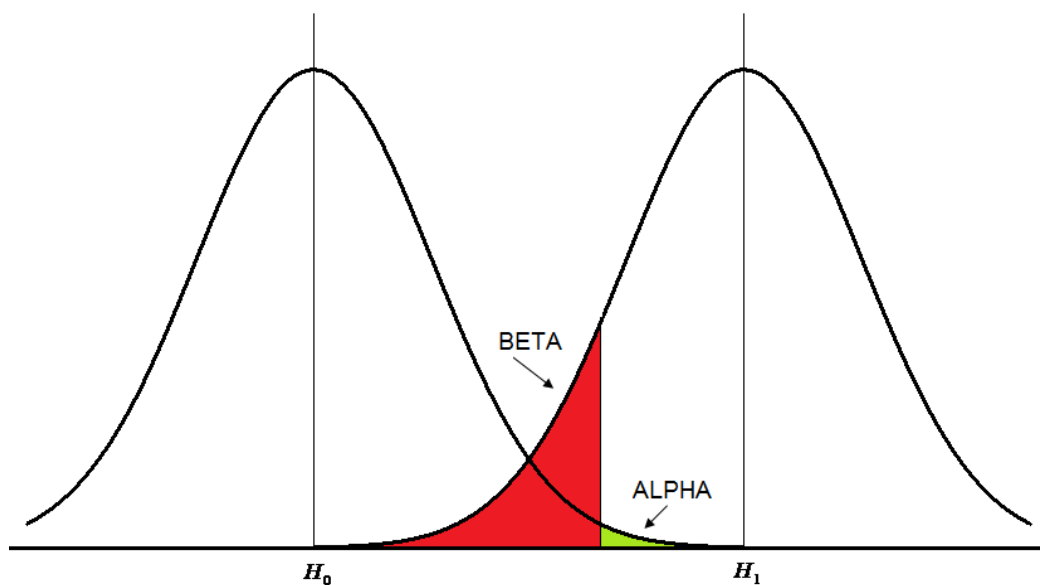


מבוא
לסטטיסטיקה
והסתברות

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 לחוויית 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 ראשון - סטטיסטיקה תיאורית – עמ' 4

הצגת נתונים
מדדי מיקום מרכזי
מדדי פיזור
מדדי מיקום יחסי
ממוצע משוקלל ושונות מצורפת
טרנספורמציה לינארית
שאלות מסכמות
תשובות סופיות

פרק שני - התפלגות נורמלית – עמ' 17

ההתפלגות הנורמלית
תשובות סופיות

פרק שלישי - רגרסיה לינארית – עמ' 20

מדד הקשר של פירסון
תשובות סופיות

פרק רביעי – התפלגות הדגימה – עמ' 23

התפלגות ממוצע המדגם
תשובות סופיות

פרק חמישי – רווחי סמך – עמ' 28

אמידה נקודתית (אומדים חסרי הטיה)
רווח סמך לתוחלת (שונות ידועה/שונות לא ידועה)
תשובות סופיות

פרק שישי – בדיקת השערות – מבחנים פרמטרים – עמ' 31

בדיקת השערות על תוחלת
בדיקת השערות הפרש תוחלות במדגמים בלתי תלויים
בדיקת השערות במדגמים מזווגים
הגישה המעשית-מובהקות התוצאה p-value
טעויות ועוצמה
הקשר בין רווח סמך לבדיקת השערות
תשובות סופיות

פרק שביעי - תרגול שאלות אמריקאיות על כל החומר - עמ' 40

פרק שמיני - ניתוח שונות חד כיוונית - עמ' 56

תרגילים

פתרונות סופיים חלקיים

פרק ראשון – סטטיסטיקה תיאורית

א. הצגת נתונים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא :

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6

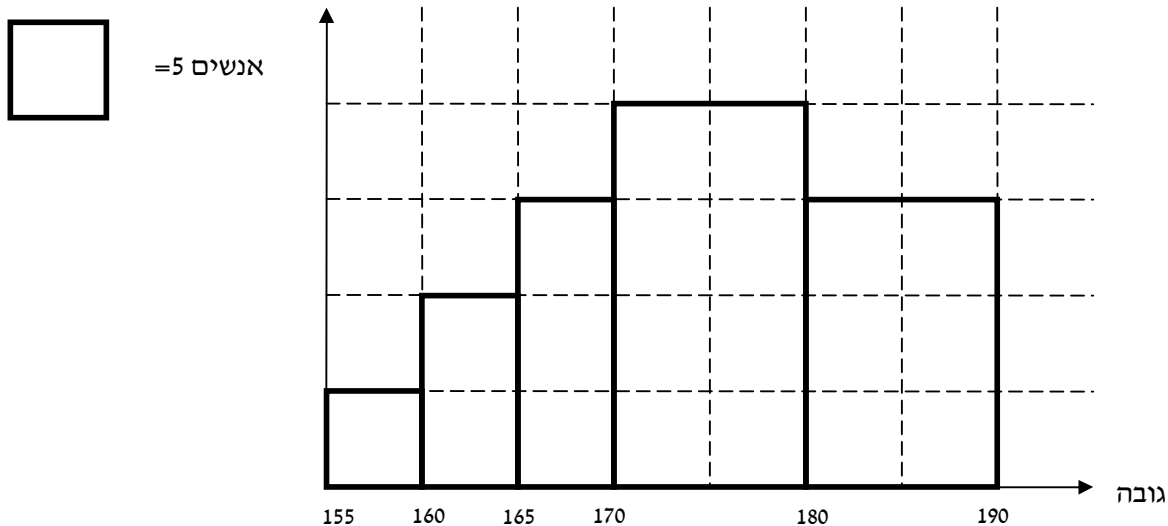
א. מהו המשתנה? האם הוא בדיד או רציף?

ב. תאר את הרשימה בטבלת שכיחויות.

ג. הוסף שכיחויות יחסיות לטבלה.

ד. תאר את הנתונים באופן גרפי.

2. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת :



א. מהו המשתנה הנחקר? האם הוא בדיד או רציף?

ב. תאר את הנתונים בטבלת שכיחויות במחלקות.

ג. הוסף שכיחות יחסית לטבלה.

ד. הוסף את הצפיפות של כל מחלקה לטבלה.

ה. מהי צורת ההתפלגות של הגבהים?

3. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

- א. תאר את ההתפלגות באופן גרפי.
 ב. מה ניתן להגיד על צורת ההתפלגות?

ב. מדדי מיקום מרכזי:

1. חשב את החציון, השכיח, אמצע הטווח והממוצע של שאלה 1 בפרק הקודם.
 2. על פי הנתונים של שאלה 2 בפרק הקודם:
 א. חשב את החציון, השכיח, אמצע הטווח והממוצע של הגבהים.
 ב. האם היה ניתן לדעת על סמך הגרף בלבד האם הממוצע יהיה קטן מהחציון?
 3. להלן התפלגות מספר מקלטי הטלויזיה שנספרו עבור כל משפחה ביישוב מסוים:

מספר משפחות	מספר מקלטים
22	0
28	1
18	2
22	3
10	4

- א. חשב את הממוצע, החציון, השכיח ואמצע הטווח של ההתפלגות.
 ב. הסבר ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלויזיה היו רוכשים מקלט אחד.
 4. על סמך הגרף בלבד וללא חישוב מפורש הסבר מה הם כל מדדי המרכז של שאלה 3 בפרק הקודם.

