



תרגיל רווחי סמך ובדיקת השערות

1

שאלה 1



כדי לאמוד את שיעור המתחסנים נגד קורונה בקרב עובדי מערכת הבריאות, נלקח מדגם של 1000 עובדים מ-6 בתי חולים ברחבי הארץ. עובדי המערכת מתחלקים לשתי קבוצות: עובדים במקצועות הבריאות ועובדי מנהל. לכל עובד צויין האם הוא או היא התחסנו נגד קורונה. אנו מעוניינים לחשב רווחי סמך ברמת סמך של 95% או לבדוק השערות ברמת מובהקות של 5%.

2

שאלה 1 סעיף א



מתוך 1000 נדגמים, 650 היו עובדים במקצועות הבריאות. ידוע כי 70% מכלל עובדי מערכת הבריאות עובדים במקצועות הבריאות, והשאר עובדי מנהל. האם נתונים אלה מתיישבים זה עם זה?

נחשב רווח סמך עבור הפרופורציה של העובדים במקצועות בריאות על פי המדגם:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.65 \pm 1.96 \cdot \sqrt{\frac{0.65 \cdot 0.35}{1000}} = (0.62, 0.68)$$

רווח הסמך אינו מכיל את 0.7 ולכן הנתונים לא מתיישבים זה עם זה.

3

שאלה 1 סעיף ב



אם רוצים לאמוד את פרופורציית המתחסנים בקרב כל עובדי מערכת הבריאות בעזרת רווח סמך ברמת סמך של 95%, מה יהיו שולי הטעות המירביים ("טעות הדגימה המירבית")?

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

שולי הטעות (margin of error) הם:

מכיוון שמראש אנו לא יודעים מה תהיה הפרופורציה במדגם, ניקח במקומה את הערך גרוע ביותר האפשרי: 0.5. ולכן שולי הטעות המירביים יהיו שווים ל-

$$z_{\alpha/2} \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}} = 1.96 \cdot \frac{0.5}{\sqrt{1000}} = 0.031$$

4

שאלה 1 סעיף ג



אם אנו רוצים לבנות רווח סמך לפרופורציה של עובדי מערכת הבריאות שהתחסנו, כך שאורך רווח הסמך יהיה 0.04, מהו גודל המדגם הדרוש?

שימו לב כי אורך רווח הסמך הוא:

$$\left(\hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) - \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) = 2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

כלומר כפליים משולי הטעות! לכן החישוב כמעט זהה לזה שבסעיף הקודם

$$2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}} = 2 \cdot 1.96 \cdot \frac{0.5}{\sqrt{n}} = 0.04$$

גודל המדגם הדרוש הוא 2401

5

שאלה 1 סעיף ד



- מתוך העובדים במקצועות הבריאות, 518 התחסנו נגד קורונה.
- בנו רווח סמך לפרופורציה של המתחסנים מקרב העובדים במקצועות הבריאות.
 - האם תדחו את ההשערה כי פרופורציה זו שונה מ-82%?

הפרופורציה של המתחסנים בקבוצת העובדים במקצועות הבריאות היא $518/650 \approx 0.797$

רווח הסמך הוא

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.797 \pm 1.96 \cdot \sqrt{\frac{0.797 \cdot 0.203}{650}} = (0.766, 0.828)$$

לא נדחה את ההשערה כי פרופורציה זו שונה מ-0.82 מאחר ורווח הסמך מכיל את הערך הזה.

6

שאלה 1 סעיף ה



מבין עובדי המנהל, 220 התחסנו נגד קורונה. בדקו את ההשערה כי שיעור המתחסנים בקרב עובדי מקצועות הבריאות גבוה משיעור המתחסנים עובדי המנהל.

כאן עלינו לבדוק השערה חד צדדית להשוואת פרופורציות של שתי אוכלוסיות.

נסמן את פרופורציית המתחסנים בקרב עובדי מקצועות הבריאות ב- p_1 ואת פרופורציית המתחסנים בקרב עובדי המנהל ב- p_2 . ההשערות הן, לכן:

$$H_0: p_1 = p_2 \quad H_1: p_1 > p_2$$

נדחה את השערת האפס אם $\hat{p}_1 > \hat{p}_2$ באופן מובהק, או באופן שקול אם $\hat{p}_1 - \hat{p}_2 > 0$ באופן מובהק, כלומר אם הפרש הפרופורציות גדול מ-

$$z_\alpha \cdot SD(\hat{p}_1 - \hat{p}_2)$$

שימו לב כי כאן מופיע z_α ולא $z_{\alpha/2}$ כי זו השערה חד צדדית.

