בדיקת השערות

The lady tasting tea

- השערת האפס: הליידי אינה יכולה להבחין בין טעם תה שנמזג לתוכו חלב ובין טעם תה שנוסף לחלב
- ניסוי: הליידי טעמה 8 ספלי תה, ולגבי כל אחד מהם חיווותה את דעתה לגבי אופן ההכנה

© Yossi Levy 2021

תהליך בדיקת ההשערות

- ניסוח השערת האפס: מבטאת בדרך כלל את הנחת היסוד, המצב הקיים או התיאוריה הקיימת
- ניסוח השערה אלטרנטיבית (אופציונלית): מבטאת את ההשערה שלנו לגבי המצב האמיתי
- ניסוח כלל החלטה
- איסוף הנתונים הדרושים לצורך קבלת ההחלטה
- החלטה האם **לדחות את השערת האפס** בהתאם **לכלל ההחלטה שקבענו מראש** ועל פי הנתונים שאספנו

דוגמה: ניסוי קליני בחולי סכרת

- אצל חולי סכרת, חשוב לשמור על איזון רמת הסוכר בדם. רמת הסוכר נמדדת על ידי מדד בשם HbA1c, וחולה נחשב ללא מאוזן אם רמה הסוכר גבוהה מ-7.5. רמת סוכר גבוהה מ-9 נחשבת למסוכנת.
- הטיפול הסטנדרטי לחולי סכרת הוא תרופה בשם metformin. אם הטיפול אינו מצליח לשמור על איזון רמת הסוכר בדם והיא עולה ל-8.5, או מתקרבת לערך זה, יש צורך בטיפול בתרופת קו שני. יש מגוון של תרופות כאלה. הן אינן יעילות עבור חלק מהחולים, ובדרך כלל תופעות הלוואי שלהן קשות יותר.
- ידוע כי לתוצאת המדידה של HbA1c יש התפלגות נורמלית עם סטיית תקן השווה ל-0.8.
- מוצעת תרופת קו שני חדשה, fixitall , שלטענת חברת התרופות יעילה יותר מהתרופות הקיימות. בעוד שתרופת קו שני מסויימת מביאה לאיזון סוכרתי של 30% מהחולים לאחר טיפול של 3 חודשים, נטען כי טיפול משולב של התרופה ופיקסיטול מביא לאיזון סוכרתי של 40% מהמטופלים.
- מוצע לבצע ניסוי קליני בו קבוצת חולים עבורם הטיפול במטפורמין נכשל, תטופל בטיפול משולב עם פיקסיטול למשך מעקב של שלושה חודשים.

© Yossi Levy 2021

www.sci-princess.info

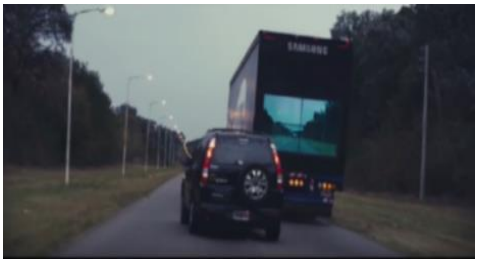
דוגמה: ניסוי קליני בחולי סכרת

- השערת האפס: $H_0: \quad p = 0.3$
- השערה אלטרנטיבית: $H_1: \quad p = 0.4$

הוחלט כי

- גודל המדגם יהיה 50 חולים
- השערת האפס תידחה אם פרופורציית החולים שיגיעו לאיזון סוכרתי תהיה לפחות 36%: $\hat{p} \geq 0.36$

שני סוגי טעות



© Yossi Levy 2021

www.sci-princess.info

שני סוגי טעות

- באשר נדרשים לקבל החלטה על סמך נתונים סטטיסטיים (שהתקבלו ממדגם או ניסוי), ניתן לחשב את ההסתברות כי תתרחש כל אחת מהטעויות
- ההסתברות לטעות מסוג ראשון מסומנת באות α.
- ההסתברות לטעות מסוג שני מסומנת באות β. מסיבות מתמטיות יותר נוח לדון בהסתברות שלא תהיה טעות מסוג שני, כלומר ב- $1-\beta$. ערך זה מכונה בשם העוצמה.

דוגמה: ניסוי קליני בחולי סכרת

- טעות מסוג ראשון: הטיפול המשולב עם פיקסיטול נמצא יעיל יותר בעוד שאינו כזה
- המשמעות: עד שהנתונים שיצטברו יראו כי אכן הטיפול המשולב אינו יעיל יותר, עשרות אלפי, ואולי מאות אלפי חולים יקבלו טיפול לא יעיל אשר עלול לסכן את בריאותם
- טעות מסוג שני: הטיפול המשולב עם פיקסיטול אמנם יעיל יותר אך הוא לא ייכנס לשימוש
- עשרות אלפי, ואולי מאות אלפי חולים שהטיפול המשולב עם פיקסיטול היה עשוי לשפר את מצבם הבריאותי לא יזכו לקבל אותו

© Yossi Levy 2021

www.sci-princess.info

פיקסיטול: ההסתברות לטעות מסוג ראשון

- טעות מסוג ראשון מתרחשת רק כאשר השערת האפס נכונה, כלומר כאשר $p = 0.3$
- במקרה זה, התוחלת והשונות של הפרופורציה במדגם הן: $E(\hat{p}) = p = 0.3$

$$V(\hat{p}) = \frac{p(1-p)}{n} = \frac{0.3 \cdot 0.7}{50}$$

- הטעות מתרחשת כאשר הפרופורציה במדגם שווה ל- 0.36 או יותר

$$\alpha = P_{H_o}(\hat{p} \geq 0.36) = P_{H_o}\left(\frac{\hat{p} - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{50}}} \geq \frac{0.36 - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{50}}}\right) = P_{H_o}(Z \geq 0.926) = 0.177$$

