

שיטות מחקר - תרגיל 6

מגישות - זוהר בוחניק 311142293 אלה דובדבן 305564866

22 במאי 2019

שאלה 1:

נגדיר את תנאי הניסוי:

$n = 24$ מספר הנבדקים, אשר חולקו לקבוצות הבאות בצורה רנדומלית:

1. ריצה

2. נסיעה באופניים

3. כדורגל

4. כדורסל

המשתנה הבלתי תלוי - ענף הספורט

המשתנה התלוי - רמת חומצת החלב בדם (המייצג את רמת הכושר הגופני הדרוש לפעילות).

(א)

נבנה סט של קונטרסטים מתוכננים התואמים את ההשערות הבאות של החוקרת:

1. השערת החוקרת: הכושר הגופני הדרוש למשחקי כדור נמוך ממקצועות אחרים

נבחר את הקונטרסטים הבאים:

$$C_1 = C_2 = 1, \quad C_3 = C_4 = -1$$

כעת נגדיר את ψ להיות:

$$\psi = \bar{x}_1 + \bar{x}_2 - \bar{x}_3 - \bar{x}_4$$

ונקבל את ההשערות הרצויות:

$$H_0 : \psi \geq 0$$

$$H_1 : \psi < 0$$

2. השערת החוקרת: הכושר הגופני הדרוש לכדורגל גבוה יותר מזה הנדרש לכדורסל

נבחר את הקונטרסטים הבאים:

$$C_1 = C_2 = 0, \quad C_3 = 1, \quad C_4 = -1$$

כעת נגדיר את ψ להיות:

$$\psi = \bar{x}_3 - \bar{x}_4$$

ונקבל את ההשערות הרצויות:

$$H_0 : \psi \leq 0$$

$$H_1 : \psi > 0$$

(ב)

נבנה קונטרסט אורתוגונלי-

$$C_1 = 1, C_2 = -1, C_3 = 0, C_4 = 0$$

כלומר:

$$\psi = \bar{x}_1 - \bar{x}_2$$

ההשערה על פי קונטרסט זה היא שהכושר הגופני הדרוש לריצה גבוה יותר מזה הנדרש לנסיעה באופניים.

$$H_0 : \psi \leq 0$$

$$H_1 : \psi > 0$$

הוא בלתי תלוי באחרים כיוון שבכל השערה אנו בודקים את הקשרים בין קבוצות שונות. בהשערה הראשונה הסתכלנו על משחקי כדור לעומת משחקים אחרים, בהשערה השנייה הסתכלנו בתוך קבוצת משחקי הכדור על רמת הפעילות הגופנית שנדרשת בכל אחד בהשוואה בניהם, ובהשערה הזאת אנו מסתכלים על קבוצת המשחקים האחרים ועל מי דורש יותר פעילות גופנית - כלומר על הקשר הפנימי בתוך הקבוצה של משחקים אחרים. ברמה המתמטית ניתן לראות כי אכן וקטורי הקונטרסטים הם אורתוגונלים אחד לשני (פירוט בסעיף ג').

(ג)

נראה את אי התלות בין שלושת הקונטרסטים שהגדרנו ע"י כך שנראה כי המכפלה הסקלרית של כל זוג קונטרסטים היא אפס:

$$\left\langle \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \right\rangle = 1 - 1 + 0 + 0 = 0$$

$$\left\langle \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \right\rangle = 0 + 0 + 0 + 0 = 0$$

$$\left\langle \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\rangle = 0 + 0 - 1 + 1 = 0$$

כנדרש.

