

# סטטיסטיקה יישומית

פרק 14 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסייה ידועות ..... 1
2. כששונויות האוכלוסייה אינן ידועות והמדגמים גדולים ..... 5
3. ניתוח פלטים ..... 8

## 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><br>$-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ |

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

## 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

כשהשונות של האוכלוסייה לא ידועות והמדגמים גדולים – רקע

| השערת האפס:<br>השערה אלטרנטיבית:                                                                                                                                     | השערת האפס:<br>השערה אלטרנטיבית:                                                                                                                                       | השערת האפס:<br>השערה אלטרנטיבית:                                                                                                                                                                                                                                             | השערת האפס:<br>השערה אלטרנטיבית:                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| $H_0 \quad \mu_1 - \mu_2 = c$<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>$H_1 \quad \mu_1 - \mu_2 \neq c$                                                                                                                                                                                                            | תנאים:<br>1. מדגמים בלתי תלויים<br>2. $\sigma_1, \sigma_2$ לא ידועות<br>3. מדגמים מספיק גדולים |
| $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$<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 $H_0$ | $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$<br><br>$-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}$<br>דוחים את $H_0$ | כלל ההכרעה:<br>אזור הדחייה של $H_0$ :                                                          |

סטטיסטי המבחן: 
$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - c}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

## חלופה אחרת לכלל הכרעה:

| נדחה $H_0$ אם מתקיים:                                                                         |                                                                                                                                                                                                          |
|-----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ | $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$<br>או<br>$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ |
| $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ |                                                                                                                                                                                                          |

## דוגמה (פתרון בהקלטה):

נרצה לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. במדגם של 46 נבחנים חיילים התקבל ממוצע 543 וסטיית תקן 123. במדגם של 50 תלמידי תיכון התקבל ממוצע 508 וסטיית תקן 178. מה המסקנה ברמת מובהקות 5%?

## שאלות

(1) חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. במדגם של 35 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות וסטיית תקן 4, ובמדגם של 45 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25 וסטיית תקן 6. האם על סמך תוצאות המדגם יש הבדל בין סוגי הברגים מבחינת רמת הקשיות שלהם? בדקו ברמת מובהקות של 5%.

(2) כדי לבדוק האם נהגים השותים תחת השפעת אלכוהול נוהגים מהר יותר מאלו שאינם שותים בוצע מדגם שבו בדקו את המהירות המקסימאלית של כל נהג בקמ"ש. להלן התוצאות:

| S  | $\bar{X}$ | גדול מדגם |                           |
|----|-----------|-----------|---------------------------|
| 20 | 80        | 70        | נהגים השותים אלכוהול      |
| 15 | 60        | 100       | נהגים שאינם שותים אלכוהול |