פיקסיטול: ההסתברות לטעות מסוג שני

- טעות מסוג שני מתרחשת רק כאשר ההשערה האלטרנטיבית נכונה, כלומר כאשר $p = 0.4$
- במקרה זה, התוחלת והשונות של הפרופורציה במדגם הן: $E(\hat{p}) = p = 0.4$

$$V(\hat{p}) = \frac{p(1-p)}{n} = \frac{0.4 \cdot 0.6}{n}$$

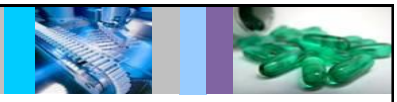
- הטעות מתרחשת כאשר הפרופורציה במדגם קטנה מ- 0.36

$$\beta = P_{H_1}(\hat{p} < 0.36) = P_{H_1}\left(\frac{\hat{p} - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{50}}} < \frac{0.36 - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{50}}}\right) = P(Z < -0.577) = 0.282$$

© Yossi Levy 2021

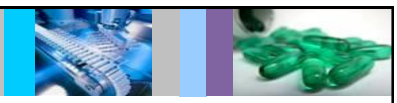
www.sci-princess.info

אסטרגיות לבדיקת השערות



- **כלל ההחלטה, גודל המדגם וההסתברויות לטעות תלויים זה בזה.**
- **יש הרבה כללי החלטה, אבל תמיד יש כלל החלטה שהוא הטוב ביותר:**
  - **אם קובעים את רמת המובהקות – כלל זה ייתן עוצמה מקסימלית**
- **לכן המשחק הוא בין רמת המובהקות, העצמה וגודל המדגם**
  - **אתם יכולים לבחור שניים, והערך של השלישי ייקבע על ידי הערכים הראשונים שבחרתם**

אסטרטגיה לבדיקת השערות



- נסחו את ההשערות ובחרו את כלל ההחלטה האופטימלי
- קבעו את הערך הנסבל הגבוה ביותר של הטעות מסוג ראשון. ערך זה נקרא רמת המובהקות של המבחן. מסיבות היסטוריות, הערך המקובל של רמת המובהקות הוא 0.05.
- מצאו הערכה להתפלגות/שונות
- חשבו את העצמה עבור ערך נתון של גודל המדגם, או את גודל המדגם הנדרש עבור ערך נתון של עצמה

© Yossi Levy 2021

www.sci-princess.info

ניסוי הפיקסיטול

- גודל המדגם היה 50 משתתפים
- ההשערות היו: $H_0: \ p = 0.3$ $H_1: \ p = 0.4$
- כלל ההחלטה היה: דחה את השערת האפס אם הפרופורציה במדגם היא לפחות 0.36
- חישבנו את ההסתברויות לשני סוגי הטעויות ומצאנו כי
  - ההסתברות לדחיה מוטעית של השערת האפס היא: $\alpha=0.177\approx18\%$
  - ההסתברות לאי-דחיה מוטעית של השערת האפס היא $\beta=0.282\approx28\%$
- ערכי טעויות אלה הינם בלתי קבילים כמעט בכל יישום

דוגמה: ניסוי הפיקסיטול – חשיבה מחדש

$$H_0: \ p = 0.3 \qquad H_1: \ p = 0.4$$

- כלל החלטה ראשוני: דחו את השערת האפס אם הפרופורציה במדגם גדולה מ-0.3 **באופן מובהק**
- כלל החלטה ראשוני: דחו את השערת האפס אם $\hat{p} > 0.3 + \Delta$
- הערך של Δ נקבע על ידי ההשערות, הנחות ההתפלגות, רמת המובהקות, העצמה, וגודל המדגם
- ההשערות והנחות ההתפלגות ידועות. נקבע את רמת המובהקות ל-$\alpha=0.05$
- איזה גודל מדגם דרוש כדי להבטיח עוצמה של $1-\beta=90\%$?

© Yossi Levy 2021

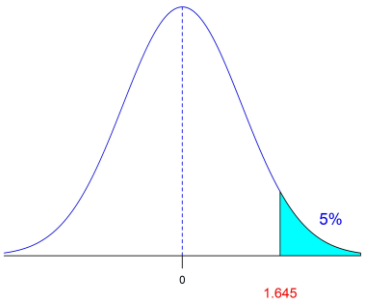
www.sci-princess.info

פיקסיטול: קביעת כלל ההחלטה – רמת המובהקות

אנו רוצים כי רמת המובהקות לא תעלה על α=0.05:

$$\alpha = P_{H_0}(\hat{p} > 0.3 + \Delta) = 0.05$$

$$\alpha = P_{H_0}(\hat{p} > 0.3 + \Delta) = P_{H_0}\left(\frac{\hat{p} - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{n}}} > \frac{0.3 + \Delta - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{n}}}\right) = P_{H_0}\left(Z > \frac{\Delta}{\sqrt{\frac{0.3 \cdot 0.7}{n}}}\right) = 0.05$$



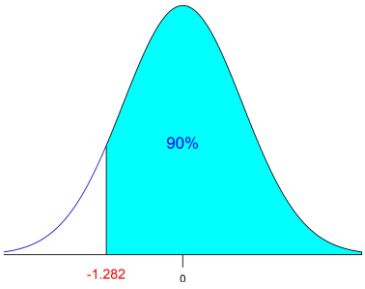
$$\frac{\Delta}{\sqrt{\frac{0.3 \cdot 0.7}{n}}} = 1.645$$