(ד)

מצורף קובץ R. תוצאות:

```

> contrast(leastsquare, weights1, adjust="none")
contrast      estimate    SE df t.ratio p.value
c(1, 1, -1, -1)      5.17 1.65 20  3.135  0.0052

> contrast(leastsquare, weights2, adjust="none")
contrast      estimate    SE df t.ratio p.value
c(0, 0, 1, -1)       1.33 1.17 20  1.144  0.2661

> contrast(leastsquare, weights3, adjust="none")
contrast      estimate    SE df t.ratio p.value
c(1, -1, 0, 0)      -0.5 1.17 20 -0.429  0.6725

```

שאלה 2:

מערך המחקר שבחרנו הוא: השפעה של מין והשכלה על-תיכונית על מדד IQ .
מהי שאלת המחקר התיאורטית?

האם יש קשר בין מין והשכלה על-תיכונית על מדד ה- IQ של אדם.

מהי האופרציונליזציה של המשתנים השונים?

נבצע מבחני IQ ונאסוף את ממוצעי הציונים ביחד עם נתוני המין וההשכלה הגבוהה של הנבדקים.

מהו סוג מערך המחקר המוצע (מתאמי/ניסויי)?

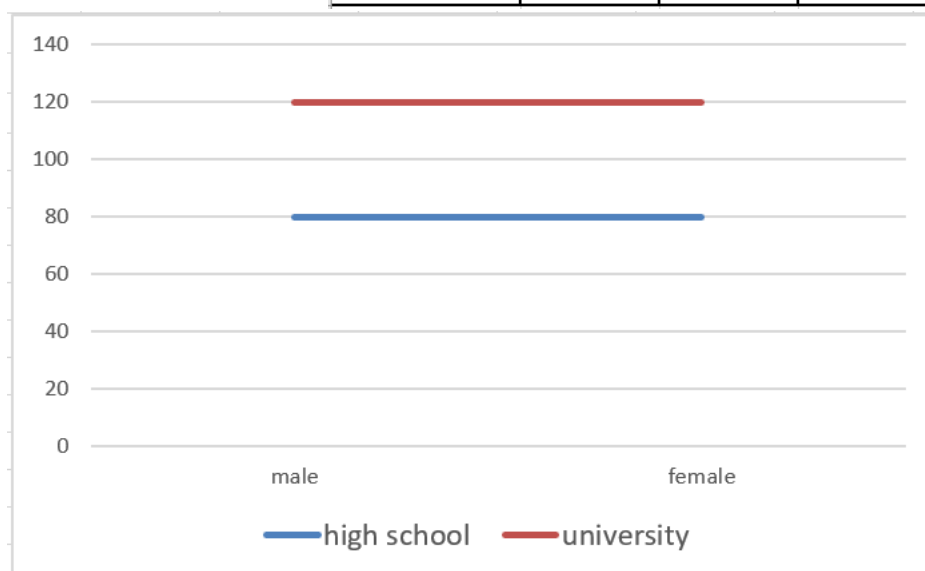
מתאמי.

האם ניתן לבצע בו הקצאה מקרית?

קיים ארטיפקט אפשרי - שכן אנו מודדים את ההשכלה הגבוהה, אשר יכולה להיות תוצר נלווה של מצב סוציו-אקונומי (הקצאה לא מקרית).

ב. יש אפקט עיקרי לשורות, ואין אפקט לעמודות ולאינטראקציה.