ג. מדדי פיזור:

1. חשב את השונות, סטיית התקן והטווח עבור הנתונים משאלה 1 פרק א'.
2. עבור הנתונים משאלה 2 פרק א' :
 - א. חשב את השונות, סטיית התקן.
 - ב. חשב את הטווח והטווח הבינרבעוני.
 - ג. אם יתווספו עוד 10 אנשים למחלקה שבה הגובה בין 155 ס"מ ל-160 ס"מ. כיצד ישתנו המדדים הבאים: הממוצע, החציון, השונות והטווח. הסבר ללא חישוב!
3. עבור נתוני שאלה 3 פרק ב' חשב את סטיית התקן.
4. נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע: 2, 3, 2, 1. חשב את השונות של חמש התצפיות.

ד. מדדי מיקום יחסי:

1. תלמידי כיתה ח' נגשו למבחן בלשון ולמבחן במתמטיקה. להלן התוצאות שהתקבלו:

המקצוע	ממוצע	סטיית תקן
לשון	74	12
מתמטיקה	80	16

עודד קיבל: 68 בלשון ו-70 במתמטיקה.

- א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?
- ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

2. בהמשך לנתוני שאלה 2 פרק א' חשב את :

- א. העשירון התחתון.
- ב. האחוזון ה-30.
- ג. הגובה ש-20% מהתצפית גדולות ממנו.
- ד. את אחוז התצפיות מתחת לגובה 158 ס"מ.
- ה. את אחוז התצפיות מעל לגובה 185 ס"מ.
- ו. את אחוז התצפיות בין גובה 170 ס"מ ל-185 ס"מ.

ה. ממוצע משוקלל ושונות מצורפת:

1. להלן נתונים לגבי ציונים במבחן באנגלית ב-3 כיתות מתוך שכבה י' בתיכון :

כיתה	ממוצע	מס' תלמידים	סטיית תקן
1	76	40	12
2	68	20	15
3	82	30	10

- א. חשב את הממוצע המשוקלל בשכבה.
- ב. חשב את השונות המצורפת בשכבה.

2. נתונות שתי קבוצות :

- בקבוצה I פי שתיים תצפיות מאשר בקבוצה II.
- הממוצע בשתי הקבוצות הוא 70.
- השונות בקבוצה I היא 100.
- השונות בקבוצה II היא 400.

- א. מצא את הממוצע של התצפיות לאחר שאוחדו שתי הקבוצות לקבוצה אחת.
- ב. מצא את סטיית התקן של התצפיות לאחר שאוחדו שתי הקבוצות לקבוצה אחת.

ו. טרנספורמציה לינארית:

1. עבור סדרת נתונים התקבל:

$$\bar{X} = 80$$

$$S = 15$$

$$MO = 70$$

הוחלט להכפיל את כל התצפיות פי-4 ולהחסיר מהתוצאה 5. חשב את המדדים הללו לאחר השינוי.

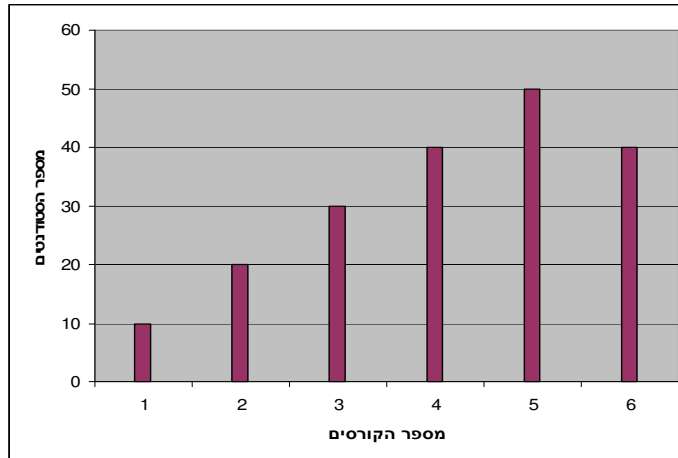
2. בחברה מסוימת השכר הממוצע הוא 40 ₪ לשעה עם סטיית תקן של 5 ₪ לשעה. הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן הם קיבלו לאחר מכן תוספת של 2 ₪ לשעה. מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.

3. בהמשך לשאלה 2 פרקים א', ב', ג', ו-ד'. הסתבר שנפלה טעות והוספו 5 ס"מ לכל התצפיות. מה הם החציון, השונות, הטווח והעשירון התחתון של התצפיות ללא הטעות.

4. בהמשך לשאלה 2 פרקים א', ב', ג', ו-ד'. מעוניינים להמיר את הגובה מס"מ למטרים. כיצד הדבר משפיע על החציון, השונות, הטווח והעשירון התחתון.

ז. שאלות מסכמות:

1. בפקולטה להנדסה אספה מזכירות הסטודנטים נתונים לגבי מס' הקורסים שכל סטודנט סיים בשנה הראשונה ללימודיו בשנת 2008.
להלן התוצאות שהתקבלו:



- א. מה המשתנה הנחקר? האם הוא בדיד או רציף?
 ב. מהי צורת ההתפלגות?
 ג. תאר את הנתונים בטבלת שכיחויות.
 ד. חשב את השכיח, החציון וטווח הציונים.

2. להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'. השתתפו במחקר 150 תלמידים.

$$\text{ממוצע הציונים שהתקבל: } \bar{X} = 7 \frac{1}{15}$$

ציון	מספר התלמידים
4	12
5	16
6	
7	38
8	
9	14
10	10

- א. השלם את השכיחויות החסרות בטבלה.
- ב. חשב את הציון החציוני, השכיח ואמצע הטווח.
- ג. חשב שונות וסטיית תקן להתפלגות הציונים.
- ד. הוחלט לשנות את סקלת הציונים ולהכפיל את הציון ב-10. למשל, ציון 8 יהפוך להיות 80. מה הממוצע ומהי השונות של הציונים בסקלה זו?

3. להלן התפלגות מספר שעות העבודה לשבוע של העובדים בחברת "סטאר".
בחברה 200 עובדים.

שכיחות	שכיחות יחסית (פרופורציה)	מספר שעות עבודה
	15%	10-20
	20%	20-30
	30%	30-40
	20%	40-50
		50-60

- א. השלם את הטבלה.
- ב. חשב את החציון, השכיח, והממוצע של התפלגות מס' שעות העבודה בחברה.
- ג. מהי סטיית התקן של מס' שעות העבודה?
- ד. מה העשירון העליון של ההתפלגות?
- ה. איזה אחוז מהעובדים עובדים מעל 45 שעות בשבוע?
- ו. מה ציון התקן של רינה שעובדת 30 שעות בשבוע?
- ז. כיצד ישתנה החציון, הממוצע וסטיית התקן אם מספר שעות העבודה המינימאלי אינו 10 אלא 15? הסבר.

4. חברה סלולארית דגמה 200 אנשים. עבור כל אדם נבדק מס' המסרונים ששלח במשך חודש. להלן ההתפלגות שהתקבלה:

מספר האנשים	מספר המסרונים
40	0-50
60	50-100
50	100-150
30	150-250
20	250-ומעלה

- א. מה אחוז האנשים ששלחו פחות מ-80 מסרונים בחודש?
- ב. מה אחוז האנשים ששלחו בין 50 ל-120 מסרונים?
- ג. הוחלט להעניק מתנה עבור $\frac{1}{4}$ מהלקוחות שמשתמשים במספר הרב ביותר של מסרונים בחודש. החל מאיזה כמות של מסרונים תחולק המתנה?
- ד. ציינו איזה מדד ניתן לחשב ואיזה לא ניתן. אם ניתן חשב:
1. ממוצע
 2. שכיח
 3. חציון
 4. שונות

5. בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.

