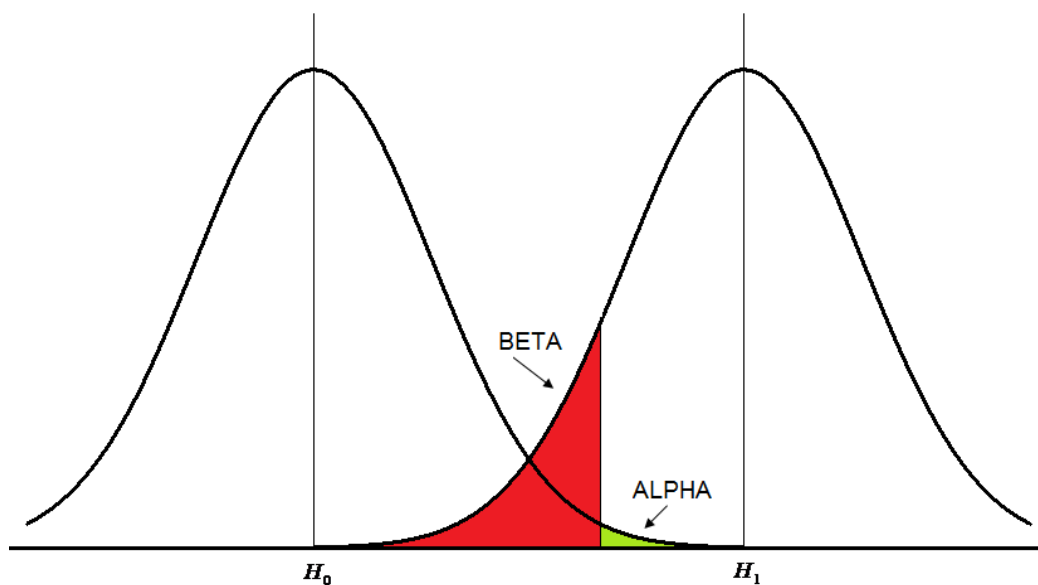


מבוא
לסטטיסטיקה
והסתברות

II

(הסקה סטטיסטית)



ברק קנדל

סטודנטים יקרים

לפניכם ספר תרגילים בנושא סטטיסטיקה והסתברות 2 (הסקה סטטיסטית).

הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה, המועבר ברשת האינטרנט On-line.

הקורס באתר כולל פתרונות מלאים לספר התרגילים, וכן את התיאוריה הרלוונטית לכל נושא ונושא.

הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי, לדוגמה לחצו כאן (www.gool.co.il/statb.html).

את הקורס בנה מר ברק קנדל, מרצה מבוקש במוסדות אקדמיים שונים ובעל ניסיון עתיר בהוראת המקצוע.

אז אם אתם עסוקים מידי בעבודה, סובלים מלקויות למידה, רוצים להצטיין או פשוט אוהבים ללמוד בשקט בבית, אנחנו מזמינים אתכם לחווית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר

www.gool.co.il



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

גול, בשביל התרגול...

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופתר - ברק קנדל ©

תוכן העניינים

פרק 1- מבחנים אפרמטריים (כולל מבחני חי בריבוע) - עמ' 3

מבחן טיב התאמה

מבחן לאי תלות

מבחן הבינום

מבחן הסימן

מבחן וילקוקסון למדגמים מזווגים

מבחן וילקוקסון למדגמים בלתי תלויים

מבחן פישר

מבחן מקנמר

תשובות סופיות

פרק 2 - ניתוח שונות חד כיוונית - עמ' 13

רקע תיאורטי

תרגילים

פתרונות סופיים חלקיים

פרק 3 - מדדי קשר - עמ' 25

מדד הקשר של ספירמן

קשר בין שני משתנים כמותיים- מקדם המתאם הלנארי (מדד הקשר של פירסון ורגרסיה)

תשובות סופיות

לתשומת לבכם, אין מענה לנושאים הבאים:

* מבחן מאן וויטני.

* קרוקאס ווליס.