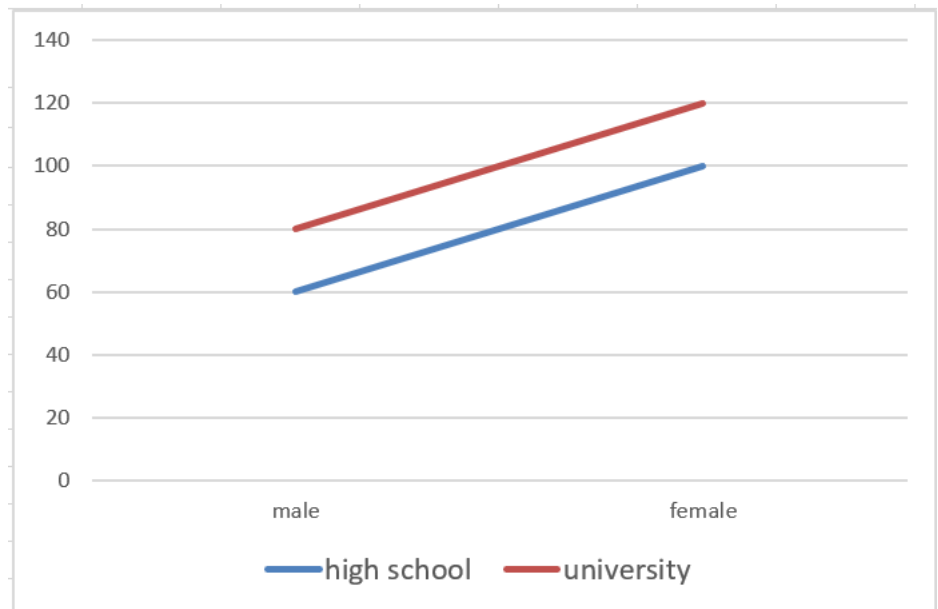
	male	female	
high school	80	80	80
university	120	120	120
	100	100	100



מה שניתן להסיק מהתוצאות זה שרמת ה- IQ אצל אנשים שלמדו בהשכלה גבוהה, גבוהה יותר משל אנשים שלא למדו בהשכלה גבוהה, וכי המין לא משפיע על רמת ה- IQ . כמו כן לא קיימת אינטראקציה בין המשתנים הבלתי תלויים שבדקנו.

ג. יש אפקט לעמודות ולשורות, אין אפקט אינטראקציה.

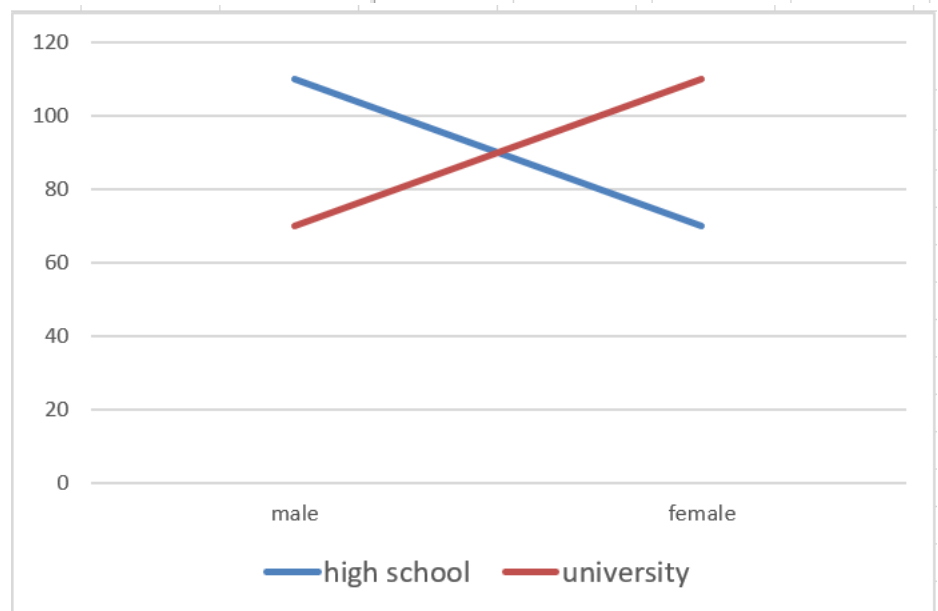
	male	female	
high school	60	100	80
university	80	120	100
	70	110	90



מה שניתן להסיק מהתוצאות זה שגם המין וגם השכלה גבוהה הם גורמים המשפיעים על רמת ה- IQ . זאת כאשר תוצאות הנשים היו גבוהות יותר משל הגברים, ואנשים שלמדו השכלה גבוהה הפגינו רמת IQ גבוהה יותר. לא נמצאה אינטרציה בין 2 הגורמים הבלתי תלויים האלו.