עבור $\alpha = 0.05$ הערך המתאים הוא $z_\alpha = 1.645$

7

שאלה 1 סעיף ה



מספר עובדי מקצועות הבריאות במדגם הוא $n_1 = 650$ ומספר עובדי המנהל במדגם הוא $n_2 = 350$. כמו כן פרופורציות המתחסנים במדגם הן: $\hat{p}_1 = 518/650 = 0.797$ $\hat{p}_2 = 220/350 = 0.629$

אנו צריכים לדעת מהי סטיית התקן של $\hat{p}_1 - \hat{p}_2$. נחשב את השונות ואחר כך נוציא שורש:

$$V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

כעת נוכל לאמוד את סטיית התקן על ידי הצבת הפרופורציות מהמדגם

$$\widehat{SD} = \sqrt{\frac{0.797 \cdot 0.203}{650} + \frac{0.629 \cdot 0.371}{350}} = 0.030$$

8

שאלה 1 סעיף ה



לסיכום, נדחה את השערת האפס לפיה שיעור המתחסנים בשתי קבוצות העובדים זהה אם הפרש הפרופורציות גדול מ-

$$z_\alpha \cdot \widehat{SD} = 1.645 \cdot 0.030 = 0.0498$$

הפרש הפרופורציות הוא $\hat{p}_1 - \hat{p}_2 = 0.797 - 0.629 = 0.168$.

ערך זה גדול מ-0.0498 ולכן נדחה את השערת האפס ונסיק כי שיעור התחסנים בקרב עובדי מקצועות הבריאות גבוה משיעור המתחסנים בקרב עובדי המינהל.

שאלה 2



חברת שיווק מעוניינת להגדיל את משך הזמן שגולשים נשארים באתר שלה. הועלתה השערה כי שינוי בעיצוב האתר יעזור להגדלת משך הזמן. לכן החברה מבצעת "AB testing": במשך תקופת זמן מסויימת, 50% מהגולשים, הנבחרים באופן מקרי, מופנים לאתר אלטרנטיבי בעיצוב החדש. לכל גולש נמדד משך הזמן בו הוא שהה באתר.

נסחו השערות המתאימות לבעיה, ונמקו מדוע אלה ההשערות שבחרתם.

נסמן ב- μ_0 את תוחלת משך הזמן שבו גולש נשאר באתר בעיצוב הישן וב- μ_1 את תוחלת משך הזמן שבו גולש נשאר באתר בעיצוב החדש.

השערת האפס תהיה $H_0: \mu_0 = \mu_1$

ההשערה האלטרנטיבית יכולה להיות $H_1: \mu_0 < \mu_1$ או $H_1: \mu_0 \neq \mu_1$

שאלה 2 סעיף ב



תארו את שני סוגי הטעויות האפשריות, והסבירו את המשמעות שלהן מנקודת המבט של החברה.

טעות מסוג ראשון: דחיה מוטעית של השערת האפס: החברה תעבור לעיצוב חדש אשר לא יועיל העלאת משך הזמן הממוצע של שהיית גולשים באתר ואולי אף ייגרם נזק

טעות מסוג שני: אי-דחיה מוטעית של השערת האפס: החברה לא תעבור לעיצוב החדש ותחמיץ את ההזדמנות להעלות את משך הזמן הממוצע של שהיית גולשים באתר

11

שאלה 3



משרד הרווחה מעוניין לאמוד את תוחלת מספר ימי המחלה שמנצלים עובדים ממעמד סוציו-אקונומי נמוך במשך שנה. הוצע לקחת מדגם של 200 עובדים, ולשאל אותם כמה ימי מחלה הם ניצלו בשנה האחרונה. ניתן להניח כי המדגם מספיק גדול. מנתוני הביטוח הלאומי ידוע כי המספר הממוצע של ימי מחלה בשנה להם זכאים עובדים אלה הוא 15 ימי מחלה, וסטיית התקן היא 2.5 ימים.

אחד החוקרים במשרד הציע לבנות רווח סמך למספר ימי המחלה הממוצע שמנצלים העובדים, ברמת סמך של 80%. חוקר אחר הציע לבדוק את ההשערה כי לפחות 30% מהעובדים מנצלים 8 ימי מחלה לכל היותר במשך שנה, ברמת מובהקות של 10%. אם יתברר כי אחוז הניצול של ימי המחלה נמוך מדי, המשרד ינקוט בפעולות כדי לברר מדוע הניצול נמוך.

