

סטטיסטיקה



תוכן העניינים

1.	סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה	1
6.	סטטיסטיקה תיאורית- הצגה של נתונים	6
14.	סטטיסטיקה תיאורית- סכימה	14
(ללא ספר)	סטטיסטיקה תיאורית- מדדי מרכז	
(ללא ספר)	סטטיסטיקה תיאורית- מדדי פיזור	
18.	סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני	18
22.	סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן	22
24.	סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה	24
26.	סטטיסטיקה תיאורית- תרשים קופסא	26
28.	סטטיסטיקה תיאורית שאלות אמריקאיות	28
34.	התפלגויות רציפות מיוחדות - התפלגות נורמלית	34
41.	התפלגויות בדידות מיוחדות -התפלגות בינומית	41
44.	הסקה סטטיסטית - הקדמה	44
47.	התפלגות הדגימה ומשפט הגבול המרכזי	47
64.	מושגי יסוד באמידה	64
69.	רווח סמך לתוחלת (ממוצע)	69
74.	מבוא לבדיקת השערות על פרמטרים	74
80.	בדיקת השערות על תוחלת (ממוצע)	80
104.	רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים	104
110.	רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים	110
112.	בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים	112
123.	בדיקת השערות לתוחלת ההפרש במדגמים מזווגים	123
127.	הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות	127

תוכן העניינים

24.	רווח סמך לפרופורציה	130
25.	רווח סמך להפרש פרופורציות	136
26.	בדיקת השערות על פרופורציה	138
27.	בדיקת השערות על הפרש פרופורציות	149
28.	נמבחן הבינום	153
29.	שאלות מסכמות בבדיקת השערות	163
30.	שאלות מסכמות על רווחי סמך	179
31.	יסודות ההסתברות	182
32.	פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים	186
33.	הסתברות מותנית-במרחב מדגם אחיד	195
34.	הסתברות מותנית - מרחב לא אחיד	198
35.	דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה	202
36.	תלות ואי תלות בין מאורעות	207
37.	שאלות מסכמות בהסתברות	211
38.	מקדם המתאם (מדד קשר) הלינארי ומובהקותו	216
39.	רגרסיה	239
40.	מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת	242

סטטיסטיקה

פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

1. סולמות מדידה.....1

סטטיסטיקה תיאורית – סיווג משתנים וסולמות מדידה:

רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתח אותם.

בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת. באותה קבוצה אנו אוספים נתונים על הישויות באותה קבוצה.

משתנה – תכונה שיכולה לקבל מספר ערכים : דעה פוליטית, מקום מגורים, גובה של אדם וכדומה.

כל ישות בקבוצה שאנו צופים בה ואוספים לגביה נתונים נקראת תצפית.

הנתונים שנאספים בדרך כלל מרוכזים במסד נתונים. במסד הנתונים כל שורה היא תצפית וכל עמודה מייצגת משתנה.

דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכביש החוף. אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה :

מין	גיל	מצב הפצוע	דופק
גבר	26.6	אנוש	40
גבר	24.5	קשה	38
אישה	32.1	קשה	50
גבר	34.9	בינוני	65
אישה	23.1	קל	89

ענו על השאלות הבאות :

הגדירו את הקבוצה שבדוגמה.

כמה תצפיות בקבוצה?

כמה משתנים בקבוצה?

כמה ערכים יש למשתנה "מין"?

את המשתנים במחקר אנו מסווגים ל-"סולמות מדידה" הדבר חשוב לדרך שבה ננתח את הנתונים בהמשך.

מיון משתנים לפי סולמות המדידה:

1. סולם שמי (nominal) – משתנה שלערכיו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות לערכים שלהם לדוגמה: צבע מועדף.
משתנה דיכוטומי (הינו מסולם שמי) אותם משתנים שיש להם רק שני ערכים אפשריים. למשל: אזרחות ישראלית: יש או אין.
2. סולם סדר (ordinal) – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר מי יותר או מי פחות אבל אין משמעות לגודל.
משתנה מסולם סדר יכול לקבל ערכים מילוליים או מספריים.
למשל, דרגה בצבא.
3. סולם כמותי (scale) – משתנה שחייב להיות מספרי, לערכים שלו בנוסף לשם ולסדר בניהם יש משמעות לערך המספרי. משתנה כמותי הוא משתנה שניתן בדרך כלל לספור או למדוד על ידי מכשיר מדידה. למשל, מספר המחשבים בדירה, שטח הדירה במ"ר.