* הסקה ברגרסייה.

* רגרסייה רבת משתנים.

פרק 1 - מבחנים אפרמטריים

א. מבחן טיב התאמה

1. במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים. התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו- 17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?

2. מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?

3. משרד החינוך טוען שבקרב השכירים במשק היחס בין השכירים בעלי השכלה נמוכה, תיכונית ואקדמאית הוא 1:2:1 בהתאמה. במדגם של 200 שכירים התקבלו 56 אנשים בעלי השכלה נמוכה, 105 בעלי השכלה תיכונית והיתר בעלי השכלה גבוהה. ע"ס תוצאות המדגם האם התפלגות ההשכלה היא כמו שמשרד החינוך מפרסם? בדוק ברמת מובהקות של 5%.

4. בפנס יש 4 סוללות. בבדיקה שנערכה ב-400 פנסים נמצאו סוללות פגומות לפי השכיחויות הבאות:

מספר הסוללות הפגומות	0	1	2	3 ומעלה
שכיחות	276	104	12	8

מעוניינים לבדוק על סמך תוצאות מדגם אלה האם הסיכוי לסוללה פגומה הוא 20%. בדוק ברמת מובהקות של 5%.

ב. מבחן אי תלות

1. במפעל עובד בשלוש משמרות. להלן מספר המוצרים הפגומים והתקינים בכל אחת מן המשמרות לפי מדגם שנעשה :

	יום	ערב	לילה
פגומים	50	60	70
תקינים	600	700	800

האם קיים קשר בין טיב המוצר למשמרת שלו? הסיקו עבור רמת מובהקות $\alpha = 0.05$.

2. בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.

3. בחנות בגדים A בדקו את התפלגות הצבעים של הבגדים הנמכרים ביום מסוים :

צבע	שחור	לבן	אדום	כחול
מספר הפריטים	15	20	15	50

כמו כן בדקו את התפלגות הצבעים בחנות שכנה B :

צבע	שחור	לבן	אדום	כחול
מספר הכדורים	60	20	10	20

- א. בדוק ברמת מובהקות של 5% האם התפלגות הצבעים בחנות A היא ביחס של 3:1:1:1 לטובת הכחול.
- ב. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החנויות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

ג. מבחן הבינום

1. במטרה לבדוק האם מטבע הוא הוגן הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל 7 כולל יוחלט שהמטבע הוגן, אחרת נחליט שהמטבע לא הוגן.
 - א. רשמו את השערות המחקר.
 - ב. מה ההסתברות לטעות מסוג ראשון?
 - ג. מהי עצמת המבחן אם במציאות אכן המטבע אינו הוגן כי הסיכוי לעץ בו הוא 20%?
2. אחוז אוכלי הגלידות בחודשי החורף הנו 30%, קיים חשש כי השנה פחת אחוז אוכלי הגלידות. לשם כך נדגמו 15 אנשים אשר מתוכם 3 אכלו גלידה בחודשי החורף.
 - א. רשום את השערות.
 - ב. מה תהיה ההחלטה ברמת מובהקות של 10%?
3. הסיכוי לזכות במשחק מזל מסוים הינו 70% מעוניינים לבדוק האם שיטה מסוימת מעלה את סיכויי ההצלחה לצורך כך מחליטים לשחק את המשחק 20 פעמים. מחליטים שאם מספר הזכיות יהיה לפחות 18 נקבל את הטענה שאכן השיטה עובדת והיא מעלה את סיכויי ההצלחה להיות 90%. מה הסיכוי לבצע טעות מסוג ראשון ומה הסיכוי לבצע טעות מסוג שני?
4. אחוז התומכים בהצעה מסוימת לפני כחמש שנים היה 40%. מעוניינים לבדוק את הטענה שאחוז התמיכה בהצעת החוק כיום אף ירד. לצורך כך דגמו 10 אנשים מתוכם אחד תמך בהצעת החוק. מהי מסקנתכם ברמת מובהקות של 5%?