פיקסיטול: קביעת כלל ההחלטה – העוצמה

אנו רוצים כי עצמת המבחן תהיה לפחות 1-β=0.9:

$$\beta = P_{H_1}(\hat{p} \le 0.3 + \Delta) = 0.1$$

$$1 - \beta = P_{H_1}(\hat{p} > 0.3 + \Delta) = P_{H_1}\left(\frac{\hat{p} - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{n}}} > \frac{0.3 + \Delta - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{n}}}\right) = P_{H_1}\left(Z > \frac{\Delta - 0.1}{\sqrt{\frac{0.4 \cdot 0.6}{n}}}\right) = 0.9$$



$$\frac{\Delta - 0.1}{\sqrt{\frac{0.4 \cdot 0.6}{n}}} = -1.282$$

© Yossi Levy 2021

www.sci-princess.info

פיקסיטול: כלל ההחלטה וגודל המדגם

מצאנו שתי משוואות:

$$\frac{\Delta - 0.1}{\sqrt{\frac{0.4 \cdot 0.6}{n}}} = -1.282 \qquad \frac{\Delta}{\sqrt{\frac{0.3 \cdot 0.7}{n}}} = 1.645$$

נפתור אותן ונקבל:

$$\Delta = 0.055 \quad n = 190.96$$

סיכום: כדי לשמור על רמת מובהקות של 5% ולהבטיח עוצמה של 90% יש צורך במדגם בגודל 191 משתתפים ולדחות את השערת האפס אם הפרופורציה במדגם גדולה מ-0.355

החדשות הטובות: לא צריך לעשות את כל התרגיל זה כל פעם מחדש ☺

סיכום

- כלל ההחלטה היה לדחות את השערת האפס כאשר הפרופורציה במדגם גדולה מ-$0.3+\Delta$
- אם נחזור על כל החישוב הנ"ל בנוסחאות בלבד, התוצעה תהיה כי כלל ההחלטה הוא לדחות את השערת האפס כאשר

$$\hat{p} > p_0 + z_\alpha \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

© Yossi Levy 2021

פיקסיטול: מה שכחנו?

בדקנו את ההשערה כי טיפול משולב מעלה את שיעור החולים המאוזנים מ-0.3 ל-0.4

- אולי הטיפול המשולב מעלה את שיעור החולים המאוזנים לערך אחר ולא דווקא ל-0.4?

$H_0: p = 0.3 \;\; H_1: p > 0.3$

- אולי הטיפול המשולב דווקא מרע את ההשפעה של הטיפול המקובל?

$H_0: p = 0.3 \;\; H_1: p \neq 0.3$

פיקסיטול: מה שכחנו?

- אולי יש צורך לבדוק את השינוי ברמת ה- HbA1c ולא את שיעור החולים המאוזנים?

$H_0: \mu_{A1c} = 8.5 \quad H_1: \mu_{A1c} = 7$

$H_0: \mu_{A1c} = 8.5 \quad H_1: \mu_{A1c} \leq 7$

$H_0: \mu_{A1c} = 8.5 \quad H_1: \mu_{A1c} \neq 8.5$

www.sci-princess.info

© Yossi Levy 2021

פיקסיטול: מה שכחנו?

- אולי יש צורך להשוות בין קבוצת חולים שמטופלת רק בטיפול המקובל לקבוצה אחרת של חולים שתטופל בטיפול המשולב?

$H_0: p_1 = p_2 \;\; H_1: p_1 > p_2$ $\qquad$ $H_0: p_1 = p_2 \;\; H_1: p_1 \neq p_2$

$H_0: \mu_1 = \mu_2 \;\; H_1: \mu_1 > \mu_2$ $\qquad$ $H_0: \mu_1 = \mu_2 \;\; H_1: \mu_1 \neq \mu_2$

$H_0: \mu_1 \leq \mu_2 \;\; H_1: \mu_1 > \mu_2$

סיכום עד כה

- שתי השערות
  - השערת האפס מייצגת את הידע הקיים
  - ההשערה האלטרנטיבית מייצגת תיאוריה חדשה שאולי תהפוך לידע חדש
- דוחים את השערת האפס אם על פי **כלל החלטה שנקבע מראש** הנתונים אינם מתיישבים עם השערת האפס **באופן מובהק**
- כלל ההחלטה מכונה בשם "מבחן סטטיסטי"

© Yossi Levy 2021

סיכום עד כה: השערות חד צדדיות על הפרופורציה באוכלוסייה

$$H_0: p = p_0 \quad H_1: p > p_0$$

כלל החלטה אופטימלי: דחה את השערת האפס כאשר הפרופורציה במדגם גדולה מ- p_0 **באופן מובהק,** כלומר כאשר

$$\hat{p} > p_0 + z_\alpha \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$H_0: p = p_0 \quad H_1: p < p_0$$

כלל החלטה אופטימלי: דחה את השערת האפס כאשר הפרופורציה במדגם קטנה מ- p_0 **באופן מובהק,** כלומר כאשר

$$\hat{p} < p_0 - z_\alpha \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

סיכום עד כה: השערות חד צדדיות על התוחלת של התפלגות נורמלית באשר השונות ידועה

$$H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0$$

כלל החלטה אופטימלי: דחה את השערת האפס כאשר ממוצע המדגם גדול מ- μ_0 **באופן מובהק,** כלומר כאשר

$$\bar{X}_n > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

$$H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0$$

כלל החלטה אופטימלי: דחה את השערת האפס כאשר ממוצע המדגם קטן מ- μ_0 **באופן מובהק,** כלומר כאשר

$$\bar{X}_n < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

© Yossi Levy 2021
www.sci-princess.info

המשך דוגמא: ניסוי הפיקסיטול

$H_0: \mu = \mu_0 = 8.5 \qquad H_1: \mu < 8.5$

כלל החלטה אופטימלי: דחה את השערת האפס כאשר ממוצע המדגם קטן מ-8.5 **באופן מובהק**

תזכורת: לערכי ההמוגלובין יש התפלגות נורמלית וסטיית התקן של מדידת ההמוגלובין היא 0.8

מחליטים לקחת מדגם של 30 חולים. מה יהיה כלל ההחלטה עבור רמת מובהקות של 5%?