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

## תשובות סופיות

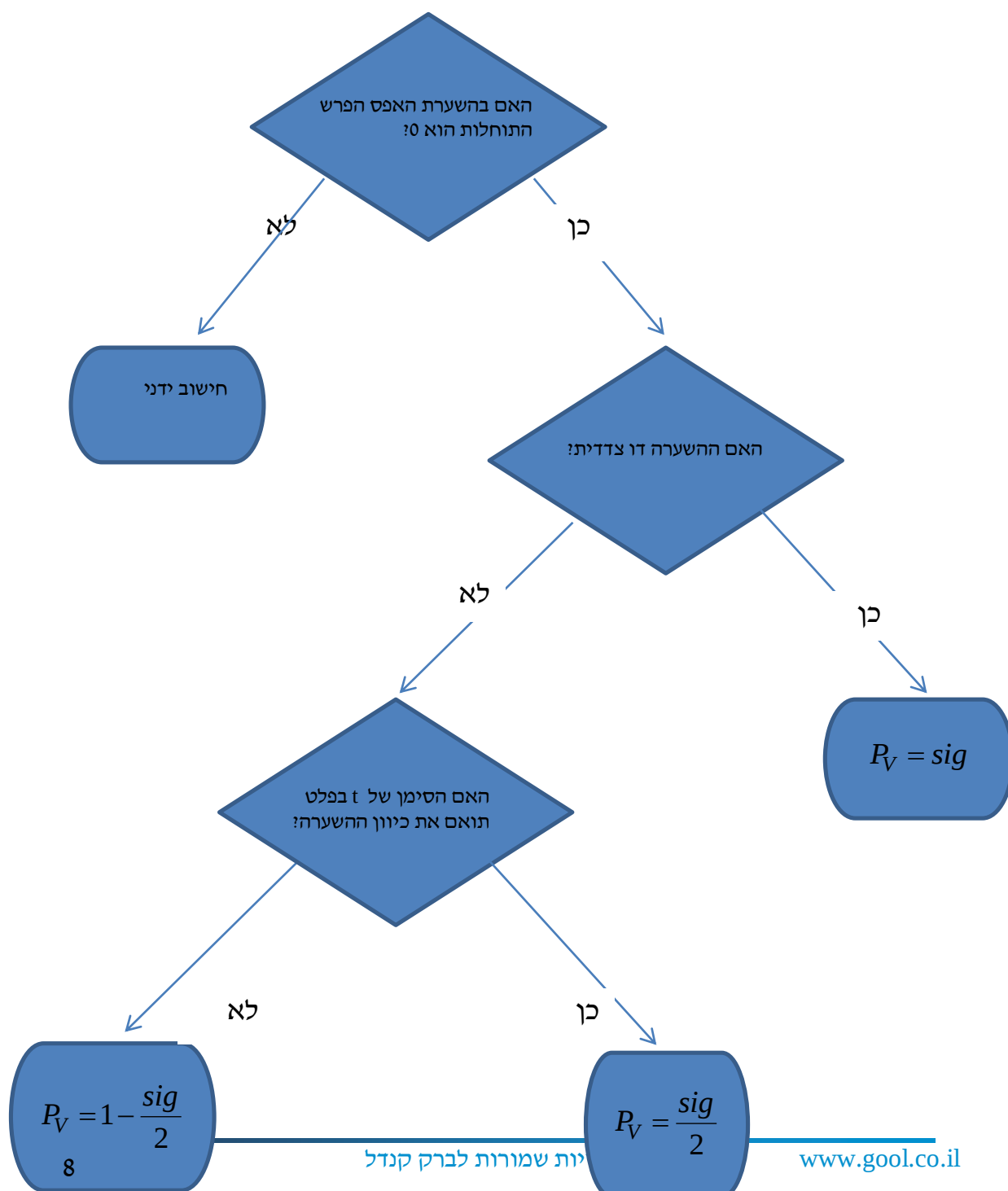
(1) נדחה את  $H_0$ .

(2) א. 0 ב. נדחה את  $H_0$ .

## בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

### ניתוח פלטים – רקע

מובהקות התוצאה על סמך הפלט:



### דוגמה (פתרון בהקלטה) :

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים משני אזורים שונים במדינה על מס' האחים והאחיות שלהם. להלן הפלט שהתקבל :

Group Statistics

|                                | Region of the United States | N   | Mean | Std. Deviation | Std. Error Mean |
|--------------------------------|-----------------------------|-----|------|----------------|-----------------|
| Number of Brothers and Sisters | North East                  | 676 | 3.76 | 2.939          | .113            |
|                                | South East                  | 410 | 4.05 | 2.993          | .148            |

Independent Samples Test

|                                |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |         |                 |                 |                       |                                           |       |
|--------------------------------|-----------------------------|-----------------------------------------|------|------------------------------|---------|-----------------|-----------------|-----------------------|-------------------------------------------|-------|
|                                |                             | F                                       | Sig. | t                            | df      | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |       |
|                                |                             |                                         |      |                              |         |                 |                 |                       | Lower                                     | Upper |
| Number of Brothers and Sisters | Equal variances assumed     | .173                                    | .677 | -1.583                       | 1084    | .114            | -.293           | .185                  | -.657                                     | .070  |
|                                | Equal variances not assumed |                                         |      | -1.576                       | 850.945 | .115            | -.293           | .186                  | -.658                                     | .072  |

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?
- ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.
- ד. מהי מובהקות התוצאה לבדיקת הטענה שההפרש הממוצע בין "South East" לבין "North East" חיובי?

## שאלות

- 1) להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים. רמת האופטימיות נמדדה בסולם ציונים של 1 עד 5.

Group Statistics

|          | GENDER | N   | Mean   | Std. Deviation | Std. Error Mean |
|----------|--------|-----|--------|----------------|-----------------|
| optimizm | MALE   | 633 | 2.6053 | .49781         | .01979          |
|          | FEMALE | 568 | 2.5503 | .48483         | .02034          |

Independent Samples Test

|          |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |          |                 |                 |                       |                                           |        |
|----------|-----------------------------|-----------------------------------------|------|------------------------------|----------|-----------------|-----------------|-----------------------|-------------------------------------------|--------|
|          |                             | F                                       | Sig. | t                            | df       | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |        |
|          |                             |                                         |      |                              |          |                 |                 |                       | Lower                                     | Upper  |
| optimizm | Equal variances assumed     | .383                                    | .536 | 1.935                        | 1199     | .053            | .05500          | .02842                | -.00076                                   | ?      |
|          | Equal variances not assumed |                                         |      | 1.938                        | 1190.977 | .053            | .05500          | .02838                | -.00068                                   | .11067 |