ד. יש אפקט אינטראקציה, אין אפקטים עיקריים.

	male	female	
high school	110	70	90
university	70	110	90
	90	90	90

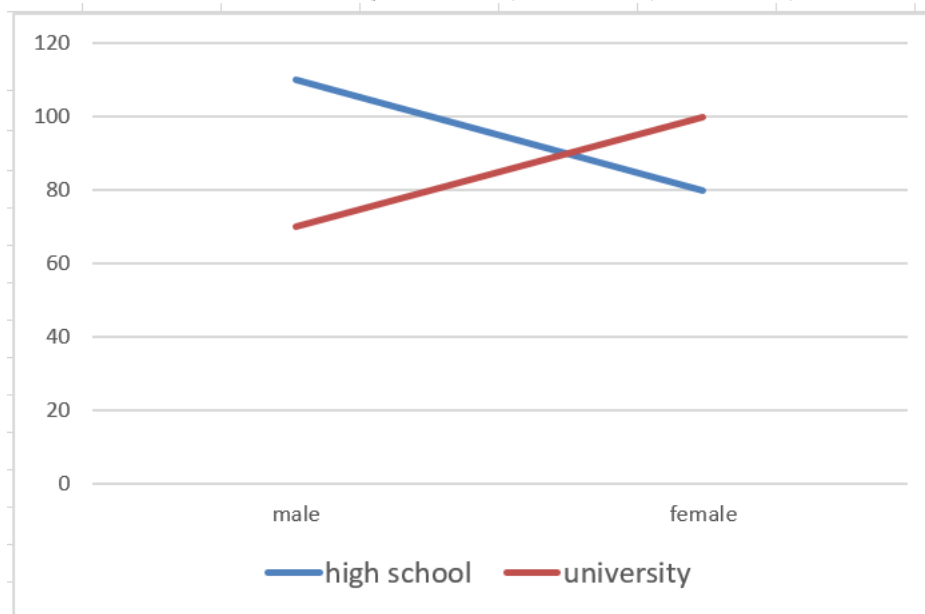


מה שניתן להסיק מהתוצאות זה שיש השפעה של מין והשכלה גבוהה על רמת ה- IQ , אמנם לא קיים אפקט ברור עבור כל אחד מבין המשתנים בנפרד. רק בשילוב בין שני המשתנים נקבל אפקט ברור על ה- IQ למשל אישה שלומדת באוניברסיטה תתאפיין

ברמת IQ גבוהה יותר, לעומת זאת עבור גבר אם הוא דווקא לא למד באוניברסיטה הוא יתאפיין ברמת IQ גבוהה יותר מאשר גבר שכן למד באוניברסיטה.

ה. יש אפקט אינטראקציה ואפקט שורות, אין אפקט עמודות.

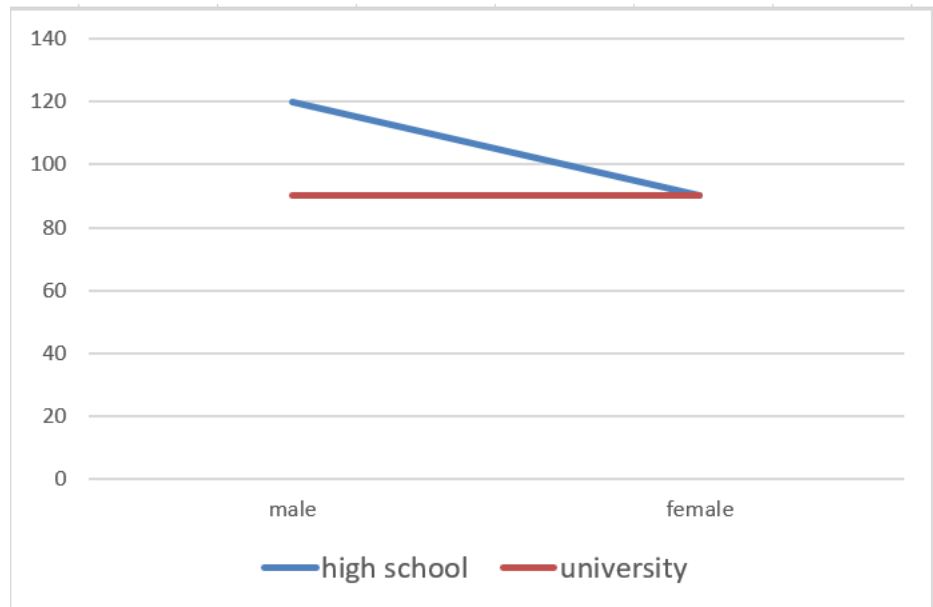
	male	female	
high school	110	80	95
university	70	100	85
	90	90	90