נדחה את השערת האפס אם:

$$\bar{X}_n < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 8.5 - 1.645 \cdot \frac{0.8}{\sqrt{30}} = 8.26$$

המשך דוגמא: ניסוי הפיקסיטול

$H_0: \mu = \mu_0 = 8.5 \qquad H_1: \mu < 8.5$

כלל ההחלטה אופטימלי: דחה את השערת האפס כאשר ממוצע המדגם קטן מ-8.26. מהי עצמת המבחן?

העצמה תלויה בערך של התוחלת כאשר ההשערה האלטרנטיבית נכונה.

יש הרבה מספרים שקטנים מ-8.5. החישוב זהה לכולם, אבל לכל ערך נקבל תוצאה אחרת

נראה את החישוב עבור $\mu_1 = 8$

$$1 - \beta = P_{H_1}(\bar{X}_n < 8.26) = P\left(\frac{\bar{X}_n - 8}{0.8/\sqrt{30}} < \frac{8.26 - 8}{0.8/\sqrt{30}}\right) = P(Z < 1.78) = 0.96$$

© Yossi Levy 2021

www.sci-princess.info

השערה דו-צדדית על פרופורציה באוכלוסייה

$H_0: p = p_0 \quad H_1: p \neq p_0$

נדחה את השערת האפס אם הפרופורציה במדגם שונה מ- p_0 **באופן מובהק**

זה יכול לקרות גם כאשר הפרופורציה במדגם קטנה מ- p_0 **באופן מובהק** וגם כאשר היא גדולה ממנו **באופן מובהק**

כדי לשמור על רמת מובהקות כוללת α, נחלק אותה בין שני הזנבות באופן שווה

נדחה את השערת האפס אם

$$\hat{p} > p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \quad \text{או אם} \quad \hat{p} < p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

הקשר בין בדיקת השערות לרווח סמך

$H_0: p = p_0 \quad H_1: p \neq p_0$

נדחה את השערת האפס אם

$$\hat{p} > p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \quad \text{או אם} \quad \hat{p} < p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

לא נדחה את השערת האפס אם $p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$

נדחה את השערת האפס אם רווח הסמך לפרופורציה לא מכיל את p_0 !

© Yossi Levy 2021

www.sci-princess.info

המשך דוגמה: פיקסיטול

נלקח מדגם בגודל של 30 מטופלים. התברר כי רמת הסוכר בדם התייצבה אצל 32% מהמטופלים. האם נדחה את השערת האפס?

נדחה את השערת האפס אם הפרופורציה במדגם גדולה מ-

$$p_0 + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.3 + 1.96 \cdot \sqrt{\frac{0.32 \cdot 0.68}{30}} = 0.47$$

0.32 לא גדול מ-0.47 ולכן לא נדחה את השערת האפס*

אלטרנטיבית: רווח הסמך לפרופורציה באוכלוסייה הוא

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.32 \pm 1.96 \cdot \sqrt{\frac{0.32 \cdot 0.68}{30}} = (0.153, 0.487)$$

רווח הסמך מכיל את 0.3 ולכן לא נדחה את השערת האפס

השערה דו-צדדית על התוחלת של התפלגות נורמלית כאשר השונות ידועה

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

נדחה את השערת האפס אם ממוצע המדגם שונה מ- μ_0 **באופן מובהק**

נדחה את השערת האפס אם

$$\bar{X}_n > \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{או אם} \quad \bar{X}_n < \mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

לא נדחה את השערת האפס אם

$$\bar{X}_n - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{X}_n + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

© Yossi Levy 2021

www.sci-princess.info

המשך דוגמא: ניסוי הפיקסיטול

$H_0: \mu = \mu_0 = 8.5 \qquad H_1: \mu < 8.5$

תזכורת: לערכי ההמוגלובין יש התפלגות נורמלית וסטיית התקן של מדידת ההמוגלובין היא 0.8

במדגם של 30 החולים, מה יהיה כלל ההחלטה עבור רמת מובהקות של 5%?

נדחה את השערת האפס אם ממוצע המדגם יהיה גדול מ-: $\mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 8.5 + 1.96 \cdot \frac{0.8}{\sqrt{30}} = 8.79$

או אם ממוצע המדגם יהיה קטן מ-8.21 (למה?)

אלטרנטיבית, נדחה את השערת האפס אם רווח הסמך לתוחלת לא יכיל את 8.5

בסיום הניסוי התברר כי ממוצע ההמוגלובין בקרב המטופלים היה 7.9. מה תהיה ההחלטה?

תרגיל: חשבו את רווח הסמך לתוחלת ובדקו כי תתקבל אותה ההחלטה

מה משפיע על העצמה של המבחן הסטטיסטי?

- כזכור, אמרנו כי גודל המדגם וההסתברויות לטעויות תלויים זה בזה
- מכאן שגודל המדגם ורמת המובהקות משפיעים על העצמה של המבחן כאשר משתמשים בכלל ההחלטה האופטימלי
- אבל יש עוד שני גורמים שצריך לקחת בחשבון:
  - השונות
  - גודל האפקט = ההפרש בין ערך הפרמטר תחת השערת האפס וערכו תחת ההשערה האלטרנטיבית
- ניתן כמובן לפתח נוסחאות לכל הדברים האלה, אבל אנחנו נסתפק בדוגמה מספרית

© Yossi Levy 2021

www.sci-princess.info

דוגמה: בדיקת השערה פשוטה על התוחלת

- מהנדס כימיה רוצה לבדוק אם שינוי בטמפרטורה של תהליך כלשהו מעלה את התפוקה של התהליך
- ידוע כי התפלגות התפוקה בתהליך הייצור הרגיל היא נורמלית, עם תוחלת השווה ל-300 גרם, וסטיית התקן שווה ל-40 גרם
- המהנדס משער כי שינוי הטמפרטורה יעלה את תוחלת התפוקה. כן הוא מעריך כי סטיית התקן לא תשתנה עם שינוי הטמפרטורה.
- המהנדס עורך 10 ניסויים עם שינוי בטמפרטורה, ומעוניין לבדוק את השערתו ברמת מובהקות של 5%.
- התפוקה הממוצעת בניסויים שערך הייתה 325 גרם. האם התהליך החדש יעיל יותר?