- א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?
- ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?
- ג. השכר בחברה מחושב על סמך הוותק. שכר בסיס של 5,000 ₪ ו-500 ₪ תוספת על כל שנת וותק לחודש. מה השכר הממוצע ומהי סטיית התקן של השכר לחודש?

6. להלן מספר טענות , עבור כל טענה ציין אם היא נכונה או לא נכונה ונמקו .

- א. בסדרה שבה כל התצפיות שוות זו לזו השונות הינה 0.
- ב. ציון התקן של החציון תמיד יהיה 0.
- ג. ציון התקן של האחוזון ה-70 בהתפלגות אסימטרית ימנית (חיובית) תמיד יהיה חיובית.
- ד. אם נוסיף תצפיות לסדרה של תצפיות, הדבר בהכרח יגדיל את הממוצע של הסדרה.
- ה. בסדרה החציון הינו 80. הוספו שתי תצפיות אחת 79 ואחת 100 לכן החציון יגדל.
- ו. אם נוסיף את הערך 4 לכל התצפיות אז סטיית התקן לא תשתנה.
- ז. אם נחלק את כל התצפיות בהתפלגות ב-2 אז השונות תקטן פי 2.
- ח. אם נגדיל את ממוצע המשכורות של עובדים בחברה אז גם השונות תגדל.

ט. תשובות סופיות לסטטיסטיקה תיאורית

פרק א' – הצגת נתונים

שאלה 1

א. בדיד

שאלה 2

א. רציף

ה. התפלגות אסימטרית
שמאלית (שלילית)

שאלה 3

א. התפלגות סימטרית

פרק ב' – מדדי מיקום מרכזי

שאלה 1

החציון: 7

השכיח: 6

אמצע הטווח: 7

הממוצע: 6.9

שאלה 2

א. החציון: 175

השכיח: 175

אמצע הטווח: 172.5

הממוצע: 174.75

שאלה 3

א. הממוצע: 1.7

החציון: 1.5

השכיח: 1

אמצע הטווח: 2

שאלה 4

הממוצע: 55

החציון: 55

אמצע הטווח: 55

השכיח: 47.5 ו-62.5

ב. הממוצע יגדל ויתר המדדים לא ישתנו.

פרק ג' – מדדי פיזור

שאלה 1

השונות: 2.19

סטיית תקן: 1.48

טווח: 6

שאלה 2

א. השונות: 69.3

סטיית תקן: 8.325

ב. טווח: 35

טווח בינרבעוני: 13.33

ג. ממוצע יקטן, חציון יקטן, שונות תגדל

וטווח לא ישתנה

שאלה 3

תשובה: 1.3

שאלה 4

תשובה: 10.8

פרק ד' – מדדי מיקום יחסי**שאלה 1**

א. לשון

ב. תשובה: 72

שאלה 2

א. תשובה: 162.5

ב. תשובה: 170

ג. תשובה: 183.33

ד. תשובה: 3%

ה. תשובה: 15%

ו. תשובה: 55%

פרק ה' – ממוצע משוקלל ושונות מצורפת**שאלה 1**

א. תשובה: 76.22

ב. תשובה: 173.5

שאלה 2

א. תשובה: 70

ב. תשובה: 14.14

פרק ו' – טרנספורמציה לינארית**שאלה 1**

הממוצע: 315

סטיית התקן: 60

השכיח: 275

שאלה 2

הממוצע: 46

השונות: 30.25

שאלה 3

החציון: 170

השונות: 69.3

הטווח: 35

שאלה 4

החציון: 1.75

השונות: 0.00693

הטווח: 0.35

העשירון התחתון: 1.625

העשירון התחתון: 157.5

פרק ז' – שאלות מסכמות**שאלה 1**

א. מספר הקורסים. בדיד.

שאלה 2

א. 20 תלמידים קיבלו ציון 6 ו-40 תלמידים קיבלו ציון 8.

ב. התפלגות אסימטרית שמאלית (שלילית)

ד. השכיח: 5

הטווח: 5

החציון: 4

ג. השונות: 2.516

סטטיית התקן: 1.586

ד. הממוצע: 70.67

השונות: 251.6

שאלה 3

ב. החציון: 35

השכיח: 35

הממוצע: 35

ג. סטיית תקן: 12.65

ד. תשובה: 53.333

ה. תשובה: 25%

ו. תשובה: -0.395

ז. חציון לא ישתנה, ממוצע יגדל

וסטיית התקן תקטן.

שאלה 4

א. תשובה: 38%

ב. תשובה: 40%

ג. תשובה: 150

ד. החציון: 100

שאלה 5

א. ממוצע לא ישתנה, סטיית התקן תקטן

ב. ממוצע לא ישתנה, סטיית התקן תגדל

ג. הממוצע: 7000 ₪, סטיית התקן: 1000 ₪

שאלה 6

א. נכון

ב. לא נכון

ג. לא נכון

ד. לא נכון

ה. לא נכון

ו. נכון

ז. לא נכון

ח. לא נכון

פרק שני - התפלגות נורמלית

1. הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ.
 - א. מה אחוז האנשים שגובהם מתחת ל- 182.4 ס"מ?
 - ב. מה אחוז האנשים שגובהם מעל 190 ס"מ?
 - א. מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
 - ב. מה אחוז האנשים שגובהם מתחת ל- 170 ס"מ?
 - ג. מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?

2. המשקל של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 60 ק"ג וסטית תקן של 8 ק"ג.
 - א. מה אחוז האנשים שמשקלם נמוך מ- 55 ק"ג?
 - ב. מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
 - ג. מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל- 70 ק"ג?
 - ד. לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ- 4 ק"ג?
 - ה. מה אחוז האנשים באוכלוסייה הזו ששוקלים מתחת ל- 140 ק"ג?

3. ציוני מבחן אינטליגנציה מתפלג נורמלית עם ממוצע 100 ושונות 225.
 - א. מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - ב. מה העשירון התחתון של ההתפלגות?
 - ג. מהו הציון ש- 20% מהנבחנים מקבלים מעליו?
 - ד. מהו האחוזון ה- 20?
 - ה. מה הרבעון התחתון?

4. אורך חיים של מכשיר מתפלג נורמלית. ידוע שמחצית מהמכשירים חיים פחות מ- 500 שעות, כמו כן ידוע ש- 67% מהמכשירים חיים פחות מ- 544 שעות.
 - א. מהו ממוצע אורך חיי מכשיר?
 - ב. מהי סטית בתקן של אורך חיי מכשיר?
 - ג. מה הסיכוי שמכשיר אקראי יחיה פחות מ- 460 שעות?
 - ד. מהו המאיון העליון של אורך חיי מכשיר?
 - ה. 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?

5. הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטית תקן של 5 דקות.
- א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?
- ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיאחר לעבודתו?

תשובות סופיות התפלגות נורמלית

<u>שאלה 1</u>	<u>שאלה 2</u>
א. 89.25%	א. 26.43%
ב. 2.28%	ב. 89.44%
ג. 0	ג. 39.44%
ד. 50%	ד. 0.383
ה. 50%	ה. 100%
<u>שאלה 3</u>	<u>שאלה 4</u>
א. 119.23	א. 500
ב. 80.77	ב. 100
ג. 112.63	ג. 0.3446
ד. 87.37	ד. 732.6
ה. 89.89	ה. 267.4
<u>שאלה 5</u>	
א. 0.1587	
ב. 0.0228	

פרק שלישי - רגרסיה ליניארית

א. מדד הקשר של פירסון – מדד הקשר הלינארי (כולל רגרסיה ליניארית

פשוטה):

1. להלן נתונים לגבי חמישה תלמידים שנגשו למבחן אמצע סמסטר ולמבחן סוף סמסטר:

ציון אמצע סמסטר	7	6	7	8	8
ציון סוף סמסטר	8	7	7	9	10

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על הקשר בין ציון אמצע סמסטר לבין ציון סוף סמסטר?
- ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?