מה שניתן להסיק מהתוצאות זה שלמשתנה הבלתי תלוי השכלה גבוהה יש השפעה על רמת ה- IQ , אמנם לא נמצא הבדל בין נשים לגברים ברמת ה- IQ במוצע. על האינטראקציה בניהם ניתן ללמוד כי המין משפיע בצורה כלשהי כאשר הוא נמדד יחד עם הפרמטר של השכלה גבוהה ונקבל אפקט על רמת ה- IQ בשילוב 2 המשתנים יחד.

ו. יש אפקט אינטראקציה, עמודות, ושורות.

	male	female	
high school	120	90	105
university	90	90	90
	105	90	97.5



מה שניתן להסיק מהתוצאות זה כי לכל משתנה בנפרד יש השפעה על רמת ה- IQ וגם לשילוב שלהם ביחד יש השפעה אחרת על רמת ה- IQ . זאת ניתן לראות לפי זה שיש הבדלים בין הממוצעים הן של השורות והן של העמודות ואת האינטראקציה ניתן לראות על ידי זה שיש למשל הבדל בין גבר שלמד בהשכלה גבוהה לבין גבר שלא, אבל אצל אישה למשל אין הבדל ברמת ההשכלה אם היא למדה בהשכלה הגבוהה או לא- כלומר לשילוב בניהם יש משמעות חדשה.

שאלה 3:

ראשית - נמלא את טבלת הממוצעים:

	כדורסל	אופניים	ריצה	
גילאים 20-30	8	3.66	3.66	5.1
גילאים 30-40	7.66	7.66	9.66	8.32
	7.83	5.66	6.66	6.716

סעיף א-

	df	SS	MS	F	F קריטי
שורות	1	$SSR = 46.658$	46.658	4.14	4.747
טורים	2	$SSC = 14.1556$	7.078	0.62795	3.885
אינטראקציה	2	$SSI_{nt} = 31.516$	15.758	1.39822	3.885
טעות	12	$SSE = 135.245$	11.27	■	■

$$SSC = \sum_{k=1}^C n \cdot R \cdot (\bar{x}_k - \bar{x})^2 = 3 \cdot 2 \cdot (6.66 - 6.716)^2 + 3 \cdot 2 \cdot (5.66 - 6.716)^2 + 3 \cdot 2 \cdot (7.83 - 6.716)^2 = 14.155608$$

$$SSE = \sum_{j=1}^R \sum_{k=1}^C \sum_{i=1}^n (x_{i,j,k} - \bar{x}_{j,k})^2 = (2 - 3.66)^2 + (3 - 3.66)^2 + (6 - 3.66)^2 + (3 - 3.66)^2 + (7 - 3.66)^2 + (1 - 3.66)^2 + (9 - 8)^2 + (9 - 8)^2 + (6 - 8)^2 + (9 - 6.66)^2 + (6 - 6.66)^2 + (14 - 6.66)^2 + (11 - 5.66)^2 + (4 - 5.66)^2 + (8 - 5.66)^2 + (7 - 6.716)^2 + (9 - 6.716)^2 + (7 - 6.716)^2 = 135.245168$$