דוגמה: שיפור תהליך

ההשערות : $H_0: \mu = \mu_0 = 300$ $H_1: \mu = \mu_1 > 300$

כלל החלטה/מבחן סטטיסטי: דוחים את העשרת האפס אם $\bar{X}_n > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

העצמה שווה ל- $1 - \beta = P_{H_1}\left(\bar{X}_n > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$

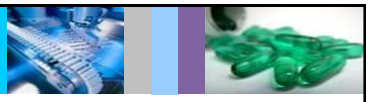
הפרמטרים שמשפיעים על גובה העצמה הם: α, σ, n, μ_1 וכמובן μ_0

ההפרש בין μ_1 ו- μ_0 נקרא בשם "גודל האפקט".

© Yossi Levy 2021

www.sci-princess.info

שיפור התהליך: בדיקת ההשערה



דוחים את השערת האפס כאשר $\bar{X}_n > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

באן: $\mu_0 = 300 \quad \sigma = 50 \quad n = 10 \quad \alpha = 0.05 \quad Z_\alpha = 1.645$

ולבן דוחים את השערת האפס כאשר $\bar{X}_n > 300 + 1.645 \cdot \frac{50}{\sqrt{10}} = 326$

שיפור התהליך: חישוב העצמה

אזהרה: את חישובי העצמה עורכים בשלב תכנון הניסוי, ולא לאחר ביצועו. זו דוגמה בלבד!

שימו לב כי אנו לא צריכים לדעת את תוצאת הניסוי כדי לחשב את העצמה.

בזכור: $\mu_0 = 300 \quad \sigma = 50 \quad n = 10 \quad \alpha = 0.05 \quad Z_\alpha = 1.645$

נחשב את העצמה עבור $\mu_1 = 330$, כלומר עבור גודל אפקט של 30

$$1 - \beta = P_{H_1}(\bar{X}_n > 326) = P_{H_1}\left(\frac{\bar{X}_n - 330}{50/\sqrt{10}} > \frac{326 - 330}{50/\sqrt{10}}\right)$$

$$= P(Z > -0.88) = 0.812$$

© Yossi Levy 2021

www.sci-princess.info

השפעת רמת המובהקות על העצמה

מה יקרה אם נעלה את רמת המובהקות ל-α=10?

$$\alpha = 0.1 \rightarrow z_\alpha = 1.28$$

לכן נדחה את השערת האפס אם ממוצע המדגם יהיה גדול מ-

$$\mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 300 + 1.28 \cdot \frac{50}{\sqrt{10}} = 320.2$$

ולכן העצמה שווה ל:

$$1 - \beta = P_{H_1}(\bar{X}_n > 320.2) = P_{H_1}\left(\frac{\bar{X}_n - 330}{50/\sqrt{10}} > \frac{320.2 - 330}{50/\sqrt{10}}\right)$$

$$= P(Z > -1.25) = 0.894$$

העלאת רמת המובהקות הגדילה את העצמה

השפעת גודל המדגם על העצמה

מה היה קורה לו גודל המדגם היה 20 במקום 10?

במקרה זה היינו דוחים את השערת האפס אם ממוצע המדגם היה גדול מ-

$$\mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 300 + 1.645 \cdot \frac{50}{\sqrt{20}} = 318.4$$

ואז העצמה הייתה שווה ל-

$$1 - \beta = P_{H_1}(\bar{X}_n > 318.4) = P_{H_1}\left(\frac{\bar{X}_n - 330}{50/\sqrt{20}} > \frac{318.4 - 330}{50/\sqrt{20}}\right)$$

$$= P(Z > -1.93) = 0.973$$

כאשר המדגם גדול יותר העצמה גבוהה יותר

© Yossi Levy 2021

www.sci-princess.info

השפעת השונות על העצמה

מה היה קורה לו סטיית התקן הייתה שווה ל-30 במקום ל-50?

במקרה זה היינו דוחים את השערת האפס אם ממוצע המדגם היה גדול מ-

$$\mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 300 + 1.645 \cdot \frac{30}{\sqrt{10}} = 315.6$$

ואז העצמה הייתה שווה ל-

$$1 - \beta = P_{H_1}(\bar{X}_n > 315.6) = P_{H_1}\left(\frac{\bar{X}_n - 330}{30/\sqrt{10}} > \frac{315.6 - 330}{30/\sqrt{10}}\right)$$

$$= P(Z > -2.57) = 0.995$$

כאשר השונות נמוכה יותר העצמה גבוהה יותר

השפעת גודל האפקט על העצמה

מהי העצמה עבור גודל אפקט של 60 במקום 30, כלומר עבור $\mu_1 = 360$?

אנו עדיין דוחים את השערת האפס אם ממוצע המדגם גדול מ-326

במקרה זה העצמה שווה ל-

$$1 - \beta = P_{H_1}(\bar{X}_n > 326) = P_{H_1}\left(\frac{\bar{X}_n - 360}{50/\sqrt{10}} > \frac{326 - 360}{50/\sqrt{10}}\right)$$

$$= P(Z > -2.15) = 0.984$$

המשמעות: יותר קל לדחות את השערת האפס אם האפקט יותר גדול

© Yossi Levy 2021

www.sci-princess.info

השוואה בין אוכלוסיות

- בדוגמת הסכרת הייתה לנו רק קבוצה אחת של חולים: כאלה שקיבלו את הטיפול המשולב.
- גם אם היה שיפור במצבם של החולים האלה, איך אנחנו יכולים לדעת שהגורם לשיפור הוא הטיפול המשולב?
- פתרון אפשרי: לקחת שתי קבוצות של חולים: אחת מהן תקבל את הטיפול הסטנדרטי, והשניה את הטיפול המשולב
- בסיום תקופת הטיפול נוכל להשוות בין
  - הפרופורציות של מספר החולים שמצבם השתפר בכל אחת מהקבוצות
  - התוחלות של רמת ההמוגלובין בכל אחת מהקבוצות
- כל זאת בהנחה שאין תלות בין שתי הקבוצות!