את המשתנה הכמותי אנו מסווגים לשני סוגים:

משתנה בדיד:

משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים כמו: מספר המחשבים בדירה.

משתנה רציף:

משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף וללא קפיצות של ערכים. למשל, שטח הדירה במ"ר.

דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכביש החוף.
אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה:

מין	גיל	מצב הפצוע	דופק
גבר	26.6	אנוש	40
גבר	24.5	קשה	38
אישה	32.1	קשה	50
גבר	34.9	בינוני	65
אישה	23.1	קל	89

סווגו כל משתנה במסד הנתונים: שמי, סדר, כמותי רציף, כמותי בדיד.

שאלות:

(1) לפניכם טבלה המסכמת נתונים לגבי סקר שנעשה היום:

שם משפחה	מצב משפחתי	מידת דתיות	מספר רכבים
כהן	רווק	חילוני	0
חדד	נשוי	חילוני	1
לביא	גרוש	מסורתי	1
פיינגולד	אלמן	חילוני	2
אבו שוקרא	נשוי	דתי	1
בן חיים	נשוי	מסורתי	0
רוטשילד	רווק	חילוני	0

א. כמה תצפיות בדוגמה זו?

ב. כמה משתנים בדוגמה זו?

ג. כמה ערכים ישנם ל-"מידת דתיות"?

ד. מהם הערכים האפשריים למשתנה "מצב משפחתי"?

(2) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן: 50 אנשים נכנסו 0 פעמים לסניף, 20 אנשים נכנסו פעם אחת לסניף, 5 אנשים נכנסו פעמיים לסניף, 5 אנשים נכנסו יותר מפעמיים.

א. הגדירו את הקבוצה בדוגמה זו.

ב. כמה תצפיות בדוגמה זו?

ג. הגדירו את המשתנה בדוגמה זו. מהו סולם המדידה שלו?

(3) במחקר רפואי התעניינו לדעת כיצד מינון תרופת "קופקס" משפיע על מספר שעות השינה של אדם. במחקר השתתף אדם אחד בשם דני שבכל יום ניתן לו מינון שונה של התרופה. הטבלה שלהלן מתארת בכל יום את מינון התרופה במ"ג שקיבל האדם וכמו כן את מספר שעות השינה שלו באותו הלילה:

מספר שעות שינה	מינון התרופה	מספר היום
6	12	1
7	14	2
7.5	16	3
6.5	18	4
8	20	5

א. כמה תצפיות נאספו במחקר?

ב. סווגו את סולם המדידה של "מינון התרופה" ?

- (4) לפניכם רשימה של משתנים.
- צינו באיזה סולם מדידה מדובר (שמי, סדר, כמותי בדיד, כמותי רציף):
- א. גובה אדם בס"מ.
 - ב. מספר ילדים למשפחה.
 - ג. מידת חרדה לפני מבחן.
 - ד. שביעות רצון משירות לקוחות בסקלה מ-1 עד 7 (1 כלל לא מרוצה עד 7 מרוצה מאד).
 - ה. השכלה.
 - ו. מספר אוטובוס.
 - ז. מקום מגורים.
 - ח. מין (1=גבר ו-2=אישה).
 - ט. מידת נעליים.
- (5) לפניכם רשימה של משתנים כמותיים.
- ציין ליד כל משתנה אם הוא רציף או בדיד:
- א. שכר עובד ב-ש.
 - ב. ציון בחינת בגרות.
 - ג. תוצאה בהטלת קובייה.
 - ד. מהירות ריצה במטר לשנייה בתחרות 100 מטר.
 - ה. שיעור התמיכה בממשלה בעיר.
- (6) גברת לוי החליטה לדגום 25 ימים של נסיעה לעבודה, כאשר בדרך לעבודתה יש 3 צמתים מרומזרים.
- ב-9 ימים הגיעה גברת לוי לעבודה מבלי לעצור באף צומת.
 - ב-9 ימים נוספים היא הצליחה לעבור בשני רמזורים ירוקים.
 - ב-5 ימים נוספים היא הצליח לעבור רק בירוק אחד.
- בשאר הימים, היא לא עברה באף רמזור ירוק.
- מעוניינים לחקור את מספר הרמזורים האדומים בהם עצרה גברת לוי.
- א. מהו המשתנה הנחקר בדוגמה זו?
 - ב. מהם הערכים האפשריים של משתנה זה?
 - ג. כמה ערכים אפשריים יש למשתנה?
 - ד. מהו סולם המדידה של המשתנה?