$$SSR = \sum_{j=1}^R n \cdot C \cdot (\bar{x}_j - \bar{x})^2 = \sum_{j=1}^R n \cdot C \cdot (\bar{x}_j - \bar{x})^2 = 3 \cdot 3 \cdot \left[(5.1 - 6.716)^2 + (8.32 - 6.716)^2 \right] = 46.65833$$

$$\begin{aligned} SS_{int} &= \sum_{j=1}^R \sum_{k=1}^C n \cdot (\bar{x}_{j,k} - \bar{x}_{j.} - \bar{x}_{.k} + \bar{x})^2 = \\ &3 \cdot \left[(8 - 5.1 - 7.83 + 6.716)^2 + (3.66 - 5.1 - 5.66 + 6.716)^2 + (3.66 - 5.1 - 6.66 + 6.716)^2 \right] + \\ &3 \cdot \left[(7.66 - 8.32 - 7.83 + 6.716)^2 + (7.66 - 8.32 - 5.66 + 6.716)^2 + (9.66 - 8.32 - 6.66 + 6.716)^2 \right] \\ &= 15.758124 + 15.758124 = 31.516248 \end{aligned}$$

דרגות חופש:

$$df_R = R - 1 = 2 - 1 = 1$$

$$df_C = C - 1 = 3 - 1 = 2$$

$$df_{int} = (R - 1)(C - 1) = 2$$

$$df_e = N - R \cdot C = 18 - 6 = 12$$

אומדים לשונות:

$$MSR = \frac{SSR}{df_R} = \frac{46.658}{1} = 46.658$$

$$MSC = \frac{SSC}{df_C} = \frac{14.1556}{2} = 7.077$$

$$MS_{int} = \frac{SS_{int}}{df_{int}} = \frac{31.516}{2} = 15.758$$

$$MSE = \frac{SSE}{df_e} = \frac{135.245}{12} = 11.27$$

ערכי F :

$$\frac{MSR}{MSE} = \frac{46.658}{11.27} = 4.14$$

$$\frac{MSC}{MSE} = \frac{7.077}{11.27} = 0.62795$$

$$\frac{MS_{int}}{MSE} = \frac{15.758}{11.27} = 1.39822$$

סעיף ב'

נסכם במילים את תוצאות ניתוח השונות ומה ניתן להסיק\לא-להסיק מבדיקת כל אחד מהאפקטים במערך: ניתן לראות כי יש אפקט שורות ואינטראקציה. זאת כיוון שהצלחנו לדחות במקרים האלה את השערת ה-0 והערך F שהתקבל היה גדול מהערך הקריטי עבור $\alpha = 0.05$. לא ניתן לדחות את השערת האפס עבור אפקט העמודות ולכן לא ניתן להסיק כלום על ההשפעה של אופי הפעילות הספורטיבית על מצב הרוח. מה שכן ניתן לראות זה שגיל המתעמל/ת משפיע על מצב הרוח, ובנוסף ניתן לראות כי קיימת אינטראקציה בין אופי הפעילות הספורטיבית וגיל המתעמלת אשר משפיעה בצורה מסוימת על מצב הרוח.

שאלה 4:

סעיף א'