ד. מבחן הסימן

1. רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב. עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב. להלן התוצאות שהתקבלו:

א	ב
62	70
74	80
68	70
94	90
82	77
67	67
65	80
84	86
78	79
80	82

בידקו ברמת מובהקות של 5%.

2. מעוניינים לבדוק האם סם מסוים משפיע על לחץ הדם. נלקחו 24 אנשים אשר נמדד להם לחץ הדם לאחר מכן ניתן להם הסם ושוב מדדו להם את לחץ הדם. לחמישה אנשים לחץ הדם לא השתנה ל 15 אנשים לחץ הדם עלה וליתר לחץ הדם ירד אחרי לקיחת הסם. מה מסקנתכם ברמת מובהקות של 5%?

3. מעוניינים לבדוק האם ההוצאות על "גיאנק פוד" בקרב הסטודנטים רבות יותר בזמן הלימודים לעומת ימי החופשה. נדגמו 15 סטודנטים מקריים, אצל 13 ההוצאות בתקופת הלימודים היו גבוהות יותר מימי החופשה ואצל 2 נמוכות יותר. מה מסקנתך בר"מ של 0.05 ?

ה. מבחן ויילקוקסון למדגמים מזווגים:

1. שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה ע"י שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים.

להלן הציונים שהתקבלו:

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

ברמת מובהקות של 5% האם אפשר לקבוע שאופה א' טוב יותר מאופה ב'?

2. סטודנטים נתבקשו לתת חוות דעתם על רמת הקושי של הקורס (סקאלה של 1-5 כאשר 5=קשה ביותר) ועל רמת הקושי של הבחינות באותה סקאלה. מעוניינים לבדוק האם קיים הבדל בין רמת הקושי הנתפסת בעיני הסטודנט על הקורס ועל המבחנים.

להלן תוצאות המדגם:

1- קושי קורס	4	3	2	4	3	2	1	5	4
2 - קושי בחינה	4	4	4	3	2	2	2	3	2

בדקו ברמת מובהקות של 5%.

1. מבחן ווילקוקסון למדגמים בלתי תלויים

1. מעוניינים להשוות בין 2 קבוצות כדורסל. נלקחו 5 משחקים מקבוצה א ושישה משחקים מקבוצה ב'. נבדק בכל משחק ועבור כל קבוצה מה מספר הנקודות שצברה בסוף המשחק.

קבוצה ב	קבוצה א
82	68
74	82
82	78
64	94
67	87
65	

בדקו ברמת מובהקות של 5% את הטענה שקבוצה ב קולעת פחות בממוצע מקבוצה א'.

2. מעוניינים לבדוק האם קורס קיץ באנגלית משפר את יכולות האנגלית לתלמידי חטיבת ביניים. נלקחו 20 ילדים בגיל חטיבת הביניים ברמת אנגלית דומה. 12 מהם נשלחו לקורס קיץ והיתר לא. בסוף הקיץ כולם נבחנו במבחן באנגלית הציון הגבוה ביותר התקבל בקרב אחד שלא עשה את הקורס ושבעת הציונים הנמוכים ביותר היו גם בקרב תלמידים שלא עשו את הקורס. מה המסקנה ברמת מובהקות של 5%?

3. במחקר לבדיקת יעילות ויטמין C נבחרו 15 מתנדבים מבין עובדי המפעל. תשעה מהם נבחרו מקרית וקיבלו טיפול שוטף בויטמין C, ואילו שאר המתנדבים (קבוצת הביקורת) קבלו גלולת סוכר. במשך שלוש שנות המחקר היו מספר ימי ההיעדרות בגלל ההצטננות: קבוצת הטיפול: 12, 16, 8, 0, 4, 3, 9, 3, 1 קבוצת הביקורת: 12, 23, 13, 28, 7, 19. בידקו ברמת מובהקות של 5% שמספר ימי המחלה במשך שלוש שנים מצטמצם ביותר מ-4 ימים עם לקיחת ויטמין C.

