# ביוסטטיסטיקה מתקדמת

## פרק 5 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי ..... 1

# ניתוח שונות חד כיוונית

## רקע תיאורטי

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $\left(\mu_{1\ldots},\mu_k\right)$ של $k$ אוכלוסיות שונות.
לכן, בניתוח שונות, <u>השערות</u> המחקר הן:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ (התוחלות של כל האוכלוסיות שוות)

$H_1:$ אחרת (לפחות שתיים מהתוחלות שונות)

**ההנחות הדרושות לביצוע התהליך:**

**2)** בכל אוכלוסייה מתוך $k$ האוכלוסיות ההתפלגות נורמלית.
**3)** כל האוכלוסיות הן עם אותה שונות $\sigma^2$.
**4)** המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor). משתנה זה הוא קטגוריאלי עם $k$ רמות (levels).
כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה:
נסמן ב- $n_i$ את גודל המדגם בקבוצה $i$.

$n = \sum_{1}^{k} n_i$ - מספר התצפיות סך הכול (בכל המדגמים).

$\overline{X}_1$ - ממוצע המדגם הראשון,.....,$\overline{X}_k$ -ממוצע המדגם ה- $k$ -י.
$\overline{X}$ - ממוצע כללי (של כל המדגמים).

סכום ריבועים בין הקבוצות: $SS_B = \sum_{i=1}^{k} n_i \left[\overline{X}_i - \overline{X}\right]^2$

סכום ריבועים בתוך הקבוצות: $SS_W = \sum_{i=1}^{k} n_i \left[n_i - 1\right] \bullet \hat{S}_i^2$

סכום ריבועים כללי: $SS_T = \sum_{i=1}^{k} \sum_{j=1}^{n_j} [X_{ij} - \overline{X}]^2$

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה:


© כל הזכויות שמורות לברק קנדל

| מקור השונות | סכום הריבועים SS | דרגות חופש $df$ | ממוצע הריבועים MS | F |
|---|---|---|---|---|
| $B$ - בין הקבוצות | $SSB$ | $k-1$ | $\frac{SSB}{k-1}$ | $\frac{MSB}{MSW}$ |
| $W$ - בתוך הקבוצות | $SSW$ | $n-k$ | $\frac{SSW}{n-k}$ | |
| $T$ - סה"כ | $SST$ | $n-1$ | | |

$$F=\frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}}\sim F\left(k-1,n-k\right)$$

אזור דחיית $H_0$ : $F>F_{(k-1,n-k)}:1-\alpha$


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש: קבוצה 1קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן".
כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר?
מהו ה"גורם" וכמה רמות יש לו?

ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.

ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

**2)** בעיר מסוימת שלושה בתי ספר תיכון. ראש העיר התעניין לבדוק האם קיים הבדל בהצלחה של בתי הספר במקצוע מתמטיקה. לצורך כך הוא דגם מספר תלמידים שנבחנו במבחן הבגרות במתמטיקה ברמה של 3 יחידות בעירו ובדק עבור כל תלמיד מה ציון הבגרות שלו במתמטיקה.
להלן הציונים שהתקבלו:

| "המתמיד" | "רבין" | "הס" |
|---|---|---|
| 78 | 98 | 85 |
| 65 | 62 | 83 |
| 70 | 55 | 74 |
| 90 | 80 | 85 |
| 56 | | 75 |

א. מהו המבחן הסטטיסטי המתאים?
רשמו את ההשערות ואת ההנחות של המבחן.

ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (factor) כמה רמות יש לו?

ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.

ד. מלאו את טבלת ANOVA.

ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.

ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

**3)** מעוניינים לבדוק האם יש הבדל בהשפעה של שיטות טפול שונות על לחץ הדם הסיסטולי (SBP) באוכלוסייה של קשישים. נבדקו 4 שיטות שונות.
בטבלה המצורפת מרוכזים ממצאי המחקר.

| השיטה | A | B | C | D |
|---|---|---|---|---|
| גודל המדגם | 12 | 14 | 8 | 12 |
| הממוצע | 178 | 172 | 180 | 182 |
| סטיית התקן | 4 | 8 | 5 | 3 |

א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.

ב. מה מסקנת המחקר ברמת מובהקות של 5%?