תשובות סופיות:

- (1) א. $n = 7$.
ג. 3.
- (2) א. לקוחות סניף 543 של בנק "רווה".
ג. $X =$ מספר הפעמים בשבוע שלקוח נכנס לסניף. כמותי בדיד.
ב. $n = 80$.
- (3) א. $n = 5$.
ב. כמותי רציף.
- (4) א. כמותי רציף.
ג. אין מספיק נתונים.
ה. אין מספיק נתונים.
ז. שמי.
ט. סדר.
- (5) א. רציף.
ג. בדיד.
ה. רציף.
- (6) א. מספר הרמזורים בהם עוצרת גברת לוי ביום בדרך לעבודה.
ב. 0, 1, 2, 3.
ד. כמותי בדיד.
ג. 4.
- ב. 4.
ד. רווק, נשוי, גרוש, אלמן.

סטטיסטיקה

פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 6

סטטיסטיקה תיאורית – הצגה של נתונים:

רקע:

דרכים להצגת נתונים שנאספו:

רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה. רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות. ההצגה הזו רלבנטית לכל סוגי המשתנים. למשל, להלן מספר החדרים בבניין בן 5 דירות: 3, 4, 3, 5, 4.

טבלת שכיחויות בדידה:

שם המשתנה- X	שכיחות – $f(x)$	שכיחות יחסית באחוזים
X_1	f_1	$\frac{f_1}{N} \cdot 100$
X_2	f_2	$\frac{f_2}{N} \cdot 100$
X_3	f_3	$\frac{f_3}{N} \cdot 100$
$\vdots$	$\vdots$	$\vdots$
X_k	f_k	$\frac{f_k}{N} \cdot 100$
סה"כ	$N = \sum_{i=1}^k f_i$	100%

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות. לא יעיל למשתנה כמותי רציף.

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

הציון - X	מספר התלמידים – השכיחות - f	F_i	$\frac{f_i}{n}$
5	2	2	$0.08=2/25$
6	4	6	$0.16=4/25$
7	8	14	$0.32=8/25$
8	5	19	$0.2=5/25$
9	4	23	$0.16=4/25$
10	2	25	$0.08=2/25$

שכיחות מצטברת – צבירה של השכיחויות.

השכיחויות F_i – השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:

 $\frac{f_i}{n}$ – איזה חלק מהתצפיות בקבוצה שוות לערך.**טבלת שכיחויות במחלקות:**

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד
וטבלת שכיחויות תהיה ארוכה מידי.

דוגמה:

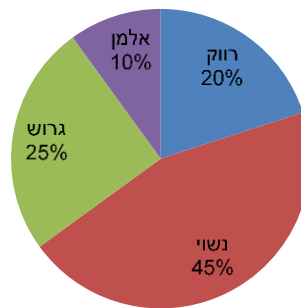
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות.
להלן ההתפלגות שהתקבלה:

מספר הילדים	זמן בדקות
20	0.5-3.5
18	3.5-9.5
14	9.5-19.5
8	19.5-29.5

דיאגרמת עוגה:

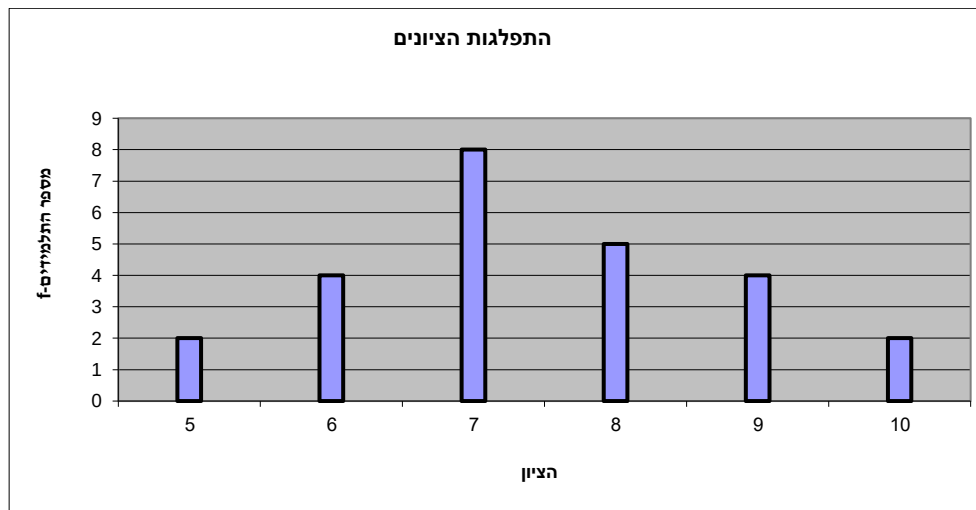
זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

התפלגות המצב המשפחתי



דיאגרמת מקלות:

הציר האופקי הוא הציר של המשתנה והציר האנכי של השכיחות, כך שהגובה של המקל מעיד על השכיחות. רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף, וכן בסולמות מדידה עבור משתנה מסולם סדר.

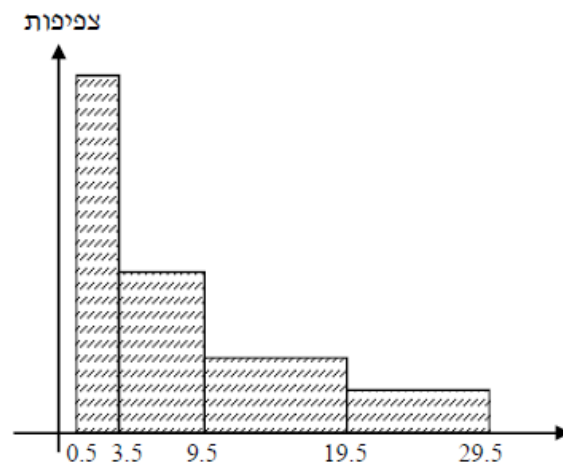


היסטוגרמה:

ההיסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות, והיא רלוונטית למשתנה כמותי רציף.

בהיסטוגרמה הציר האופקי הוא הציר של המשתנה והציר האנכי הוא הציר של הצפיפות. הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה, והיא נותנת את מספר התצפיות הממוצע בכל מחלקה ליחידה. אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך בצפיפות.

צפיפות	מצטברת	שכיחות	אמצע	רוחב	X
6.6667	20	20	2	3	0.5 - 3.5
3	38	18	6.5	6	3.5 - 9.5
1.4	52	14	14.5	10	9.5 - 19.5
0.8	60	8	24.5	10	19.5 - 29.5



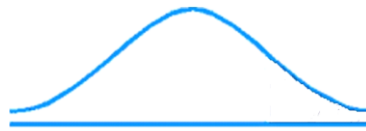
פוליגון – מצולעון:

אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של התפלגות המשתנה.

צורות התפלגות נפוצות:

התפלגות סימטרית פעמונית

רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.

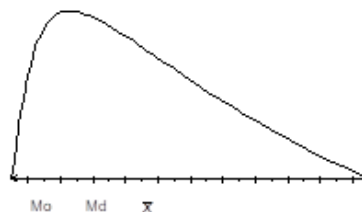


ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

התפלגות אסימטרית ימנית (חיובית)

רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. לדוגמה, שכר במשק.