התוצאות שקיבלנו:

```
> two_way_aov <- aov(len ~ supp * dose, data = my_data)
> summary(two_way_aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
supp	1	205.4	205.4	15.572	0.000231	***
dose	2	2426.4	1213.2	92.000	< 2e-16	***
supp:dose	2	108.3	54.2	4.107	0.021860	*
Residuals	54	712.1	13.2			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

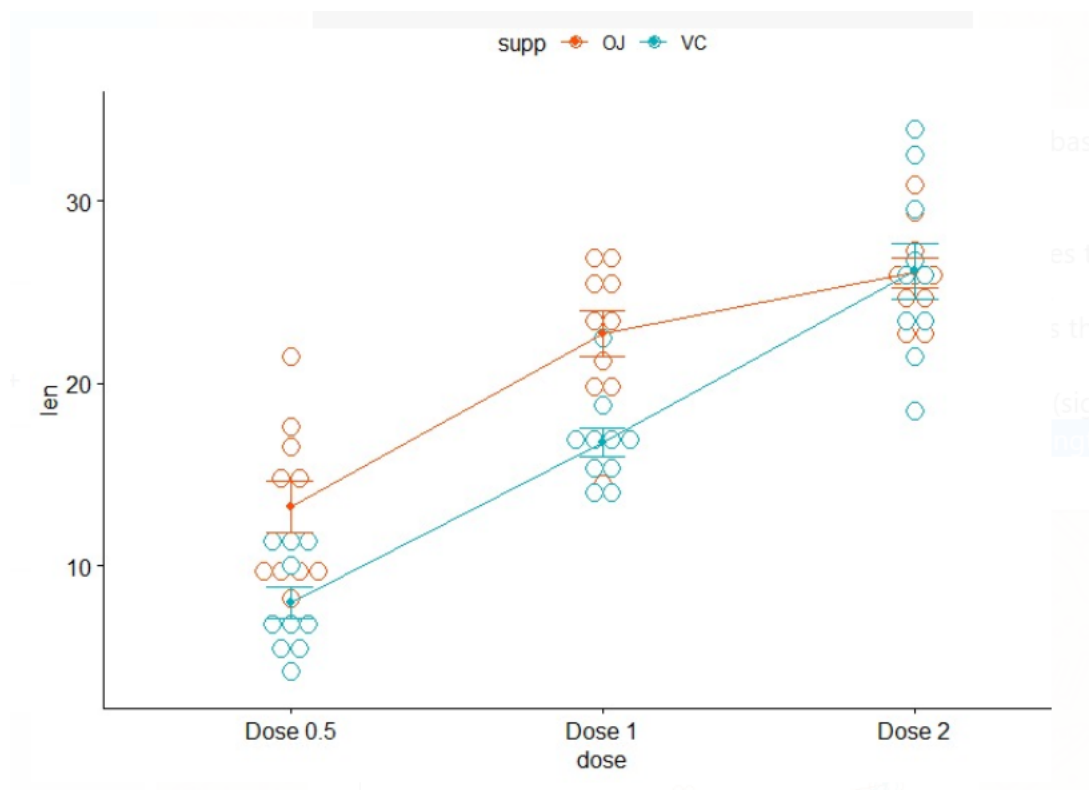
אפקט	ערכי F	דרגות חופש	p -value	האם האפקט שנמצא מובהק?
supp	15.572	1	0.000231	כן
dose	92	2	$2e-16$ (מספר מאוד קרוב לאפס)	כן
אינטראקציה	4.107	2	0.021860	כן

- $supp$ מייצג את הדרך שבה נצרך הויטמין C כאשר OJ מייצג מיץ תפוזים ו- VC מייצג צריכה דרך האוכל.
- $dose$ מייצג את המינון שבו קיבלנו את הויטמין C - כאלו שקיבלו מידי יום מינון של 5.0 מ"ל גרם של ויטמין C , כאלו שקיבלו מינון של 1 מ"ג וכאלו שקיבלו מינון של 2 מ"ג

מה ניתן או לא ניתן להסיק מהתוצאות?

ניתן לדחות את השערת האפס בשני המשתנים, כלומר יש השפעה של הדרך שבה נצרך הויטמין C על צמיחת השיניים של שרקנים וגם יש השפעה של המינון שבו נצרך ויטמין C לצמיחת השיניים של שרקנים. בעבור האינטראקציה ניתן לדחות את השערת האפס ונובע כי הקשר בין המינון שבו נצרך ויטמין C לצמיחת השיניים של שרקנים תלוי בדרך שבה הוא נצרך.

גרף אינפורמטיבי של ההשפעה של המינון על הצמיחה של השן עבור 2 שיטות הצריכה:



סעיף ב'

חוקר ב' משתמש ב-ANOVA חד גורמי, זאת בשונה מחוקר א' שמשמש במבחן ANOVA דו גורמי. לכן במעברים בין המבחנים הדברים שיש לחשב משתנים, כלומר $SSR, SSC, SSE, SSInt$ מתחלפים עכשיו בחישוב רק של SSB, SSW וכן בחישוב ה- MS קורה שינוי דומה. לכן בשאלה מה קורה ל- SS ול- MS נניח כי הכוונה היא מה היו הערכים שלהם כאשר אנו מסתכלים על מבחן דו גורמי עם עמודה אחת.

נניח אם כן כי מתקיים $C = dose, R = supp$ כאשר יש לנו רק עמודה אחת (כלומר אין מינונים שונים של ויטמין C כי החוקר לא מעוניין בהם)