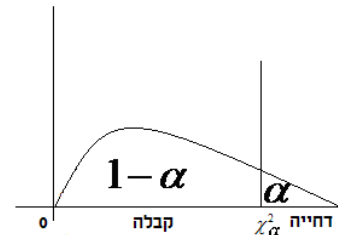
ז. מבחן פישור

1. מעוניינים לבדוק האם שיעורי עזר יעילים בשיפור ההישגים . נלקחו 2 כיתות בנות 15 תלמידים כל אחת. בכיתה אחת נתנו שיעורי עזר ובכיתה שנייה לא נתנו שיעורי עזר. בכיתה שנתנו שיעורי עזר 1 נכשל ובכיתה ללא שיעורי עזר 3 נכשלו. מה המסקנה ברמת מובהקות של 5%?
2. מעוניינים לבדוק האם תרופה אנטידכאונית מסוימת משפיעה על מצב הרוח. נלקחו 28 אנשים שהתלוננו על דיכאון ברמה דומה הם חולקו באקראי לשתי קבוצות, כך ש-16 קיבלו התרופה והיתר היוו קבוצת ביקורת וקבלו פלסיבו. כעבור 3 חודשים נבדק מצבם הנפשי. בקרב לוקחי התרופה רק 2 התלוננו על דיכאון ובקרב לוקחי הפלסיבו 6. מה המסקנה ברמת מובהקות של 10%?

ח. מבחן מקנמר

מתי משתמשים במבחן זה ?

1. המדגם מזווג.
2. המשתנה הנחקר דיכוטומי (משתנה המקבל שני ערכים בלבד).
3. השערה דו צדדית.
4. מספר התצפיות עם שינוי לפחות 20.



כלל ההכרעה: מתבסס על התפלגות χ^2 ודרגת חופש $df = 1$ דחיה תמיד בצד ימין הסטטיסטי:

$$\chi^2 = \frac{(B-C)^2}{B+C} \quad \text{כאשר } B, C - \text{שכיחויות שבהן חל שינוי.}$$

במבחן מקנמר כמו במבחן הסימן מתעלמים מהתצפיות שלא חל בהן שינוי
אם תנאי 3 או 4 לא מתקיימים ניתן לבצע מבחן הסימן במקום.

שאלות:

1. פוליטיקאי הופיע אמש בטלוויזיה והוא מעוניין לבדוק האם התוכנית השפיעה על אמון הציבור בו. לצורך כך בוצע סקר שבו נשאל הצופה האם הוא תפס את הפוליטיקאי כאמין לפני התוכנית והאם הוא תפס אותו אמין לאחר התוכנית.
להלן התוצאות שהתקבלו: (המספרים מייצגים מספר צופים)

לפני			
לא	אמין		
21	12	אמין	אחרי
17	7	לא אמין	

מה המסקנה ברמת מובהקות של 5%?

2. חברת משקאות יצאה בקמפיין שנוי במחלוקת. החברה מעוניינת לבדוק האם הקמפיין השפיע על הרגלי הצריכה. במחקר שבו השתתפו 50 נשאלים 30 טענו שלא שינוי את הרגלי הצריכה. 15 טענו שהחלו לרכוש את המשקה בעקבות הקמפיין ו-5 טענו שהפסיקו לרכוש את המשקה בעקבות הקמפיין.

א. מה מסקנתכם ברמת מובהקות של 2.5%?

ב. בדוק ברמת מובהקות של 5% שהקמפיין היטיב עם החברה מבחינת הרגלי הצריכה.