השוואת פרופורציות

- בטיפול הסטנדרטי , אחוז החולים שמצבם משתפר פחות או יותר ידוע, והוא שווה בערך ל-30%
- ההשערה היא כי הטיפול המשולב מעלה את הפרופורציה של החולים שמצבם השתפר
- נסמן ב-p_1 את הפרופורציה של החולים שמצבם משתפר בטיפול הסטנדרטי, וב-p_2 את הפרופורציה של החולים שמצבם משתפר לאחר טיפול משולב
- ההשערות הן לכן: $H_0: p_1 = p_2 \quad H_1: p_1 < p_2$
- נדחה את השערת האפס רק אם **הפרופורציה של מספר החולים בניסוי שמצבם השתפר לאחר טיפול משולב גדולה באופן מובהק מהפרופורציה של מספר החולים בניסוי שמצבם השתפר לאחר טיפול סטנדרטי**
- במילים אחרות: נדחה את השערת האפס אם הפרש הפרופורציות גדול מאפס באופן מובהק

© Yossi Levy 2021
www.sci-princess.info

השוואת פרופורציות

- ההשערות: $H_0: p_1 = p_2 \quad H_1: p_1 < p_2$
- דוחים את השערת האפס אם $\hat{p}_2 - \hat{p}_1$ גדול מאפס באופן מובהק
- לכן אנו צריכים לחשב את התוחלת וסטיית התקן של היצור הזה:

$$E(\hat{p}_1) = p_1 \qquad E(\hat{p}_2) = p_2 \qquad E(\hat{p}_2 - \hat{p}_1) = p_2 - p_1$$

- תחת השערת האפס תוחלת זו שווה לאפס

השוואת פרופורציות

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

$$V(\hat{p}_2 - \hat{p}_1) = V[\hat{p}_2 + (-\hat{p}_1)] = V[\hat{p}_2 + (-\hat{p}_1)]$$

$$= V(\hat{p}_2) + V(-\hat{p}_1) = V(\hat{p}_2) + V(\hat{p}_1)$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

www.sci-princess.info

© Yossi Levy 2021

השוואת פרופורציות

- נדחה לכן את השערת האפס באשר

$$\hat{p}_2 - \hat{p}_1 > 0 + z_\alpha \cdot SD(\hat{p}_2 - \hat{p}_1) \quad = z_\alpha \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

המשך הדוגמה

$$n_1 = 29 \;\; n_2 = 31 \quad \hat{p}_1 = 0.33 \quad \hat{p}_2 = 0.39 \qquad \hat{p}_2 - \hat{p}_1 = 0.06$$

$$SD(\hat{p}_2 - \hat{p}_1) = \sqrt{\frac{0.33 \cdot 0.67}{29} + \frac{0.39 \cdot 0.61}{31}} = 0.124$$

$$z_\alpha \cdot SD(\hat{p}_1 - \hat{p}_2) = 1.645 \cdot 0.124 = 0.203$$

© Yossi Levy 2021

www.sci-princess.info

השוואת פרופורציות – השערה דו צדדית

- ההשערות $H_0: p_1 = p_2 \quad H_1: p_1 \neq p_2$
- נדחה את השערת האפס אם $\hat{p}_2 - \hat{p}_1$ שונה מאפס באופן מובהק
- כלומר: נדחה את השערת האפס אם $\hat{p}_2 - \hat{p}_1$ גדול מאפס באופן מובהק או אם $\hat{p}_2 - \hat{p}_1$ קטן מאפס באופן מובהק
- כלומר אם

$$\hat{p}_2 - \hat{p}_1 > z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

או

$$\hat{p}_2 - \hat{p}_1 < -z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

בדוגמה שלנו:

$$z_{\alpha/2} \cdot SD(\hat{p}_1 - \hat{p}_2) = 1.96 \cdot 0.124 = 0.242$$

השוואת תוחלות: השערה חד צדדית

- ההשערות הן: $H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 > \mu_2$
- נסמן ב-X את התצפיות מהקבוצה הראשונה וב-Y את התצפיות מהקבוצה השניה
- נדחה את השערת האפס אם הממוצע של הקבוצה הראשונה גדול מהממוצע של הקבוצה השניה באופן מובהק. באופן שקול: אם הפרש הממוצעים $\bar{X}_{n_1} - \bar{Y}_{n_2}$ גדול מאפס באופן מובהק:
- קל לחשב כי: $E\left(\bar{X}_{n_1} - \bar{Y}_{n_2}\right) = \mu_1 - \mu_2$ ותחת השערת האפס תוחלת זו שווה לאפס

© Yossi Levy 2021
www.sci-princess.info

השוואת תוחלות: חישוב סטיית התקן

- אנו צריכים לחשב את $SD\left(\bar{X}_{n_1} - \bar{Y}_{n_2}\right)$
- ידוע כי $V\left(\bar{X}_{n_1}\right) = \frac{\sigma_1^2}{n_1} \quad V\left(\bar{Y}_{n_2}\right) = \frac{\sigma_2^2}{n_2}$
- ובמקודם $V\left(\bar{X}_{n_1} - \bar{Y}_{n_2}\right) = V\left(\bar{X}_{n_1}\right) + V\left(\bar{Y}_{n_2}\right)$
- ולכן $SD\left(\bar{X}_{n_1} - \bar{Y}_{n_2}\right) = \sqrt{V\left(\bar{X}_{n_1}\right) + V\left(\bar{Y}_{n_2}\right)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