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

$$\begin{aligned}\sum_{i=1}^{20} Y_i &= 200 & \sum_{i=1}^{20} X_i &= 240 \\ \sum_{i=1}^{20} Y_i^2 &= 2080 & \sum_{i=1}^{20} X_i^2 &= 2960 \\ \sum_{i=1}^{20} X_i Y_i &= 2464\end{aligned}$$

- א. חשב את מדד הקשר הלינארי בין X ל- Y .
- ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה.
- 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:

$$\begin{aligned}S_y &= 5 & S_x &= 2 \\ \bar{Y} &= 8 & \bar{X} &= 14 \\ COV(X, Y) &= 7.5\end{aligned}$$

- 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. איזה אחוז משוונות ההכנסה מוסבר באמצעות ההשכלה של האדם?

4. 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 ש. 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 שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.
- ד. אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 1.
- ה. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

פתרונות**א. מדד הקשר של פירסון – מדד הקשר הלינארי****שאלה 1**

א.

ב. 0.871

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

ג. 14.6 שנים

ד. 56.25%

שאלה 4

א. נכון

ב. לא נכון

ג. לא נכון

ד. נכון

ה. נכון

פרק רביעי - התפלגות הדגימה

א. התפלגות ממוצע המדגם

1. להלן התפלגות מספר מקלטי הטלויזיה למשפחה בישוב מסוים:

מספר המקלטים	מספר המשפחות
0	500
1	2500
2	3500
3	3000
4	500
סך הכול $N = 10000$	

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

- א. בנו את פונקציית ההסתברות של x .
- ב. חשבו את התוחלת, השונות וסטיית התקן של x .
- ג. אם נדגום 4 משפחות מהישוב מה תהיה התוחלת, מהי השונות ומהי סטיית התקן של ממוצע המדגם?

2. אם נטיל קובייה פעמיים ונתבונן בממוצע התוצאות שיתקבלו, מה תהיה התוחלת ומה תהיה סטיית התקן של ממוצע זה?

3. משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.

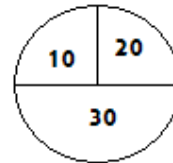
א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?

נתון שביום מסוים נולדו 4 תינוקות.

- ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
- ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
- ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?

4. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועיות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלי.
- א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
- ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה פחות ממנו?
- ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
5. נפח היין בבקבוק מתפלג נורמאלי עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?
- ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?
- ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?
- ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
6. משתנה מתפלג נורמאלי עם תוחלת 80 וסטיית תקן 4.
- א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
- ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
- ג. הסבר את ההבדל בתשובות של שני הסעיפים.

7. בקזינו ישנה רולטה. על הרולטה רשומים המס' הבאים כמוראה בשרטוט :



- אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.
- א. בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד.
- ב. מה התוחלת ומה השונות של סכום הזכייה?
- ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?
- ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה"כ יזכה ב-1050 ₪ ומעלה?
8. לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪. מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?
9. מטילים קובייה 50 פעמים בכל פעם מתבוננים בתוצאה של הקובייה. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72 ב-50 ההטלות?
10. אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ
- א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?
- ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?
- ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. העזר במשפט הגבול המרכזי.

תשובות סופיות להתפלגות הדגימה

פרק א' - התפלגות ממוצע מדגם

שאלה 1

א.

$$E(\bar{X}) = 3.5$$

$$\sigma(\bar{X}) = 1.208$$

4	3	2	1	0	x
0.05	0.3	0.35	0.25	0.05	P(x)

$$\sigma = 0.973 \quad \sigma^2 = 0.9475 \quad \mu = 2.05 \quad \text{ב.}$$

$$V(\bar{X}) = 0.237 \quad E(\bar{X}) = 2.05 \quad \text{ג.}$$

$$\sigma(\bar{X}) = 0.487$$

שאלה 3

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

שאלה 4

א. 0.0465

ב. 32.29

ג. 0.2628

שאלה 6

א. 0.5468

ב. 0.6826

שאלה 5

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

שאלה 8

א. 0.0475

שאלה 7

א.

30	20	10	
0.5	0.25	0.25	P(x)

ב. התוחלת: 22.5

השונות: 68.75

ג. התוחלת: 22.5

השונות: 13.75

ד. 0.8997

שאלה 10

א. 0.9772

ב. 0.0228

ג. 271

שאלה 9

0.1814

פרק חמישי - רווחי סמך

א. אמידה נקודתית (אומדים חסרי הטיה)

1. נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו :
168, 184, 192, 171, 180, 177, 187, 168, 177 ו-175.
- א. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
- ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
- ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.

2. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן

$$\text{התוצאות שהתקבלו : } \sum_{i=1}^{20} X_i^2 = 1502.2 \quad \sum_{i=1}^{20} X_i = 162$$

- א. אמדו את השכר הממוצע של השכירים במשק.
- ב. אמדו את סטיית התקן של שכר השכירים במשק.

ב. רווח סמך לתוחלת

ב.1. שונות האוכלוסיה ידועה

1. מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
- א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
- ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
- ג. הסבר כיצד ומדוע השתנה רווח הסמך.
- ד. מהי טעות התקן של האומד?

2. בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (99,105). שחזרו את :
 - א. ממוצע המדגם.
 - ב. שגיאת האמידה המקסימאלית.
 - ג. רמת הסמך.

3. מעוניינים לאמוד את המשקל הממוצע של רכיב מסוים. ידוע שהמשקל מתפלג נורמאלי עם סטיית תקן 12.
- א. מהו גודל המדגם המינימאלי שיש לקחת אם מעוניינים ששגיאת האמידה המקסימאלית תהיה 2 גרם ברמת סמך של 95%.
- ב. ביצעו מדגם שאת גודלו מצאתם בסעיף א והתקבל ממוצע של 23 גרם. בנו רווח סמך ברמת סמך של 95% למשקל הממוצע של הרכיב.
4. זמן החלמה מאנגינה מתפלג עם סטיית תקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.
- א. בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
- ב. מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
- ג. מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.
5. משתנה מקרי מתפלג נורמאלי עם סטיית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רווח סמך ברמת סמך של 98% שאורכו לא יעלה על 2?