12

שאלה 3 סעיפים א-ב



התברר כי המספר הממוצע של ימי המחלה שנוצלו על ידי העובדים במדגם הוא 7.3 ימים. בנו את רווח הסמך שמציע חוקר א.

שימו לב כי כאן $1 - \alpha = 0.8$ ולכן $z_{\alpha/2} = 1.282$. לכן רווח הסמך הוא

$$\bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 7.3 \pm 1.282 \cdot \frac{2.5}{\sqrt{200}} = (7.07, 7.53)$$

על פי תשובתכם לסעיף א, מה ההסתברות כי התוחלת שווה ל-8 ימים?

תשובה שגויה: הערך 8 לא נמצא בתוך רווח הסמך ולכן ההסתברות היא 0

תשובה שגויה: רמת הסמך היא 80% ולכן ההסתברות היא 0.8

תשובה נכונה: התוחלת μ אינה משתנה מקרי ולכן אין משמעות לשאלה

שאלה 3 סעיף ג



נסחו את ההשערות שמציע החוקר השני. תארו את הטעויות האפשריות ואת ההשלכות שלהן, הן מבחינת המשרד והן מבחינת העובדים.

נסמן ב- p את הפרופורציה של העובדים המנצלים 8 ימי מחלה לכל היותר במשך שנה.

$$H_0: p = 0.3$$

השערת האפס היא כי פרופורציה זו שווה ל-0.3:

$$H_1: p < 0.3$$

ההשערה האלטרנטיבית מייצגת את מה שמשרד הרווחה מעוניין לבדוק, כלומר האם אחוז הניצול של ימי המחלה נמוך מדי, ולכן ההשערה האלטרנטיבית היא

שאלה 3 סעיף ד



מה כלל ההחלטה הדרוש לבדיקת ההשערות של החוקר השני?

נדחה את השערת האפס אם הפרופורציה במדגם קטנה מ-0.3 באופן מובהק, כלומר כאשר

$$\hat{p} < p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1 - p_0)}{n}} = 0.3 - 1.282 \cdot \sqrt{\frac{0.3 \cdot 0.7}{200}} = 0.258$$

15

שאלה 3 סעיף ה



בהנחה שאחוז הניצול האמיתי הוא 20%, הסבירו כיצד תחשבו את העוצמה של המבחן. יש לפרט את החישובים עד לנקודה בה יש לחשב תא ההסתברות שמשתנה מקרי נורמלי מקבל ערך מסויים (כלומר, עד שיש צורך לחשב הסתברות מהסוג $P(Z < a)$ או $P(Z > a)$)

$$1 - \beta = P_{H_1}(\hat{p} < 0.258) = P_{H_1}\left(\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2 \cdot 0.8}{200}}} < \frac{0.258 - 0.2}{\sqrt{\frac{0.2 \cdot 0.8}{200}}}\right) = P(Z < 2.051) = 0.979$$

16

שאלה 4



בארצות הברית, כל אזרח נדרש להגיש דו"ח שנתי למס הכנסה. מילוי הדו"ח מסובך, ואזרחים רבים מבליים זמן רב בהכנתו או פונים לחברות מסחריות המכינות עבורם את הדו"ח ומגישות אותו לרשויות בשם הלקוח.

חברת רואי חשבון מעוניינת לאמוד את תוחלת משך הזמן הנדרשת כדי למלא את הדו"ח בעיר בו היא פועלת. לשם כך היא בחרה מדגם של 120 איש, והתברר כי משך הזמן הממוצע שנדרש לנבחרים במדגם למלא את הדו"ח הוא 23.6 שעות. סטיית התקן ידועה ושווה ל-7 שעות. ניתן להניח כי גודל המדגם מספיק גדול.

17

שאלה 4 סעיפים א-ב



בנו רווח סמך לתוחלת משך השעות הנדרש כדי למלא את הדו"ח

שימו לב כי לא צויינה רמת הסמך! במציאות זהו מצב שכיח. נחליט כי רמת הסמך תהיה 95%. לכן רווח הסמך הוא

$$\bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 23.6 \pm 1.96 \cdot \frac{7}{\sqrt{120}} = (22.3, 24.9)$$

מהם שולי הטעות של רווח הסמך?