השוואת תוחלות

- בסופו של דבר נדחה את השערת האפס אם $\bar{X}_{n_1} - \bar{Y}_{n_2} > 0 + z_\alpha \cdot SD\left(\bar{X}_{n_1} - \bar{Y}_{n_2}\right)$

www.sci-princess.info
© Yossi Levy 2021

השוואת תוחלות – השערה דו צדדית

- ההשערות הן $H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$
- נדחה את השערת האפס אם הפרש הממוצעים שונה מאפס באופן מובהק
- כמקודם, נדחה את השערת האפס אם

$$\bar{X}_{n_1} - \bar{Y}_{n_2} > z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{או} \quad \bar{X}_{n_1} - \bar{Y}_{n_2} < -z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

דוגמה: ניסוי הפיקסיטול

- היו שתי קבוצות:29 חולים קיבלו טיפול משולב, ו-31 קיבלו טיפול סטנדרטי
- סטיית התקן של רצת ההמוגלובין בשתי הקבוצות שווה ל-0.8
- בסוף התקופה ממוצע ההמוגלובין בקבוצת הטיפול המשולב היה 7.8 ובקבוצת הטיפול הסטנדרטי הוא היה 8.3
- סטיית התקן של הפרש הממוצעים היא $\sqrt{\frac{0.8^2}{29} + \frac{0.8^2}{31}} = 0.21$
- ולכן נדחה את ההשערה החד צדדית אם הפרש הממוצעים גדול מ-

$$1.645 \cdot \sqrt{\frac{0.8^2}{29} + \frac{0.8^2}{31}} = 1.645 \cdot 0.21 = 0.34$$

- האם נדחה את ההשערה הדו-צדדית?

© Yossi Levy 2021
www.sci-princess.info

מדגמים מזווגים

- בדוגמה של ניסוי הפיקסיטול הסתכלנו על רמת ההמוגלובין בסוף תקופת הטיפול
- הבעיה: לא לקחנו בחשבון את רמת ההמוגלובין בתחילת הטיפול
- מתקבל על הדעת כי מצבם של חולים שהתחילו ברמת המוגלובין גבוהה במיוחד אכן ישתפר, אבל עדיין יישארו עם רמת המוגלובין גבוהה.
- פתרון: להסתכל על השינוי של כל חולה ברמת ההמוגלובין מתחילת תקופת המעקב לסופה
- במצב כזה יש לכל חולה שתי תצפיות: רמת ההמוגלובין בתחילת המעקב ורמת ההמוגלובין בסוף המעקב

מדגמים מזווגים: דוגמה

	month0 - X	month6 - Y	diff Z=X-Y
1	8.50	7.59	-0.91
2	8.50	7.66	-0.84
3	9.82	9.69	-0.13
4	8.50	8.16	-0.34
5	8.50	8.20	-0.30
6	10.02	9.86	-0.16
7	8.83	8.62	-0.21
8	8.50	6.76	-1.74
9	8.50	7.35	-1.15
10	8.50	7.69	-0.81
11	9.85	8.99	-0.86
12	8.90	8.07	-0.82
13	9.17	8.21	-0.97
14	8.50	8.52	0.02
15	8.50	7.66	-0.84
16	10.49	9.59	-0.90

- ממוצעים
  - חודש 0: 8.87
  - חודש 6: 8.0
  - הפרש: 0.87-
- שונויות
  - חודש 0: 0.39
  - חודש 6: 0.92
  - הפרש: 0.37

© Yossi Levy 2021

מדגמים מזווגים - פתרון

- נחשב לכל חולה את ההפרש ונבדוק השערה על ההפרש
- נדחה את השערת האפס אם הממוצע של Z קטן מאפס באופן מובהק
- בדוגמה שלנו: סטיית התקן של Z היא 0.61
- לכן נדחה את השערת האפס אם הממוצע של Z קטן מ- $-1.645 \cdot 0.61 = -1.003$

- האם נדחה את ההשערה הדו צדדית?

The lady tasting tea

- השערת האפס: הליידי אינה יכולה להבחין בין טעם תה שנמזג לתוכו חלב ובין טעם תה שנוסף לחלב
- ניסוי: הליידי טעמה 8 ספלי תה, ולגבי כל אחד מהם חיוותה את דעתה לגבי אופן ההכנה

© Yossi Levy 2021

www.sci-princess.info

בדיקת השערות – p-value

- תזכורת – ניסוי הליידי הטועמת תה
- 10 ספלים" תה עם חלב או חלב עם תה?
- השערת האפס: הליידי לא יודעת להבחין בין הטעמים

- כלל החלטה – נדחה את השערת האפס אם הליידי תזהה נכונה 8 ספלים לפחות

- אם נדחה את השערת האפס כאשר הליידי מזהה נכונה 8 ספלים, בוודאי נדחה אותה אם תזהה 9 ספלים או 10 ספלים.

בדיקת השערות – p-value

p

p

p

p

p

p

© Yossi Levy 2021

בחזרה אל הליידי

$H_0: p = 0.5 \quad X \sim B(10,0.5) \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

$P(X = 7) = 0.1172 \cdot (0.5)^3 = 0.1172 \quad P(X \geq 7) = 0.1719$

$P(X = 8) = 0.0439 \quad P(X \geq 8) = 0.0547$

$P(X = 9) = 0.0098 \quad P(X \geq 9) = 0.0107$

$P(X = 10) = 0.0010 \quad P(X \geq 10) = 0.0010$

חישוב הפי-ואליו - פיקסיטול

$H_0: p = 0.3 \quad H_1: p > 0.3$

- תזכורת: דוחים את השערת האפס כאשר הפרופורציה במדגם גדולה מ-0.3 באופן מובהק
- מבין 50 משתתפים בניסוי, 17 הגיעו לאיזון סוכרתי. מהו הפי-ואליו?