2.ב. שונות האוכלוסיה לא ידועה

1. זמן התגובה מתפלג נורמאלי. במטרה לאמוד את תוחלת זמן התגובה נדגמו 4 תצפיות. להלן התוצאות בשניות: 3, 4.5, 4.6, 5.2, 7.4. בנו רווח סמך ברמת סמך של 95% לממוצע זמן התגובה באוכלוסיה.
2. ציוני מבחן אינטליגנציה מתפלגים נורמאלי. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 ו סטיית תקן מדגמית 13.
- א. בנו רווח סמך לממוצע הציונים באוכלוסיה ברמת ביטחון של 95%.
- ב. חזרו על סעיף א' אם סטיית התקן הנתונה היא סטיית התקן האמיתית של כלל הנבחנים.
- ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.
3. נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

תשובות סופיות לאמידה נקודתית ורווחי סמך

פרק א' - אמידה נקודתית

<u>שאלה 1</u>	<u>שאלה 2</u>
א. 177.9	א. 8.1
ב. 64.1	ב. 3.16
ג. 0.4	

פרק ב' - רווח סמך לתוחלת

שונות האוכלוסיה ידועה

<u>שאלה 1</u>	<u>שאלה 2</u>
א. $223.42 < \mu < 236.58$	א. 102
ב. $222.16 < \mu < 237.84$	ב. 3
ד. 4	ג. 0.9544
<u>שאלה 3</u>	<u>שאלה 4</u>
א. 139	א. $3.58 < \mu < 4.42$
ב. $21 < \mu < 25$	ב. יקטן פי 2
	ג. גדל
<u>שאלה 5</u>	
780	

שונות האוכלוסיה לא ידועה

<u>שאלה 1</u>	<u>שאלה 2</u>
$4.39 < \mu < 5.51$	א. $96.63 < \mu < 107.37$
	ב. $96.90 < \mu < 107.10$
<u>שאלה 3</u>	
$3.149 < \mu < 3.351$	

פרק שישי - בדיקת השערות - מבחנים פרמטרים

א. בדיקת השערות על תוחלת

שונות האוכלוסיה ידועה

1. ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהייה 15 מה מסקנתכם ברמת מובהקות של 5%?

2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.

א. מה מסקנתכם ברמת מובהקות של 2.5%?

ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?

3. מהנדס האיכות מעוניין לבדוק אם מכונה מכוילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?

שונות האוכלוסיה לא ידועה

1. משך זמן ההחלמה בלקיחת אנטיביוטיקה מסוימת הוא 120 שעות בממוצע עם סטיית תקן לא ידועה. מעוניינים לבדוק האם אנטיביוטיקה אחרת מקטינה את משך זמן ההחלמה. במדגם של 5 חולים שלקחו את האנטיביוטיקה האחרת התקבלו זמני ההחלמה הבאים: 90, 95, 100, 80, 125 שעות. מה מסקנתכם ברמת מובהקות של 5%. מהי ההנחה הדרושה לצורך הפתרון.

2. ציוני מבחן IQ מתפלגים נורמלית בארצות הברית עם ממוצע 100. במדגם על 20 נבחנים ישראלים התקבלו התוצאות הבאות.

$$\bar{x} = 102$$

$$\hat{S} = 16$$

האם בישראל ממוצע הציונים הוא כמו בארצות הברית? בדקו ברמת מובהקות של 2%
הניחו הנחות אם יש צורך.

ג. בדיקת השערות להפרש תוחלות במדגמים בלתי תלויים

שוניות האוכלוסייה ידועות

1. 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%.

2. 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.
האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית בדקו ברמת מובהקות של 10%. מה ההנחה הדרושה לצורך פתרון?

שוניות האוכלוסייה לא ידועות (בהנחה שהן שוות)

1. במדגם של 10 ישראלים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:

$$\begin{aligned}n_1 &= 10 \\ \sum x_i &= 1020 \\ \sum x_i^2 &= 105120\end{aligned}$$

במדגם של 14 אמריקאים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:

$$\begin{aligned}n_1 &= 14 \\ \sum x_i &= 1386 \\ \sum x_i^2 &= 138644\end{aligned}$$

בדקו האם קיים הבדל בממוצע הציונים בבחינת ה-IQ בין ישראל לארה"ב. רשמו את כל ההנחות הדרושות. ברמת מובהקות של 5%.

2. להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

1-100W	2-60W	הקבוצה
956	1007	$\bar{x}$
72	80	$\hat{S}$
15	13	n

בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפיתרון.

ד. בדיקת השערות על מדגמים מזווגים

1. במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

המדינה	X	Y
ארה"ב	1.5	1.4
קנדה	2.1	2
הולנד	2.2	1.9
פולין	3	3.1
מצרים	3.5	3.2
סין	3.2	3.2
יפן	4.2	4.2

בהנחה והמחירים מתפלגים נורמלית בכל חברה, בדקו ברמת מובהקות של 5% האם קיים הבדל בין החברות מבחינת המחירים במוצק.

2. מכון המכין לפסיכומטרי טוען שהוא מעלה את ממוצע הציונים ביותר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

לפני	506	470	420	640	670	390	500	590
אחרי	570	540	430	610	680	510	520	580

מה מסקנתכם ברמת מובהקות 5%? הניחו שציוני פסיכומטרי מתפלג נורמלית.

ה. הגישה המעשית – מובהקות התוצאה – **P-VALUE** – $\alpha_{\min}$

1. השכר הממוצע במשק בשנת 2006 היה 7800 ₪ עם סטיית תקן 2000 ₪. במדגם כיום של מאה עובדים התקבל שכר ממוצע 7500 ₪. מטרת המחקר היא לבדוק האם קיום חלה שחיקה בשכר כיום. עבור אילו רמות מובהקות שיבחר החוקר יוחלט לקבל את טענת המחקר?

$$2. \quad \begin{aligned} H_0 : \mu &= 70 \\ H_1 : \mu &> 70 \end{aligned}$$

לפניך השערות של מחקר :

המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות :

$$n = 100$$

$$\bar{x} = 74$$

מהי האלפא המנימלית לדחיית השערת האפס?

3. הגובה של מתגייסים לצה"ל מתפלג נורמלית. במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות :

$$\bar{x} = 176.2$$

$$\sum (x_i - \bar{x})^2 = 2832$$

מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוה מ-174 ס"מ באופן מובהק. מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6% ?

ו. טעויות ועוצמה

1. לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.
- א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.
- ב. בהמשך לסעיף א מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?
- ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העוצמה של הכלל מסעיף א?

2. להלן נתונים על תהליך של בדיקת השערות על תוחלת:

$$H_0 : \mu = 200$$

$$H_1 : \mu \neq 200$$

$$\sigma = 30$$

$$n = 225$$

- א. רשום כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.
- ב. בהמשך לסעיף א מהי העוצמה אם התוחלת שווה ל-195?
- ג. הסבר ללא חישוב איך העוצמה תשתנה אם רמת המובהקות תהיה 5%?

3. מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומוודדים את קוטרים, בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.
- א. רשום את ההשערות ואת ככל ההכרעה ברמת מובהקות של 5%.
- ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?
- ג. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.
- ד. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

ז. הקשר בין רווח סמך לבדיקת השערות

1. חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10 מ"ג. נלקח מדגם של 100 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היא 112.6 מ"ג לסמ"ק.

בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו. א.

ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 113 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבר. ב.

2. חוקר רצה לבדוק את ההשערות הבאות :

$$H_0 : \mu = 90$$

$$H_1 : \mu \neq 90$$

החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא : (87,97). אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.

3. במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלגים נורמלית בנו רווח סמך לממוצע ההפרשים שיבדוק את השערות המחקר ברמת מובהקות של 5%

להלן התוצאות :

המדינה	X	Y
ארה"ב	1.5	1.4
קנדה	2.1	2
הולנד	2.2	1.9
פולין	3	3.1
מצרים	3.5	3.2
סין	3.2	3.2
יפן	4.2	4.2

תשובות סופיות לבדיקת השערות

פרק א' – בדיקת השערות לתוחלת

שונות האוכלוסייה ידועה

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

שונות האוכלוסייה לא ידועה

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

פרק ב' – בדיקת השערות להפרש תוחלות

שונות האוכלוסייה ידועות

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

שונות האוכלוסייה לא ידועות

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

פרק ג' – בדיקת השערות במדגמים מזווגים

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

--	--

פרק ד' – מובהקות התוצאה

<u>שאלה 1</u>	<u>שאלה 2</u>
מעל 0.0668	0.0228
	<u>שאלה 3</u>
	נקבל H_0

פרק ה' – טעויות ועוצמה

<u>שאלה 1</u>	<u>שאלה 2</u>
ב. נדחה H_0	ב. 0.8051
ג. 1	

פרק ו' – הקשר בין בדיקת השערות לרווח סמך

<u>שאלה 1</u>	<u>שאלה 3</u>
(110.64, 114.56)	(-0.241, 0.041)

פרק שביעי - שאלות אמריקאיות על כל חומר הלימוד.