ג. האם יש צורך לבצע השוואות מרובות?


© כל הזכויות שמורות לברק קנדל

4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.
האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות?
בדקו ברמת מובהקות של 5%.

| האופה | ניר | מוזס | שלום |
|---|---|---|---|
| סכום הזמנים | 206 | 212 | 182 |
| סכום ריבועי הזמנים | 10644 | 11250 | 8982 |

5) להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות.
מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

**ANOVA**

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 10.317 | 3 | 3.439 | 1.361 | .279 |
| Within Groups | 60.648 | 24 | 2.527 | | |
| Total | 70.964 | 27 | | | |

6) להלן טבלת ANOVA בטבלה הושמטו חלקים.
השלימו את החלקים בטבלה שהושמטו ומסומנים באותיות.

**ANOVA**

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 357.450 | ב | ג | ה | .000 |
| Within Groups | א | 17 | ד | | |
| Total | 522.950 | 19 | | | |


© כל הזכויות שמורות לברק קנדל

7) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים ל שלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה ( dosage). המינונים שניתנו הם : 10 מ״ג, 20 מ״ג ו-30 מ״ג. לאחר שעה מזמן לקיחת התרופה ניבדק קצב פעימות הלב של כל אדם (pulse). הנתונים הוזנו לתוכנה סטטיסטית והתקבלו התוצאות הבאות:

**ANOVA**

pulse

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 414.400 | 2 | 207.200 | 19.733 | .000 |
| Within Groups | 126.000 | 12 | 10.500 | | |
| Total | 540.400 | 14 | | | |

**pulse**

Tukey HSD[a]

| dosage | N | Subset for alpha = 0.05 | |
|---|---|---|---|
| | | 1 | 2 |
| 30.00 | 5 | 71.0000 | |
| 20.00 | 5 | | 80.2000 |
| 10.00 | 5 | | 83.4000 |
| Sig. | | 1.000 | .299 |

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

## Post Hoc Tests

**Multiple Comparisons**

pulse
Tukey HSD

| (I) dosage | (J) dosage | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 10.00 | 20.00 | 3.20000 | 2.04939 | .299 | -2.2675 | 8.6675 |
| | 30.00 | 12.40000* | 2.04939 | .000 | 6.9325 | 17.8675 |
| 20.00 | 10.00 | -3.20000 | 2.04939 | .299 | -8.6675 | 2.2675 |
| | 30.00 | 9.20000* | 2.04939 | .002 | 3.7325 | 14.6675 |
| 30.00 | 10.00 | -12.40000* | 2.04939 | .000 | -17.8675 | -6.9325 |
| | 20.00 | -9.20000* | 2.04939 | .002 | -14.6675 | -3.7325 |

*. The mean difference is significant at the 0.05 level.

א. בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.

ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

ג. האם יש צורך במחקר בהשוואת מרובות. נמקו!

ד. לטבלת הANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת ״טוקי״. ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?


© כל הזכויות שמורות לברק קנדל

8) בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו. ענו על הסעיפים הבאים:

א. כמה בתי ספר יש בעיר?

ב. כמה תלמידים השתתפו בסך הכול במחקר?

ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%

ד. בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

**Oneway**

**ANOVA**

grade

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 7799.600 | 4 | 1949.900 | 13.586 | .000 |
| Within Groups | 2870.400 | 20 | 143.520 | | |
| Total | 10670.000 | 24 | | | |


