

# עקרונות ויישומים בניתוח סטטיסטי

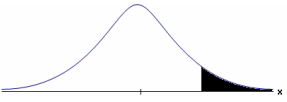
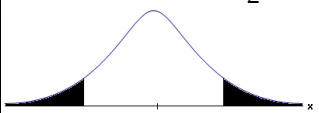
פרק 19 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסייה ידועות ..... 1
2. כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות ..... 5

## בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונות של האוכלוסייה ידועות – רקע

| $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 > c$                                                                                                      | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 < c$                                                                                                        | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 \neq c$                                                                                                                                                                                                                   | השערת האפס:<br>השערה אלטרנטיבית:                                                              |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| מדגמים בלתי תלויים<br>$\sigma_1, \sigma_2$ ידועות<br>$X_1, X_2 \sim N$<br>או מדגמים מספיק גדולים                                                                    |                                                                                                                                                                       |                                                                                                                                                                                                                                                                                     | תנאים:                                                                                        |
| $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$<br><br>$Z_{1-\alpha}$<br>דוחים את $H_0$ | $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$<br><br>$-Z_{1-\alpha}$<br>דוחים את $H_0$ | או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$<br>$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$<br><br>$-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$<br>דוחים את $H_0$ | כלל ההכרעה:<br>אזור הדחייה של $H_0$<br>$-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ |

סטטיסטי המבחן: 
$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

## חלופה אחרת לכלל הכרעה:

| נדחה $H_0$ אם מתקיים:                                                                                   |                                                                                                                                                                                                                              |
|---------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ | $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$<br>או<br>$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |
| $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |                                                                                                                                                                                                                              |

התפלגות הפרש הממוצעים:  $\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{התקנון:}$$

## דוגמה (פתרון בהקלטה):

בשנת 2004 הפער בין השכר הממוצע של הגברים לנשים היה 3000 ₪ לטובת הגברים. מעוניינים לבדוק האם כיום הצטמצם הפער בין הגברים לנשים מבחינת השכר הממוצע. נדגמו 100 עובדים גברים. שכרם הממוצע היה 9,072 ₪. נדגמו 80 עובדות, שכרן הממוצע היה 7809 ₪. לצורך פתרון נניח שסטיות התקן של השכר ידועות ושוות ל-2000 ₪ באוכלוסיית הנשים ו-3000 ₪ באוכלוסיית הגברים. מה המסקנה ברמת מבוהקות של 5%?

## שאלות

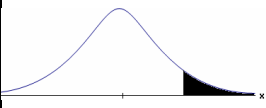
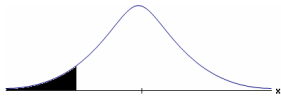
- (1) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלה נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות. לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.
- (2) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.
- (3) במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.
  - א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.
  - ב. מה תהיה המסקנה ברמת מובהקות של 8%?
- (4) במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע. מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו-40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים ו-12 ס"מ אצל הגברים.
  - א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?
  - ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

**תשובות סופיות**

- (1) נדחה  $H_0$ .
- (2) לא נדחה את  $H_0$ .
- (3) א. 0      ב. נדחה את  $H_0$ .
- (4) א. נדחה את  $H_0$ , אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס"מ.  
ב. 0.6331

## בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

| השערת האפס:<br>השערה אלטרנטיבית:<br>תנאים:                                                                     | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 > c$                                                                                                                                 | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 < c$                                                                                                                                  | $H_0 \quad \mu_1 - \mu_2 = c$<br>$H_1 \quad \mu_1 - \mu_2 \neq c$                                                                                                                                                                                                                                                               |
|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. מדגמים בלתי תלויים<br>2. $\sigma_1, \sigma_2$ לא ידועות אך שוות<br>3. המשתנים בכל אוכלוסייה מתפלגים נורמלית | $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$<br><br>$t_{1-\alpha, n_1+n_2-2}$<br>דוחים את $H_0$ | $t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$<br><br>$-t_{1-\alpha, n_1+n_2-2}$<br>דוחים את $H_0$ | $t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$<br><br>$-t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ $t_{1-\frac{\alpha}{2}, n_1+n_2-2}$<br>דוחים את $H_0$ |

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

סטטיסטי המבחן:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

השונות המשוקללת:

## חלופה אחרת לכלל הכרעה:

| נדחה $H_0$ אם מתקיים:                                                                                       |                                                                                                                                                                                                                                                      |
|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ | $\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$<br>או<br>$\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ |
| $\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ |                                                                                                                                                                                                                                                      |

## דוגמה (פתרון בהקלטה):

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשתמשים בה כיום לבניית בניינים. לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות מהסוג החדש. להלן תוצאות המדגם:

טמפרטורת ההתכה הממוצעת במתכת הישנה 1170 מעלות עם אומד חסר הטיה לשונות  $S^2 = 200$ .

טמפרטורת ההתכה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות  $S^2 = 260$ .

נניח לצורך פתרון שטמפרטורת ההתכה מתפלגת נורמאלית עם אותה שונות במתכות השונות. בדקו ברמת מובהקות של 5%.

## שאלות

- (1) להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר):

|     |    |    |     |    |     |     |      |
|-----|----|----|-----|----|-----|-----|------|
| 120 | 94 | 90 | 130 | 95 | 112 | 120 | 2012 |
|     | 69 | 74 | 105 | 91 | 82  | 100 | 2013 |

בדקו שבשנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012 עבור רמת מובהקות של 5%.  
הניחו ששטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שונות.

- (2) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

| המדינה              | ישראל   | ארה"ב   |
|---------------------|---------|---------|
| גודל המדגם          | 15      | 15      |
| סכום הציונים        | 1560    | 1470    |
| סכום ריבועי הציונים | 165,390 | 147,560 |

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים לאמריקאים מבחינת ממוצע הציונים במבחן ה-IQ לטובת ישראל.  
רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (3) להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

| 100W | 60W  | הקבוצה    |
|------|------|-----------|
| 956  | 1007 | $\bar{X}$ |
| 72   | 80   | $S$       |
| 15   | 13   | $n$       |

- א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.  
ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג 100?  
ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.



**תשובות סופיות**

- (1) נדחה את  $H_0$ .
- (2) הנחות:
1. סטיות התקן שוות.
  2. המשתנים מתפלגים נורמלית.
- נקבל את  $H_0$ .
- (3) א. נדחה את  $H_0$ . ב. רמת מובהקות של לפחות 5%.
- ג. לא נדחה את  $H_0$ .