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופתר - ברק קנדל ©

תשובות סופיות - מבחנים אפרמטריים

פרק א' - מבחן טיב התאמה

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נקבל H_0
<u>שאלה 3</u>	<u>שאלה 4</u>
נקבל H_0	נדחה H_0

פרק ב' - מבחן לאי תלות

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נדחה H_0
<u>שאלה 3</u>	
א. נקבל H_0	
ב. נדחה H_0	

פרק ג' - מבחן הבינום

<u>שאלה 1</u>	<u>שאלה 2</u>
ב. 0.0781	נקבל H_0
ג. 0.1678	
<u>שאלה 3</u>	<u>שאלה 4</u>
א. 0.0355	נדחה H_0
ב. 0.3231	

פרק ד' - מבחן הסימן

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נדחה H_0
<u>שאלה 3</u>	
נדחה H_0	

פרק ה' - מבחן ווילקוקסון למדגמים מזווגים

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה H_0	נקבל H_0

פרק ו' - מבחן ווילקוקסון למדגמים בלתי תלויים

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה H_0	נדחה H_0
<u>שאלה 3</u>	
נקבל H_0	

פרק ז' - מבחן פישר

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל H_0	נדחה H_0

פרק ח' - מבחן מקנמר

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה H_0	א. נקבל H_0
	ב. נדחה H_0

פרק 2 - ניתוח שונות חד ביוונית

רקע תיאורטי:

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $(\mu_1, \dots, \mu_k)$ של k אוכלוסיות שונות.

ולכן בניתוח שונות השערות המחקר הן:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad (\text{התוחלות של כל האוכלוסיות שוות})$$

$$H_1: \quad \text{אחרת} \quad (\text{לפחות שתיים מהתוחלות שונות})$$

ההנחות הדרושות לביצוע התהליך הן:

1. בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

2. כל האוכלוסיות הן עם אותה שונות σ^2 .

3. 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 .

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

$$\bar{X}_1 - \text{ממוצע המדגם הראשון}, \dots, \bar{X}_k - \text{ממוצע המדגם ה-}k.$$

$$\bar{X} - \text{ממוצע כללי (של כל המדגמים)}.$$

$$SS_B = \sum_{i=1}^k n_i [\bar{X}_i - \bar{X}]^2 \quad \text{סכום ריבועים בין הקבוצות}$$

$$SS_W = \sum_{i=1}^k [n_i - 1] \cdot \hat{S}_i^2 \quad \text{סכום ריבועים בתוך הקבוצות}$$

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

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה :

טבלת ניתוח שונות

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		

$$F = \frac{SS_B / (k-1)}{SS_W / (n-k)} \sim F(k-1, n-k)$$

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

תרגילים:

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 1 קיבלה "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:

בית הספר	"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?

4. שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.

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

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות? בדקו ברמת מובהקות של 5%.

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.

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			

מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

6. להלן טבלת ANOVA בטבלה הושמטו חלקים. השלם 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			

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 נבדק 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			

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.

Homogeneous Subsets

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.

- א. בדוק 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. נמק!
- ד. לטבלת ה ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

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			

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	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	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.

פתרונות סופיים חלקיים - ניתוח שונות חד כיוונית

2. אם חישבת נכון ה F הסטטיסטי יוצא: 0.58

3. נדחה את השערת האפס.

4. להלן טבלת הניתוח השונות המתקבלת:

	Sum of Squares	df	Mean Square	F
Between Groups	126.000	2	63.000	.756
Within Groups	750.000	9	83.333	
Total	876.000	11		

5. נקבל את השערת האפס.

6. א. 165.5 ב. 2 ג. 178.725 ד. 9.375 ה. 18.36

7. א. נדחה את השערת האפס. ב. לא משתנה. ג. כן

8. א. 5 ב. 25 ג. כן

פרק 3 - מדדי קשר

מדד הקשר של ספירמן:

1. בתוכנית "רוקדים עם כוכבים" שני שופטים נתנו ציון ל-7 הזוגות שהשתתפו בתוכנית:

מספר זוג	1	2	3	4	5	6	7
ציון שופט א'	7	8	6	8	9	5	6
ציון שופט ב'	8	8	7	8	9	5	7

האם קיים קשר בין שתי הערכות השופטים? נמק והסבר!

2. משרד רצה לבחון האם קיים קשר בין מידת המוטיבציה של העובדים שלו לבין מספר החיסורים של העובדים בחודש עבודה. להלן התוצאות שהתקבלו:

מספר חיסורים	מידת מוטיבציה
0	גבוהה
4	נמוכה
2	בינונית
5	נמוכה
1	גבוהה

האם קיים קשר בין רמת המוטיבציה של העובד ומספר החיסורים שלו? חשב באמצעות מדד הקשר המתאים והסבר.