שולי הטעות הם ה-"פלוס/מינוס", כלומר

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{7}{\sqrt{120}} = 1.25$$

18

שאלה 4 סעיף ג



החברה מעוניינת ברווח סמך שבו שולי הטעות יהיו שווים לשעה אחת. ציינו שתי אפשרויות בהן ניתן לקבל רווח סמך כזה. איזה אפשרות תעדיפו? נמקו.

אפשרות אחת: להגדיל את גודל המדגם כך ש- $1.96 \cdot \frac{7}{\sqrt{n}}$ יהיה שווה ל-1. במקרה זה ל-189 במקום 120.

אפשרות שניה: להוריד את רמת הסמך. אם למשל נרד מ-95% ל-90%, אז במקום 1.96 יהיה לנו 1.645 ואז שולי הטעות יהיו 1.051. צריך להוריד עוד קצת, ואפשר לחשב בדיוק כמה.

19

שאלה 5



חברת ביטוח מעוניינת לדעת איזה אחוז מהנהגים משתמשים בטלפון סלולרי בזמן הנהיגה. הוצע לקחת מדגם ולשאול כל נהג שעלה במדגם האם הוא משתמש בטלפון בזמן הנהיגה.

מהי ההתפלגות של מספר הנהגים במדגם הטוענים כי אינם משתמשים בטלפון בזמן הנהיגה?

בהנחה הסבירה שמדובר במדגם מייצג, זוהי התפלגות בינומית עם n השווה לגודל המדגם ו- p השווה לפרופורציה של הנהגים המשתמשים בטלפון הסלולרי בזמן הנהיגה מתוך אוכלוסיית הנהגים.

20

שאלה 5 סעיף ב



מהו גודל המדגם הנדרש כדי להבטיח ברמת סמך של 95% כי הטעות תהיה 3% לכל היותר?

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

זהו חישוב דומה לזה שנעשה בשאלה 1, סעיף ב. שולי הטעות הם:

מכיוון שמראש אנו לא יודעים מה תהיה הפרופורציה במדגם, ניקח במקומה את הערך גרוע ביותר האפשרי: 0.5. כאן אנו דורשים שולי הטעות המירביים יהיו שווים ל-0.03, כלומר

$$1.96 \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}} = 0.03$$

נפתור את המשוואה עבור n ונמצא כי גודל המדגם הדרוש הוא 1067

שאלה 5 סעיף ג



מנהלי החברה דורשים כי רמת הסמך תהיה 99% ולא 95%, אך אינם מוכנים להגדיל את תקציב המדגם. מה יהיו ההשלכות לגבי אורך רווח הסמך?

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

אורך רווח הסמך הוא בסופו של דבר

אם רוצים להגדיל את רמת הסמך, $z_{\alpha/2}$ יהיה גדול יותר, יותר – כי יהיה צריך להשאיר זנב קטן יותר.

אם גודל המדגם לא ישתנה וגם האמדן לפרופורציה לא ישתנה באופן משמעותי אז כל הביטוי הזה יגדל ולכן אורך רווח הסמך יגדל

שאלה 5 סעיף ד



בסופו של דבר הוחלט לקחת מדגם של 400 נהגים. מתוכם, 320 טענו כי לעולם אינם משתמשים בטלפון הסלולרי בזמן הנהיגה. חשבו את רווח הסמך ברמת הסמך של 95%.

הפרופורציה במדגם היא $0.8 = 320/400$. רווח הסמך הוא

$$0.8 \pm 1.96 \cdot \sqrt{\frac{0.8 \cdot 0.2}{400}} = (0.78, 0.82)$$

23

שאלה 6



חוקר רצה לבדוק איזה אחוז מבין הסטודנטים באוניברסיטה שוהים ברשתות החברתיות יותר מ-3 שעות ביום. החוקר משער כי נתון זה הוא לפחות 70%.

$$H_0: p = 0.7 \quad H_1: p \geq 0.7$$

נסחו את השערות של החוקר

החוקר דגם 60 סטודנטים מהפקולטה ושאל כל אחד מהם כמה שעות ביום הם מבליים ברשתות החברתיות. הוא השתמש בנתונים אלה כדי לבדוק השערות שלו, וקיבל ערך p-value השווה ל-0.0606.

מה רמת המובהקות של המבחן הסטטיסטי בו השתמש החוקר? הנתונים בשאלה אינם מספיקים

מה תהיה מסקנת החוקר אם רמת המובהקות היא 5%?

$0.05 < 0.0606$ ולכן החוקר אינו יכול לדחות את השערת האפס

24