הנתונים הבאים מתייחסים לשאלות 1-4

פסיכולוגים צפו במשך שבוע שלם בהתנהגותם של 28 ילדים בגן חובה. לאחר מכן נאלצו לדווח על רמת הביטחון העצמי של כל ילד בסקלה של 1 עד 5. כאשר 5 נחשב לרמת בטחון עצמי גבוהה ו-1 לרמת בטחון עצמי נמוכה. להלן סיכום התוצאות:

מספר הילדים	בטחון עצמי
6	1
7	2
10	3
4	4
1	5

שאלה 1

מהו סולם המדידה של המשתנה הנחקר?

- א. שמי.
- ב. סדר.
- ג. רווח.
- ד. מנה.

שאלה 2

מהי הדרך הגרפית המתאימה ביותר כדי לתאר את הנתונים?

- א. טבלת שכיחויות.
- ב. דיאגרמת מקלות.
- ג. היסטוגרמה.
- ד. דיאגרמת עוגה.

שאלה 3

מהו השכיח של התפלגות הנתונים שנאספו?

- א. 2
- ב. 1
- ג. 3

ד. 10

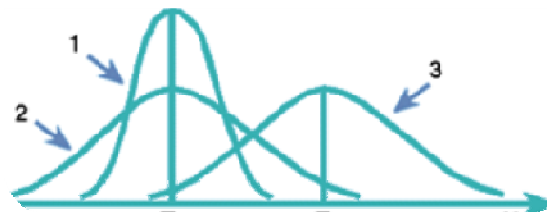
שאלה 4

התווסף עוד ילד עם רמת בטחון עצמי נמוכה לכן סטיית התקן של המשתנה הנחקר כתוצאה מההוספה:

- א. תגדל
- ב. תקטן
- ג. לא תשתנה
- ד. אין לדעת

הנתונים הבאים מתייחסים לשאלות 6-10

להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל בניהן.

**שאלה 6**

לאיזו התפלגות הממוצע הגבוה ביותר?

- א. 1
- ב. 2
- ג. 3
- ד. אין לדעת.

שאלה 7

לאיזו התפלגות השכיח הגדול ביותר?

- א. 1
- ב. 2
- ג. 3
- ד. אין לדעת

שאלה 8

במה התפלגות 1 ו 2 זהות?

- א. בעשירון העליון.
- ב. בממוצע.
- ג. בשונות.
- ד. אף אחת מהתשובות אינה נכונה.

שאלה 9

איזה מהמשפטים הבאים נכון לגבי התפלגות מספר 3?

- א. הממוצע שווה לחציון בהתפלגות.
- ב. הטווח שווה לטווח הבין רבעוני.
- ג. העשירון התחתון שווה לעשירון העליון.
- ד. סטיית התקן היא אפס.

שאלה 10

לאיזו התפלגות סטיית התקן הקטנה ביותר?

- א. 1
- ב. 2
- ג. 3
- ד. אין לדעת.

הנתונים הבאים מתייחסים לשאלות 11-15 :

מוכר החליט לתת 20% הנחה לכל המוצרים שבחנות שלו.

נסמן ב- X המחיר של מוצר לפני ההנחה בש"ח וב- Y את המחיר של המוצר אחרי ההנחה בש"ח.

המוכר חישב את המדדים הבאים לפני ההנחה :

ממוצע	80
חציון	70
שונות	300
טווח	48

כמו כן הוא חישב גם את כל הנתונים לגבי המשתנה Y .

שאלה 11

מה יהיה הממוצע של המחירים בש"ח אחרי ההנחה?

א. 16

ב. 64

ג. 80

ד. 70

שאלה 12

מה יהיה טווח המחירים בש"ח אחרי ההנחה?

א. 9.6

ב. 38.4

ג. 48

ד. 70

שאלה 13

מה תהיה השונות של המחירים אחרי ההנחה?

א. 300

ב. 60

ג. 240

ד. 192

שאלה 14

מהו מקדם ההשתנות (CV) של המחירים לפני ההנחה?

- א. 3.75
- ב. 0.267
- ג. 0.2165
- ד. 4.619

שאלה 15

אם המוכר יחשב את מקדם המתאם על X ו Y התוצאה שתתקבל תהיה?

- א. 0
- ב. 1
- ג. -1
- ד. אין לדעת.

שאלה 16

בהתפלגות אסמטרית ימנית סטיית התקן יותר גדולה מאשר בהתפלגות אסמטרית שמאלית.

- א. הטענה תמיד נכונה.
- ב. הטענה תמיד אינה נכונה בהכרח.
- ג. אין מספיק נתונים כדי לדעת.

שאלה 17

ביחס לציר המספרים רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים:

- א. בערכים הגבוהים.
- ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
- ג. בערכים הנמוכים.
- ד. לא ניתן לדעת.

שאלה 18

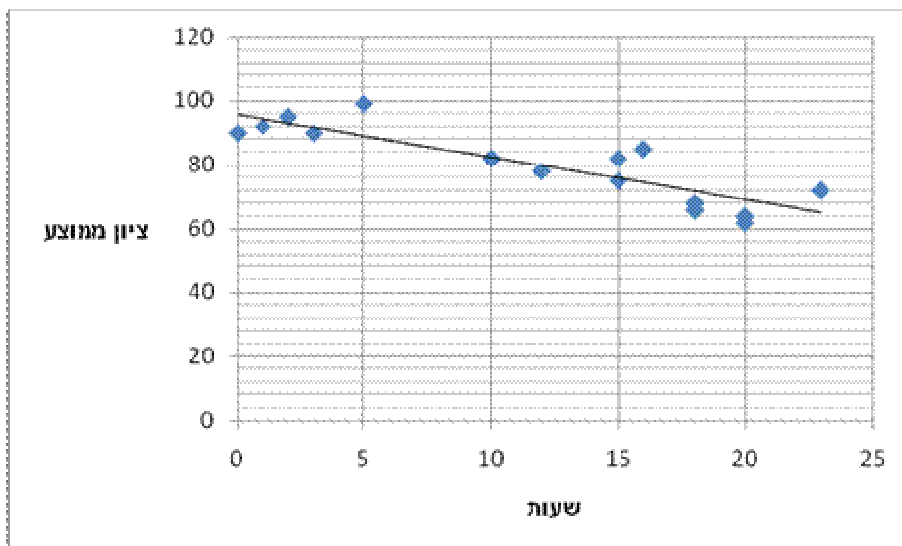
הוספת גודל קבוע לכל תצפיות סדרת נתונים.

- א. תגדיל את סטיית התקן.
- ב. תקטין את סטיית התקן.
- ג. לא תשנה את סטיית התקן.

ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 19-21

חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר בעזרת האקסל דיאגרמת פיזור. החוקר אף הוסיף לדיאגרמה את קו המגמה המתאים לנתונים.



שאלה 19

מיהו המשתנה הבלתי תלוי?

- א. ציון ממוצע.
- ב. מספר שעות לבילוי.
- ג. מספר הסטודנטים.

שאלה 20

מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

(הסתמכו על הנתונים ולא על דעתכם האישית...)