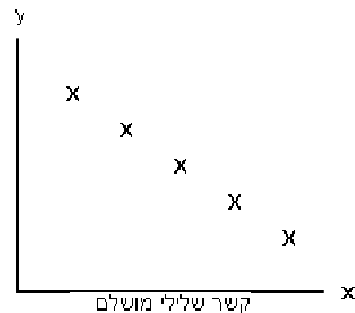
3. אם $r_s = 1$ הדבר אומר שערכי X תמיד שווים לערכי Y. האם הטענה נכונה? הסבר.

מדד הקשר הלינארי - קשר בין שני משתנים כמותיים

רקע

המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X מסבירה את ההכנסה שלו Y . במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו.

בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים.



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבדוק עד כמה קיים קשר לינארי בין שני המשתנים. הממד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שניראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

מקדם מתאם זה מקבל ערכים בין -1 ל 1 .

מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה: $y = bx + a$.

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לינארי מלא בו השיפוע b יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לינארי מלא בו השיפוע b שלילי (מקדם מתאם -1).

מתאם חיובי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר.



מקדם המתאם יסומן באות r .

כדי לחשב את מקדם המתאם, יש לחשב את סטיות התקן של כל משתנה ואת השונות המשותפת.

$$\text{cov}(x, y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} : \text{שונות משותפת}$$

$$\hat{S}_x^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} : \text{שונות של המשתנה } X$$

$$\hat{S}_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} : \text{שונות המשתנה } Y$$

$$r_{x,y} = \frac{\text{cov}(x, y)}{\hat{S}_x \cdot \hat{S}_y} : \text{מקדם המתאם הלינארי}$$

בשלב השלישי, במידה וקיים קשר חזק בין שני המשתנים נהוג לבצע ניבויי לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר.

מדובר בקו שמנבא את Y על סמך X . השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים.

a- בעצם נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא נקרא החותך של הקו.

b- הוא שיפוע הקו נותן בכמה בעצם Y משתנה כאשר X גדל ביחידה אחת על גבי קו הניבויים. להלן המשוואות למציאת הפרמטרים של קו הרגרסיה:

$$\tilde{Y} = bX + a$$

$$b = r \frac{\hat{S}_y}{\hat{S}_x}$$

$$a = \bar{Y} - b\bar{X}$$

טרנספורמציה לינארית והשפעתה על מקדם המתאם

טרנספורמציה לינארית בין אם נעשית על X ובין אם נעשית על y , או בין אם נעשית על שניהם, אינה משנה את עוצמת הקשר. היא עלולה רק לשנות את כיוונו אם השיפועים של שתי הטרנספורמציות שוני סימן.

תרגילים:

1. להלן נתונים לגבי שישה תלמידים שנגשו למבחן. בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

מספר חיסורים	ציון
2	80
1	90
0	90
2	70
3	70
4	50

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?
- ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?
- ג. הסבר ללא חישוב כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

2. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200 \quad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \quad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) = 60.8$$

- א. חשב את מדד הקשר הלינארי בין X ל- Y . מיהו המשתנה התלוי?
- ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבר את משמעות הפרמטרים של קו הרגרסיה.
- ג. משפחת כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

3. נסמן ב- X את ההשכלה של אדם בשנות למוד. נסמן ב- Y את הכנסתו באלפי ₪.

במחקר התקבלו התוצאות הבאות :

$$\hat{S}_y = 5 \quad \hat{S}_x = 2$$

$$\bar{Y} = 8 \quad \bar{X} = 14$$

$$COV(X, Y) = 7.5$$

א. חשב את מדד הקשר של פירסון בין ההשכלה להכנסה.

ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?

ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000 ₪?

