

הערה כללית: ערכי  $Z_\alpha$  שייטכן ותידרשו לדעת את ערכם מפורטים בסיום התרגיל.

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%.
  - 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?
  - ב. אם רוצים לאמוד את פרופורציית המתחסנים בקרב כל עובדי מערכת הבריאות בערת רווח סמך ברמת סמך של 95%, מה יהיו שולי הטעות המירביים ("טעות הדגימה המירבית")?
  - ג. אם אנו רוצים לבנות רווח סמך לפרופורציה של עובדי מערכת הבריאות שהתחסנו, כך שאורך רווח הסמך יהיה 0.04, מהו גודל המדגם הדרוש?
  - ד. מתוך העובדים במקצועות הבריאות, 518 התחסנו נגד קורונה.
    - i. בנו רווח סמך לפרופורציה של המתחסנים מקרב העובדים במקצועות הבריאות.
    - ii. האם תדחו את ההשערה כי פרופורציה זו שונה מ-82%?
  - ה. מבין עובדי המנהל, 220 התחסנו נגד קורונה. בדקו את ההשערה כי שיעור המתחסנים בקרב עובדי מקצועות הבריאות גבוה משיעור המתחסנים עובדי המנהל.
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.
  - א. נסחו השערות המתאימות לבעיה, ונמקו מדוע אלה ההשערות שבחרתם.
  - ב. תארו את שני סוגי הטעויות האפשריות, והסבירו את המשמעות שלהן מנקודת המבט של החברה.
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.
    - א. התברר כי המספר הממוצע של ימי המחלה שנוצלו על ידי העובדים במדגם הוא 7.3 ימים. בנו את רווח הסמך שמציע חוקר א.
    - ב. על פי תשובתכם לסעיף א, מה ההסתברות כי התוחלת שווה ל-8 ימים?

- ג. נסחו את ההשערות שמציע החוקר השני. תארו את הטעויות האפשריות ואת ההשלכות שלהן, הן מבחינת המשרד והן מבחינת העובדים.
- ד. מה כלל ההחלטה הדרוש לבדיקת ההשערות של החוקר השני?
- ה. בהנחה שאחוז הניצול האמיתי הוא 20%, הסבירו כיצד תחשבו את העוצמה של המבחן. יש לפרט את החישובים עד לנקודה בה יש לחשב את ההסתברות שמשנתה מקרי נורמלי מקבל ערך מסויים (כלומר, עד שיש צורך לחשב הסתברות מהסוג  $P(Z < a)$  או  $P(Z > a)$ )
4. בארצות הברית, כל אזרח נדרש להגיש דו"ח שנתי למס הכנסה. מילוי הדו"ח מסובך, ואזרחים רבים מבלים 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.
- א. בנו רווח סמך לתוחלת משך השעות הנדרש כדי למלא את הדו"ח
- ב. מהם שולי הטעות של רווח הסמך?
- 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.
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.
- א. מהי ההתפלגות של מספר הנהגים במדגם הטוענים כי אינם משתמשים בטלפון בזמן הנהיגה?
- ב. מהו גודל המדגם הנדרש כדי להבטיח ברמת סמך של 95% כי הטעות תהיה 3% לכל היותר?
- 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?
- 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%.
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%.
- א. נסחו את ההשערות של החוקר
- ב. החוקר דגם 60 סטודנטים מהפקולטה ושאל כל אחד מהם כמה שעות ביום הם מבלים ברשתות החברתיות. הוא השתמש בנתונים אלה כדי לבדוק ההשערות שלו, וקיבל  $p\text{-value}=0.0606$ .
- i. מה רמת המובהקות של המבחן הסטטיסטי בו השתמש החוקר?
- ii. מה תהיה מסקנת החוקר אם רמת המובהקות היא 5%?

ערכי  $Z_\alpha$ :

$$Z_{0.2}=0.84, Z_{0.10}=1.282, Z_{0.05}=1.645, Z_{0.025}=1.96, Z_{0.01}=2.33$$