© כל הזכויות שמורות לברק קנדל

## Post Hoc Tests

**Multiple Comparisons**

grade

Scheffe

| (I) school | (J) school | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| 1.00 | 2.00 | 5.40000 | 7.57681 | .971 | -20.2543 | 31.0543 |
| | 3.00 | 36.80000* | 7.57681 | .003 | 11.1457 | 62.4543 |
| | 4.00 | 36.40000* | 7.57681 | .003 | 10.7457 | 62.0543 |
| | 5.00 | -2.60000 | 7.57681 | .998 | -28.2543 | 23.0543 |
| 2.00 | 1.00 | -5.40000 | 7.57681 | .971 | -31.0543 | 20.2543 |
| | 3.00 | 31.40000* | 7.57681 | .011 | 5.7457 | 57.0543 |
| | 4.00 | 31.00000* | 7.57681 | .013 | 5.3457 | 56.6543 |
| | 5.00 | -8.00000 | 7.57681 | .888 | -33.6543 | 17.6543 |
| 3.00 | 1.00 | -36.80000* | 7.57681 | .003 | -62.4543 | -11.1457 |
| | 2.00 | -31.40000* | 7.57681 | .011 | -57.0543 | -5.7457 |
| | 4.00 | -.40000 | 7.57681 | 1.000 | -26.0543 | 25.2543 |
| | 5.00 | -39.40000* | 7.57681 | .001 | -65.0543 | -13.7457 |
| 4.00 | 1.00 | -36.40000* | 7.57681 | .003 | -62.0543 | -10.7457 |
| | 2.00 | -31.00000* | 7.57681 | .013 | -56.6543 | -5.3457 |
| | 3.00 | .40000 | 7.57681 | 1.000 | -25.2543 | 26.0543 |
| | 5.00 | -39.00000* | 7.57681 | .001 | -64.6543 | -13.3457 |
| 5.00 | 1.00 | 2.60000 | 7.57681 | .998 | -23.0543 | 28.2543 |
| | 2.00 | 8.00000 | 7.57681 | .888 | -17.6543 | 33.6543 |
| | 3.00 | 39.40000* | 7.57681 | .001 | 13.7457 | 65.0543 |
| | 4.00 | 39.00000* | 7.57681 | .001 | 13.3457 | 64.6543 |

*. The mean difference is significant at the 0.05 level.

## Homogeneous Subsets

**grade**

Scheffe[a]

| school | N | Subset for alpha = 0.05 1 | Subset for alpha = 0.05 2 |
|---|---|---|---|
| 3.00 | 5 | 45.0000 | |
| 4.00 | 5 | 45.4000 | |
| 2.00 | 5 | | 76.4000 |
| 1.00 | 5 | | 81.8000 |
| 5.00 | 5 | | 84.4000 |
| Sig. | | 1.000 | .888 |

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 5.000.


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות

**1)** א. משתנה בלתי תלוי: סוג התרופה.
משתנה תלוי: הזמן עד להשפעת התרופה בדקות.

ב. ניתוח שונות חד כיווני
$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$.

ג. 1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

**2)** א. המבחן לניתוח שונות חד כיוונית.
$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונות.

ב. גודל המדגם: 14. משתנה ב"ת: בית הספר, בעל 3 רמות.

ג. $\bar{X} = 71.8, \hat{S} = 12.93$ , $\bar{X} = 73.75, \hat{S} = 19.29$ , $\bar{X} = 80.4, \hat{S} = 5.46$ .

ד. להלן טבלה:

| מקור השונות | SS | df | MS | F |
|---|---|---|---|---|
| B | 200.6 | 2 | 100.3 | |
| W | 1904.75 | 11 | 173.2 | |
| סה"כ | 2105.35 | 13 | | 0.58 |

ה. $F > 3.98$ .

ו. נקבל את $H_0$ .

**3)** א. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1: otherwise$

ב. נדחה את $H_0$ .

ג. כן.

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.


© כל הזכויות שמורות לברק קנדל

4) נקבל את $H_0$ : נכריע שאין הבדל מובהק בין האופים מבחינת תוחלת זמן הכנה.

5) $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1: otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

נקבל את $H_0$ : לסוג סוללה אין השפעה של תוחלת החיים ברמת ביטחון של 10%.

6) להלן טבלה:

**ANOVA**

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 357.450 | 2 ב | 178.725 ג | 18.36 ה | .000 |
| Within Groups | 165.5 א | 17 | 9.735 ד | | |
| Total | 522.950 | 19 | | | |

7) א. $H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונות.

נדחה את $H_0$ : ברמת ביטחון של 5% קיים הבדל במינונים השונים מבחינת תוחלת הדופק.

ב. ראה וידאו. ג. כן. ד. $\mu_{20} = \mu_{10} > \mu_{30}$ .

8) א. 5 ב. 25

ג. נדחה את $H_0$ : יש לפחות שני בתי ספר בעיר עם תוחלת רמת ציונים שונה.

ד. $(\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3)$ .


© כל הזכויות שמורות לברק קנדל