- א. ככל שמבליים יותר הציון נוטה לרדת.
- ב. אין קשר בין שעות הבילוי לציון.
- ג. ככל שמבליים פחות הציון נוטה לרדת.
- ד. ככל שהציון יורד הסטודנט מבלה פחות.

שאלה 21

איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?

- א. 0.85
- ב. 0.15
- ג. - 0.85
- ד. - 0.15

שאלה 22

סטיית התקן של משתנה מסוים X הייתה 2. הוחלט לבצע טרנספורמציה למשתנה לפי הקשר הבא: $Y = 3X - 2$. שונות Y אחרי הטרנספורמציה היא:

- א. 4
- ב. 6
- ג. 10
- ד. 12
- ה. 36

הנתונים הבאים מתייחסים לשאלות 23-25

בכיתה 30 סטודנטים אותם 30 נבחנו במבחן באנגלית ובמבחן בסטטיסטיקה. להלן פלט לגבי ציונים:

סטטיסטיקה	אנגלית	
80	90	ממוצע
100	121	שונות

שאלה 23

באיזה מקצוע להתפלגות הציונים פיזור יחסית יותר גבוה?

- א. אנגלית.
- ב. סטטיסטיקה
- ג. אותו פיזור בשני המקצועות באופן יחסי.
- ד. אין מספיק נתונים כדי לענות על השאלה.

שאלה 24

יערה קיבלה 92 באנגלית ו82 בסטטיסטיקה. באיזה מקצוע היא יותר טובה יחסית לכיתה?

- א. אנגלית.
- ב. סטטיסטיקה
- ג. אותו דבר יחסית.

ד. אין מספיק נתונים כדי לענות על השאלה.

שאלה 25

עודד שקיבל 80 בסטיסטיקה העתיק בבחינה. הוחלט לחשב מחדש את השונות של הציונים בסטיסטיקה בלעדיו. השונות החדשה :

- א. תקטן
- ב. תגדל
- ג. לא תשתנה
- ד. אין לדעת

שאלה 26

חושב הטווח הבין רבעוני עבור התפלגות מסוימת והתקבלה התוצאה אפס. לכן :

- א. לפחות 50% מהתצפיות זהות.
- ב. סטיית התקן היא אפס.
- ג. ההתפלגות היא סימטרית.
- ד. מצב זה כלל לא יתכן.

שאלה 27

נתונה התפלגות של משתנה כלשהו.

- א. הטווח של 20% התצפיות הגבוהות ביותר שווה לטווח של 20% התצפיות הנמוכות ביותר.
- ב. הטווח של 50% התצפיות המרכזיות הינו הטווח הבין רבעוני.
- ג. הרבעון העליון שווה לרבעון התחתון.
- ד. הטווח הבין רבעוני הוא מחצית מהטווח.

הנתונים הבאים מתייחסים לכל השאלות 28-29

חוקר רצה לחקור את הקשר הקווי שבין הציון במבחן הרשות בסטטיסטיקה ומימון לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות : הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.

שאלה 28

על פי משוואת הרגרסיה של שעת הכנה נוספת משפרת את ציון המבחן ב :

- א. 1.5 נקודות.
- ב. 0.53 נקודות.
- ג. 0.66 נקודות.
- ד. 1.20 נקודות.
- ה. 0.96 נקודות.

שאלה 29

על פי משוואת הרגרסיה תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון :

- א. 29.
- ב. 0.
- ג. 33.
- ד. 24.
- ה. 26.

שאלה 30

אם מקדם המתאם בין שני משתנים הוא שלילי אזי :

- א. הערכים של המשתנים הם שליליים.
- ב. ככל שמשתנה אחד עולה השני עולה.
- ג. ככל שמשתנה אחד יורד השני יורד.
- ד. קיימת טרנספורמציה לינארית שלילית בין שני המשתנים.
- ה. אף טענה אינה נכונה.

שאלה 33

ככל שההתפלגות הנורמאלית חדה וצרה יותר במרכזה אזי :

- א. השונות שלה יותר גבוהה
- ב. הממוצע שלה יותר גבוה
- ג. היא מייצגת אנשים גבוהים יותר
- ד. השונות שלה נמוכה יותר
- ה. החציון שלה גבוה יותר

שאלה 34

נתונה סדרה של N מדידות שלא כולן זהות.

נניח ששתי מדידות נוספות צורפו לסדרה ושתייהן זהות לממוצע הסדרה. האם וכיצד תשנה הוספת שני הערכים החדשים את שונות הסדרה?

- א. שונות הסדרה תקטן
- ב. שונות הסדרה תגדל
- ג. לא ניתן לדעת, זה תלוי במספר התצפיות
- ד. לא ניתן לדעת, זה תלוי בערכו של הממוצע

שאלה 35

הוותק הממוצע של עובדי מפעל מסוים הוא 12 שנים וסטיית התקן של הוותק 8 שנים. בעוד 3 שנים – אם כל העובדים ימשיכו לעבוד במפעל ולא יתווספו עובדים חדשים – נקבל כי :

- א. הממוצע 15 שנים וסטיית התקן 8 שנים
- ב. הממוצע 12 שנים וסטיית התקן 11 שנים
- ג. הממוצע 15 שנים וסטיית התקן 11 שנים
- ד. הממוצע 12 שנים וסטיית התקן 8 שנים

שאלה 36

שני סטודנטים עזבו את החוג לכלכלה. הציון של כל אחד מהם היה שווה לציון הממוצע. כיצד תשפיע עזיבתם על הממוצע ושונות ציוני התלמידים הנותרים ? אם הממוצע לפני העזיבה היה 80 והשונות 100.

- א. הממוצע לא ישתנה והשונות תגדל.
- ב. הממוצע לא ישתנה והשונות תקטן.
- ג. הממוצע לא ישתנה והשונות לא תשתנה.
- ד. הממוצע יקטן והשונות תגדל.
- ה. הממוצע יגדל והשונות תקטן.

שאלה 37

החציון של סדרת נתונים מסוימת הוא 90. הוסיפו שתי תצפיות נוספות: 100 ו-20, לכן החציון:

- א. יקטן.
- ב. יגדל.
- ג. לא ישתנה.
- ד. לא ניתן לדעת.

שאלה 38

סטיית התקן של המשכורות בחברה הנה 3000 ₪ אם נוסף לכל עובדי החברה 200 ₪ לשכר אז:

- א. סטיית התקן תגדל אך אין לדעת בכמה.
- ב. סטיית התקן תגדל בהכרח ב – 200 ₪.
- ג. סטיית התקן לא תשתנה.
- ד. סטיית התקן תקטן.
- ה. לא ניתן לדעת.

שאלה 40

ממוצע של סידרת נתונים הנה 50 וסטיית התקן 10. אם נוסף עוד שתי תצפיות שערכן 50 סטיית התקן:

- א. תקטן.
- ב. תגדל.
- ג. לא תשתנה.
- ד. אין לדעת.

שאלה 41

בהתפלגות אסימטרית עם זנב ימני ציון התקן של הרבעון התחתון:

- א. בהכרח שלילי.
- ב. בהכרח חיובי.
- ג. אפס.
- ד. לא ניתן לדעת.

שאלה 42

אם השונות של המשתנה שווה אפס. מה ניתן לומר על המשתנה?

א. עולה. ב. יורד. ג. קבוע. ד. נורמלי. ה. לא ניתן לדעת

שאלה 43

נתון משתנה מקרי W עם שונות 10.

מה תהיה השונות אם נכפיל את ערכי המשתנה W פי 2?