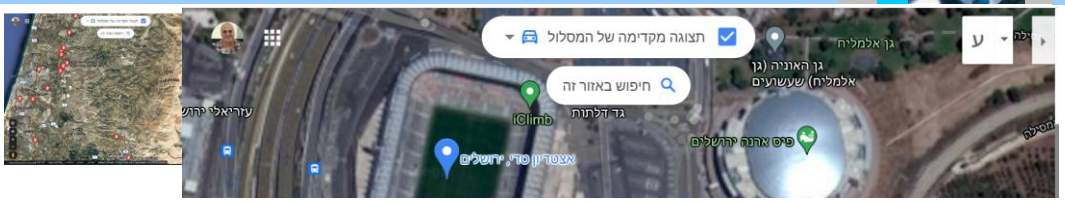
$$p - value = P_{H_0}(\hat{p} \geq 0.34)$$

$$= P_{H_0}\left(\frac{\hat{p} - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{50}}} \geq \frac{0.34 - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{50}}}\right) = P(Z \geq 0.617) = 0.2685$$

© Yossi Levy 2021

www.sci-princess.info

המשמעות המספרית של הפי-ואליו



אין משמעות לערך המספרי של הפי-ואליו!

הפי-ואליו אינו...

- הפי-ואליו אינו ההסתברות כי השערת האפס נכונה
- הפי-ואליו אינו ההסתברות כי תוצאות הניסוי התקבלו "במקרה" ("by chance")
- הפי-ואליו אינה ההסתברות כי בחזרה על הניסוי תתקבל אותה תוצאה כמותית
- הפי-ואליו אינו ההסתברות כי בחזרה על הניסוי תתקבל אותה תוצאה איכותית
- הפי-ואליו אינו ההסתברות לדחייה מוטעית של השערת האפס
- פי-ואליו הקרוב בערכו לרמת המובהקות אינו מעמיד בספק את ההחלטה המתקבלת
- פי-ואליו שערכו נמוך במיוחד (או גבוה במיוחד) אינו מחזק יותר את ההחלטה המתקבלת
- ערכו של הפי-ואליו אינו מעיד על גודל האפקט או חשיבותו

© Yossi Levy 2021

תוצאה מובהקת מול תוצאה משמעותית

American Journal of Epidemiology
Copyright © 1999 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved
Vol. 150, No. 9
Printed in U.S.A.

Sex Ratios, Family Size, and Birth Order

Robert J. Biggar,[1] Jan Wohlfahrt,[2] Tine Westergaard,[2] and Mads Melbye[2]

In many countries, the male:female ratio at birth has varied significantly over the past century, but the reasons for these changes have been unclear. The authors observed a close parallel between decreasing family size and declining male:female sex ratio in Denmark from 1960 to 1994. To explain this finding, they examined the sex ratio and birth order of 1,403,021 children born to 700,030 couples. Overall, 51.2% of the first births were male. However, families with boys were significantly more likely than expected to have another boy (biologic heterogeneity). By the fourth birth to families with three prior boys, 52.4% were male. The increase varied directly with the number of prior boys (p for trend = 0.0007). Furthermore, couples with boys were more likely to continue to have children. In summary, the authors found that the declining male:female ratio in Denmark and probably other European populations is mainly attributable to three effects: declining family size, biologic heterogeneity, and child sex preference. Why families with boys are more likely to have additional boys is unknown. *Am J Epidemiol* 1999;150:957–62.

birth order; family characteristics; infant; reproduction; sex distribution; sex ratio

and Public Health
Vol. 150, No. 9
Printed in U.S.A.
lbye[2]
d significantly over the past century, but the reasons
a close parallel between decreasing family size and
994. To explain this finding, they examined the sex
ratio and birth order of 1,403,021 children born to 700,030 couples. Overall, 51.2% of the first births were male.
However, families with boys were significantly more likely than expected to have another boy (biologic
heterogeneity). By the fourth birth to families with three prior boys, 52.4% were male. The increase varied directly
with the number of prior boys (p for trend = 0.0007). Furthermore, couples with boys were more likely to continue
to have children. In summary, the authors found that the declining male:female ratio in Denmark and probably
other European populations is mainly attributable to three effects: declining family size, biologic heterogeneity,
and child sex preference. Why families with boys are more likely to have additional boys is unknown. *Am J Epidemiol* 1999;150:957–62.
birth order; family characteristics; infant; reproduction; sex distribution; sex ratio

במהלך תקופה של 35 שנים המסכן הנולד של בנים שמרה היה גדול ב-7.2 מסכן הילדים שהיו
צריכים להיוולד תחת השערת האיש

תוצאה מובהקת מול תוצאה משמעותית

Association of Plate Design With Consumption of Fruits and Vegetables Among Preschool Children

Enter Search Te

Emily M. Melnick, MPH[1]; Men

Research Le

October 201

Associa Consur ables A

Emily M. Melnick
» Author Affiliati

JAMA Pediatr. 20

an increase in the amount of fruits taken or consumed (Figure). Children in our cohort took a greater mean amount of vegetables per child per day during the intervention period (43.51 g; 95% CI, 37.97-49.05 g) compared with baseline (29.69 g; 95% CI, 27.20-32.18 g), an increase of 13.82 g ($P < .001$). They also consumed a greater mean amount of vegetables per child per day during the intervention period (28.17 g; 95% CI, 24.03-32.32 g) compared with the baseline period (20.63 g; 95% CI, 17.81-23.44 g), an increase of 7.54 g ($P = .001$). Intervention also interac-

www.sci-princess.info
© Yossi Levy 2021