- א. האם ניתן להניח ששוונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
- ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
- ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.
- ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

(2) פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר. להלן הפלט שהתקבל:

### Group Statistics

|       |   | N  | Mean     | Std. Deviation | Std. Error Mean |
|-------|---|----|----------|----------------|-----------------|
| grade | A | 9  | 96.8889  | 9.40006        | 3.13335         |
|       | B | 11 | 108.4545 | 11.46616       | 3.45718         |

| Independent Samples Test |                             |                                         |      |                              |        |                 |                 |                                           |                    |
|--------------------------|-----------------------------|-----------------------------------------|------|------------------------------|--------|-----------------|-----------------|-------------------------------------------|--------------------|
|                          |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                                           |                    |
|                          |                             |                                         |      |                              |        |                 |                 | 95% Confidence Interval of the Difference |                    |
|                          |                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference                     |                    |
| grade                    | Equal variances assumed     | .206                                    | .656 | -2.428                       | 18     | .026            | -11.56566       | 4.76333                                   | Lower Upper        |
|                          | Equal variances not assumed |                                         |      | -2.479                       | 17.997 | .023            | -11.56566       | 4.66583                                   | -21.36832 -1.76299 |

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק

ברמת מובהקות של  $\alpha = 0.05$ .

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של  $\alpha = 0.01$ .

3) כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז. לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

## T-Test

Group Statistics

| מקום המגורים       | N  | Mean    | Std. Deviation | Std. Error Mean |
|--------------------|----|---------|----------------|-----------------|
| גיל הנישואין עיר א | 50 | 24.8072 | 1.38978        | .19654          |
| עיר ב              | 50 | 23.0131 | 1.62070        | .22920          |

Independent Samples Test

|              |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                       |                                           |         |
|--------------|-----------------------------|-----------------------------------------|------|------------------------------|--------|-----------------|-----------------|-----------------------|-------------------------------------------|---------|
|              |                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |         |
|              |                             |                                         |      |                              |        |                 |                 |                       | Lower                                     | Upper   |
| גיל הנישואין | Equal variances assumed     | .330                                    | .567 | 5.942                        | 98     | .000            | 1.79415         | .30193                | 1.19497                                   | 2.39332 |
|              | Equal variances not assumed |                                         |      | 5.942                        | 95.772 | .000            | 1.79415         | .30193                | 1.19480                                   | 2.39350 |

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.
- ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

## 4) להלן פלט של תוכנת SPSS :

## T-Test

|   | N  | Mean    | Std. Deviation | Std. Error Mean |
|---|----|---------|----------------|-----------------|
| X | 26 | 36.3077 | 13.23259       | 2.59513         |
| Y | 24 | 46.4583 | 20.96369       | 4.27920         |

## Independent Samples Test

|                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                       |                                           |         |
|-----------------------------|-----------------------------------------|------|------------------------------|--------|-----------------|-----------------|-----------------------|-------------------------------------------|---------|
|                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |         |
|                             |                                         |      |                              |        |                 |                 |                       | Lower                                     | Upper   |
| Equal variances assumed     | 4.446                                   | .040 | -2.164                       | ???    | .044            | -10.15064       | ???                   | -20.03781                                 | -.26347 |
| Equal variances not assumed |                                         |      | -2.038                       | 38.267 | .048            | ???             | 5.00462               | -20.27964                                 | -.02164 |

א. השלימו את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של  $X$  לזה של  $Y$ ?ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של  $X$  גדולה מהתוחלת של  $Y$ ?ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של  $X$  קטנה מהתוחלת של  $Y$ ?

## תשובות סופיות

- (1) א. נקבל את  $H_0$  ונכריע שיש שוויון שוניות.  
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.  
 ג. 0.11076  
 ד.  $2.5665 \leq \mu \leq 2.6441$
- (2) א. מבחן T להפרש ממוצעים במדגמים בלתי תלויים.  
 ב. מבחן T למדגמים מזווגים.  
 ג. נקבל את  $H_0$ , נקבע לקיום שוויון שוניות.  
 ד. נקבל את  $H_0$ , לא נקבל את טענת הפסיכולוגים.
- (3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.  
 ב.  $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$  ג. כן.
- (4) א. 10.15, 4.69, -48  
 ב. 0.04  
 ג. 0.978  
 ד. 0.022