א. 20. ב. 10. ג. 400. ד. 40. ה. 0

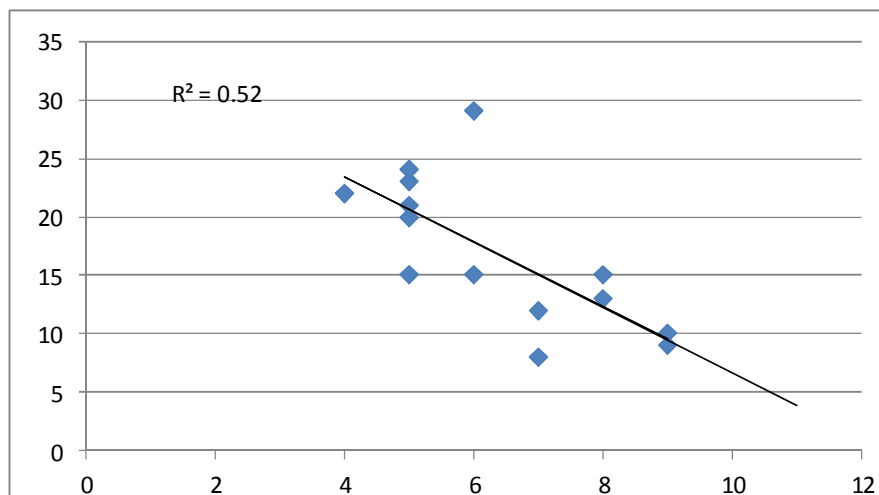
שאלה 44

נמצא שקיים מקדם מתאם חיובי בין הציון בעברית לציון בחשבון בבחינה לכן :

- א. הדבר מעיד שהציונים בכתה היו חיוביים.
- ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
- ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
- ד. אף אחת מהתשובות לא נכונה.

הנתונים הבאים מתייחסים לשאלות 46-48

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים X (משתנה בלתי תלוי-בציר האופקי) ו- Y (משתנה תלוי), כמו כן הועבר קו הרגרסיה וחושב ריבוע מקדם המתאם.



שאלה 46

לאור הנתונים המופיעים בדיאגרמה איזה מבין הערכים הבאים מתאים להיות התוצאה של מקדם המתאם שתופעל על הנתונים?

- א. 0.52
- ב. -0.52
- ג. -0.72
- ד. 0.72

שאלה 47

מה תהיה התוצאה הכי מתאימה לפרמטר b ברגרסיה?

- א. 0.52
- ב. 2.79
- ג. -2.79
- ד. -0.52

שאלה 48

מהו טווח התפלגות התצפיות של המשתנה הבלתי תלוי X ?

- א. 5
- ב. 12
- ג. 6.5
- ד. 7

הנתונים הבאים מתייחסים לשאלות 49-51

במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים:

מספר פועלים	תפוקה	
15	48	ממוצע
2	10	סטיית תקן

שאלה 49

איזו טענה מהטענות הבאות נכונה?

- א. המספר המקסימלי של העובדים במפעל הוא 17 עובדים.
- ב. התפוקה הכוללת במשך ה- 40 ימים הללו הייתה 192,000 מצברים.
- ג. הטווח של התפלגות תפוקת המצברים הוא 20 מאות.

ד. אף אחת מהטענות לא נכונה.

שאלה 50

לפי קריטריון CV (מקדם ההשתנות)

- הפיזור באופן יחסי שווה בין התפוקה היומית לכמות הפועלים העובדים ביום.
- הפיזור יחסית יותר גדול עבור התפוקה היומית מאשר עבור מספר הפועלים ביום.
- הפיזור יחסית יותר גדול עבור מספר הפועלים ביום מאשר עבור התפוקה היומית.
- אין מספיק נתונים כדי לחשב את CV.

שאלה 51

באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.

מה יותר חריג באותו היום יחסית לשאר הימים שנבדקו נתוני התפוקה או כמות הפועלים?

- חריגים באותה מידה.
- כמות הפועלים.
- התפוקה.
- חסרים נתונים כדי לדעת זאת.

שאלה 52

התפלגות הציונים במבחן מסוים היא סימטרית לכן:

- סטיית התקן של הציונים היא אפס.
- הציון החציוני שווה לציון הממוצע.
- העשירון העליון שווה לעשירון התחתון של הציונים.
- כל הטענות בשאר הסעיפים לא נכונות.

שאלה 53

מקדם המתאם בין ההכנסה לבין ההוצאה של 10 משפחות חושב והתקבל 0.7. אם חל גידול של 5% בהכנסת האוכלוסייה כולה וגידול של 7% בהוצאה שלה, אזי מקדם המתאם בין ההכנסה החדשה של 10 המשפחות הני"ל:

- לא ישתנה וישאר 0.7.
- יהפוך להיות 0.7 -.
- אין מספיק נתונים כדי לדעת מה יהיה מקדם המתאם.
- אפשר לדעת רק מה יהיה מקדם המתאם באוכלוסייה כולה.
- בין 0.7 ל- 0.7-.

שאלה 54

איזה מהמשפטים הבאים אינו נכון?

- א. אם מוסיפים קבוע לתצפיות הדבר לא משפיע על פיזור הנתונים.
- ב. בהתפלגות סימטרית הממוצע שווה לשכית.
- ג. אם כל התצפיות זהות סטיית התקן בהכרח אפס.
- ד. הכפלה בקבוע משנה את סטיית התקן.

שאלה 55

איזה מהמשפטים הבאים נכון?

- א. הטווח הבין רבעוני הוא אפס רק אם כל הצפיות זהות.
- ב. הרבעון העליון שווה לרבעון התחתון בהתפלגות סימטרית.
- ג. בהתפלגות סימטרית החציון שווה לממוצע.
- ד. 90% מהתצפיות נמצאות מעל האחוזון התשעים.

שאלה 58

הכפלה בגודל קבוע לכל תצפיות סדרת נתונים.

- א. תגדיל את סטיית התקן.
- ב. תקטין את סטיית התקן.
- ג. לא תשנה את סטיית התקן.
- ד. לא ניתן לדעת.

שאלה 59

בעיר "חולית", בקיץ, כמות הגשם היורד בחודש מתפלג נורמלית עם תוחלת 10 מ"מ וסטיית התקן 2, ובחורף עם תוחלת 10 מ"מ וסטיית התקן 3. איפה יש יותר סיכוי שירד יותר מ 12 מ"מ גשם?

- א. בקיץ
- ב. בחורף
- ג. סיכוי שווה.
- ד. לא ניתן לדעת.

שאלה 60

בהתפלגות שבה המאון ה-40 שווה לממוצע. ציון התקן של הממוצע יהיה :

- א. חיובי.
- ב. שלילי.
- ג. אפס.
- ד. לא ניתן לדעת.

דף פתרונות:

1	ב	11	ב	21	ג		41	א	51	ב
2	ב	12	ב	22	ה		42	ג	52	ב
3	ג	13	ד	23	ב	33	ד	43	ד	א
4	א	14	ג	24	ב	34	א	44	ב	ב
		15	ב	25	ב	35	א		55	ג
6	ג	16	ג	26	א	36	א	46	ג	
7	ג	17	ג	27	ב	37	ג	47	ג	
8	ב	18	ג	28	ד	38	ג	48	א	ד
9	א	19	ב	29	א			49	ב	ב
10	א	20	א	30	ה	40	א	50	ב	ג

פרק שמיני - ניתוח שונות חד כיוונית

רקע תאורטי:

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 קטגוריאל עם k רמות (levels).

כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה:

נסמן ב- n_i את גודל המדגם בקבוצה i .

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

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

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

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

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

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

$$SST = SSB + SSW$$

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