ד. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

4. מבחן בנוי מחלק כמותי ומחלק מילולי. ממוצע הציון בחלק המילולי הנו 100 עם סטית תקן

20 ובחלק הכמותי ממוצע הציון 90 עם סטיית תקן 15. מקדם המתאם בין שני הציונים הוא 0.9.

א. חשבו את השונות המשותפת בין ציוני שני חלקי הבחינה.

ב. אם יעלו את כל הציונים בחלק המילולי ב- 20%, מה יהיה מקדם המתאם בין

הציון המילולי החדש לציון הכמותי ובין הציון המילולי הישן לציון המילולי

החדש?

ג. נגדיר משתנה חדש W להיות המרחק של הציון בחשיבה מילולית מהציון

המקסימאלי בבחינה- 150. מצא את מקדם המתאם בין הציון המילולי ל- W ובין

W ל-ציון הכמותי.

5. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק!

א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם

מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר

הדירה בדולרים הוא יקבל 1.

ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ $S_x = S_y = 1$ לכן מדד הקשר של פירסון

יהיה 1.

ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של

פירסון יהיה 0.

6. במחקר רצו לבדוק את הקשר בין גודל המנוע של מכונית (Engine size) לבין כוח הסוס שלו (Horsepower) להלן הפלט שהתקבל:

Descriptive Statistics			
	Mean	Std. Deviation	N
Horsepower	185.95	56.700	156
Engine size	3.061	1.0447	156

Correlations			
		Horsepower	Engine size
Horsepower	Pearson Correlation	1	.837**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	498313.590	7687.787
	Covariance	3214.926	49.599
	N	156	156
Engine size	Pearson Correlation	.837**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	7687.787	169.151
	Covariance	49.599	1.091
	N	156	156

** . Correlation is significant at the 0.01 level (2-tailed).

- א. מצא את קו הניבויים לניבוי כוח הסוס על סמך גודל המנוע.
- ב. מה יהיה הניבוי לכוח הסוס של מכונית עם גודל מנוע 3?
- ג. איזו התפלגות יותר הומוגנית. זו של התפלגות כוח הסוס של המכונית או זו של גודל המנוע? הסבר באמצעות חישוב.

7. במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון X

בדם החולה לרמת ההורמון Y שלו. לצורך כך מדדו את רמת

ההורמונים ההלו עבור חמישה חולים.

להלן הפלט שהתוכנה הוציאה :

Descriptive Statistics		
	Mean	N
X	16.00	5
Y	15.40	?

Correlations			
		X	Y
X	Pearson Correlation	1	?
	Sum of Squares and Cross-products	44.000	49.000
	Covariance	11.000	?
	N	5	5
Y	Pearson Correlation		1
	Sum of Squares and Cross-products	49.000	59.200
	Covariance	12.250	14.800
	N	5	5

א. מלא את המספרים החסרים בפלט (במקום סימני השאלה).

ב. מצא את קו הניבוי לניבוי ההורמון Y על סמך ההורמון X.

ג. התברר ונפלה טעות ויש להוסיף 1 לכל ערכי הX. חזור על סעיף הקודם לאחר השינוי.

תשובות**מדד הקשר של ספירמן****שאלה 1**

0.973

שאלה 3

לא נכון

שאלה 2

-0.85

מדד הקשר של פירסון - מדד הקשר הלינארי**שאלה 1**

א. בהקלטה

ב. -0.9325

שאלה 2

א. 0.8

ב. $\tilde{Y} = 0.8X + 0.4$

ג. 12.4

שאלה 3

א. 0.75

ב. 4.25 אלפי ש"ח

שאלה 4

א. 270

ב. בין הציון המילולי הישן לחדש: 1

בין הציון המילולי החדש לכמותי: 0.9

ג. בין W ל ציון המילולי: -1

בין W לציון הכמותי: -0.9

שאלה 5**שאלה 6**א. $\tilde{Y} = 45.43X + 46.89$

ב. 183.18

ג. כוח סוס

ג. לא נכון

ד. נכון

שאלה 7

א. בהקלטה

ב. $\tilde{Y} = 1.11X - 2.36$