפורמציה מאשר רווח הסמך הראשון

דוגמה: מגורים בדירה שכורה

- במשרד השיכון רוצים לדעת מהו אחוז המשפחות בנות 4 נפשות לפחות המתגוררות בדירה שכורה.
- הלמ"ס דוגם 400 משפחות, ומתברר כי 115 מהמשפחות שנדגמו מתגוררות בדירה שכורה.
- האמדן לפרופורציה באוכלוסייה הוא הפרופורציה במדגם, כלומר 0.2875
- מהו רווח סמך לפרופורציה באוכלוסייה?
- X, מספר המשפחות הגרות בשכירות הוא משתנה מקרי בינומי כאשר *n=400*, ו-*p* אינו ידוע.
- כלומר הפרמטר שאנו רוצים לאמוד הוא *p* - הפרופורציה באוכלוסייה

מגורים בדירה שכורה – המשך הדוגמא

האמד לפרופורציה באוכלוסייה הוא $\hat{p} = X/n$

כפי שכבר חישבנו: $V(\hat{p}) = p(1-p)/n \qquad E(\hat{p}) = p$

לפי משפט הגבול המרכזי (הקירוב הנורמלי להתפלגות הבינומית), ל- $(\hat{p}-p)/\sqrt{p(1-p)/n}$ יש התפלגות נורמלית סטנדרטית אם המדגם מספיק גדול

באופן דומה לחישוב שעשינו עבור מציאת רווח סמך לממוצע המדגם, רווח סמך לפרופורציה ברמת סמך 1-α יהיה

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$$

בדוגמה זו: רווח סמך לפרופורציה ברמת סמך 95% הוא

$$0.2875 \pm 1.96 \cdot \sqrt{0.2875(1-0.2875)/400} = (0.243, 0.332)$$

דוגמה: סקר בחירות

טלוויזיה | חדשות LIVE תאגיד השידור הישראלי כאן

הסקר נערך בקרב 1,406 גברים ונשים בגילאי 18 ומעלה, דגימה משולבת אינטרנטית וטלפונית, כולל מגזר ערבי, בין 17 במארס 2021 עד ל-18 במארס. מספר האנשים שהתבקשו לענות על הסקר הוא 7,542 איש. טעות הדגימה היא 2.6%

$$1/\sqrt{1406} \approx 0.0266$$

חישוב שולי הטעות ("טעות הדגימה המירבית")

כזכור, רווח הסמך לפרופורציה הוא $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$

שולי הטעות הם: $z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$

אנחנו לא יודעים מה ערכו של p, או שאולי אנחנו רוצים לאמוד מספר פרופורציות, ולכן נחליף את p בערך שייתן את סטיית התקן הגבוהה ביותר, שהוא 0.5:

$$z_{\alpha/2} \cdot \sqrt{0.5 \cdot 0.5/n} \ = z_{\alpha/2} \cdot 0.5 \cdot \sqrt{n}$$

אם רוצים רמת סמך של 95% משתמשים "במספר הקסם" 1.96 ומקבלי כי שולי הטעות המרביים הם:

$$1.96 \cdot 0.5/\sqrt{n}$$

בדוגמה של סקר "כאן" שולי הטעות המרביים הם אכן:

$$1.96 \cdot 0.5/\sqrt{1406} = 0.0261$$

חישוב גודל המדגם לאמידת פרופורציה

שולי הטעות בחישוב רווח סמך לפרופורציה הם: $z_{\alpha/2} \cdot \sqrt{p(1-p)/n}$

כמקודם, מאחר ו-p אינו ידוע נחליף אותו בערך בו $p(1\text{-}p)$ מקסימלי ונקבל: $error = 0.5 \cdot z_{\alpha/2}/\sqrt{n}$

כדי לחשב את גודל המדגם נבחר את ערך z המתאים לרמת הסמך שאנו רוצים ואת שולי הטעות המרביים הרצויים ונפתור את המשוואה.

לדוגמא: אם רוצים ששולי הטעות יהיו 2.5% ברמת סמך 90%, אז ערך z המתאים הוא $z_{0.05} = 1.645$

ולכן

$$1.645 \cdot 0.5/\sqrt{n} = 0.025$$

$$\sqrt{n} = \frac{1.645 \cdot 0.5}{0.025} = 32.9 \qquad n \approx 32.9^2 = 1082.41 \qquad n = 1083$$

שימו לב כי גודל המדגם אינו תלוי בגודל האוכלוסייה!

חישוב גודל המדגם לאמידת התוחלת

שולי הטעות בחישוב רווח סמך לתוחלת הם $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

באופן עקרוני אין בעיה לחשב את n אם יודעים את הערך של σ

באופן מעשי הערך של σ בדרך כלל אינו ידוע

לכן יש צורך בהערכה כלשהי לערך σ

שימו לב כי גם כאן גודל המדגם אינו תלוי בגודל האוכלוסייה!

בקרת איכות – המשך הדוגמה

- במפעל התרופות החליטו כי אין בעיה אם הסטייה מהמשקל של 30 מ"ג תהיה 1 מ"ג למעלה או למטה.
- המשמעות היא ששולי הטעות צריכים להיות באורך של 1 מ"ג. רמת הסמך צריכה להיות 95%. ידוע כי סטיית התקן היא 2. מה גודל המדגם הנדרש?

$$z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} = 1$$

$$1.96 \cdot \frac{2}{\sqrt{n}} = 1$$

$$\sqrt{n} = 2 \cdot 1.96$$

$$n = (2 \cdot 1.96)^2 = 15.37$$

$$n = 16$$

בקרת איכות – המשך הדוגמה

- הנתון כי סטיית התקן היא 2 הוא בעצם אמדן מנתוני ייצור שנאספו לאורך זמן.
- לכן, ליתר ביטחון, החליטו להשתמש באמדן 3 עבור ערך סטיית התקן. מה גודל המדגם הדרוש?

$$1.96 \cdot \frac{3}{\sqrt{n}} = 1$$

$$n = (3 \cdot 1.96)^2 = 34.57$$

$$n = 35$$