מקור השינוי	גדל\קטן\ללא שינוי	
הערך מוכפל במספר העמודות אשר ירד ל-1	קטן	SSR
אנו עוברים למצב בו יש עמודה יחידה, ולכן ממוצע העמודה שווה לממוצע הכללי	מתאפס	SSC
להבנתנו, זהו המקביל של ה- SSW , ולכן לא אמור להשתנות	ללא שינוי	SSE
עפ"י נוסחת החישוב - הסכומים בסוגריים שווים ולכן $(\bar{x}_{j,k} - \bar{x}_{j.}) + (-\bar{x}_{.k} + \bar{x}) = 0$	מתאפס	$SSInt$
מספר הדרכים אשר נבדקו עבור ה- $supp$ לא השתנו	ללא שינוי	df_R
מקטינים את האיבר שמחוסר מ- N	גדל	df_e
בהתאמה ל- $SSC, SSInt$	מתאפסות	df_C, df_{int}
SSR קטן, df_R לא משתנה	קטן	MSR
בהתאמה ל- SSC	מתאפס	MSC
בהתאמה ל- $SSInt$	מתאפס	$MSInt$
SSE לא משתנה, df_e גדל	קטן	MSE
בהתאמה ל- $SSC, SSInt$	מתאפסים	F_C, F_{int}
MSE לא משתנה, MSR קטן	קטן	F_R

סעיף ג'

נערוך את הניתוח של חוקר ב' ב-R.

- תוצאות הניתוח ומסקנת החוקר.

הגשת הפלט:

```
> summary(one_way_aov)
              Df Sum Sq Mean Sq F value Pr(>F)
supp           1    205   205.35   3.668 0.0604 .
Residuals     58   3247    55.98
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

הסבר במילים:

נשווה בין התוצאות מניתוח זה לתוצאות מסעיף ב':

- ערך ה- p - $value$ גדל מאוד כך שממבחן זה לא ניתן לדחות את השערת האפס עבור $\alpha = 5\%$.

- ערך ה- F קטן משמעותית מ-15.57 ל-3.668

- ערכי ה- SS וה- MS לא השתנו כמעט (סביב ה-205)

נשים לב כי לפי מבחן זה **לא** ניתן לדחות את השערת האפס, כלומר לא ניתן להגיד כלום על הקשר בין הדרך שבה נצרך ויטמין C לבין צמיחת השיניים של שרקנים. ההבדלים בתוצאות נובעים משינוי סוג המבחן. כלומר כאשר עברנו למבחן חד גורמי השתנו הערכים של SSW ושל SSB כמו שראינו בסעיף הקודם וזה גרר שהערך F שהתקבל נהיה קטן יותר ויותר קשה היה לדחות את ההשערה.

שאלה 5:

הקוד של זוהר - 2825265

הקוד של אלה - 4331094