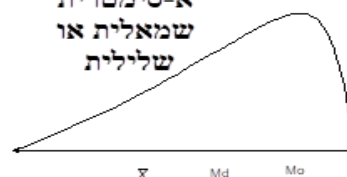
התפלגות א-סימטרית
ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית)

רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. לדוגמה, אורך חיים.

התפלגות
א-סימטרית
שמאלית או
שלילית



שאלות:

(1) בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות : 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו-25 לא צפו בטלוויזיה בזמן הסקר.

א. רשמו את טבלת השכיחות ואת השכיחות היחסית.

ב. תארו את הנתונים באופן גרפי.

(2) להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

מספר התלמידים	המקצוע
44	מתמטיקה
20	תנ"ך
12	אנגלית
26	היסטוריה

א. מהו המשתנה הנחקר?

ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

(3) להלן התפלגות ההשכלה במקום עבודה מסוים:

מספר העובדים	השכלה
60	נמוכה
120	תיכונית
20	אקדמאית

א. מהו המשתנה הנחקר?

מאיזה סולם הוא?

ב. תארו את הנתונים באופן גרפי.

(4) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא :
6, 5, 8, 7, 6, 7, 8, 6, 7, 5, 8, 4, 6, 10, 9, 8, 6, 7.

א. מהו המשתנה? האם הוא בדיד או רציף?

ב. תארו את הרשימה בטבלת שכיחויות.

ג. הוסיפו שכיחויות יחסיות לטבלה.

ד. תארו את הנתונים באופן גרפי.

תשובות סופיות:

(1) א. להלן טבלה: ב. עיין גרף מלא בסרטון הוידאו.

x	$f(x)$	$\frac{f(x)}{n}$	%
ערוץ 1	25	$\frac{25}{200}$	12.5%
ערוץ 10	25	$\frac{25}{200}$	12.5%
ערוץ 2	75	$\frac{75}{200}$	37.5%
כבלים	50	$\frac{50}{200}$	25%
לא צפו	25	$\frac{25}{200}$	12.5%
סה"כ	200	1	100%

- (2) א. מקצוע מועדף. ב. 19.6%.
- (3) א. משתנה נחקר: השכלה, סוג: סדר. ב. עיין גרף מלא בסרטון הוידאו.

4) א. המשתנה : ציון, משתנה בדיד.
ד. עיין גרף מלא בסרטון הוידאו.

ב+ג. להלן טבלה :

x	$f(x)$	$\frac{f(x)}{n}$	%
4	1	$\frac{1}{20}$	5%
5	2	$\frac{2}{20}$	10%
6	6	$\frac{6}{20}$	30%
7	4	$\frac{4}{20}$	20%
8	4	$\frac{4}{20}$	20%
9	2	$\frac{2}{20}$	10%
10	1	$\frac{1}{20}$	5%
סה"כ	20	20	100%

סטטיסטיקה

פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי 14

סטטיסטיקה תיאורית – סכימה:

רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת לסכום של תצפיות: $\sum_{i=1}^n X_i$.

נסביר את צורת הרישום על ידי הדוגמה הבאה:

i	X_i
1	5
2	0
3	1
4	3
5	2

(הסבר מלא מופיע בסרטונים באתר).

שאלות:

- 1) בבניין 5 דירות. לכל דירה רשמו את מספר החדרים שיש בדירה (X), ומספר הנפשות החיות בדירה (Y). חשבו:

מספר דירה	X	Y
1	2	1
2	3	1
3	2	2
4	4	3
5	3	2

א. $\sum_{i=1}^3 X_i$

ב. $\sum_{i=1}^5 Y_i$

ג. $\sum_{i=1}^4 X_i$

ד. $\left(\sum_{i=1}^4 X_i \right)^2$

ה. $\sum X_i$

ו. $\sum X_i Y_i$

ז. $\sum (X_i) \sum (Y_i)$

- (2) נתון לוח ערכי המשתנים X_i ו- Y_i , כאשר: $i = 1, 2, \dots, 6$, ונתונים הקבועים:
 $a = 2$, $b = 5$. חשבו את הנוסחאות הבאות:

i	1	2	3	4	5	6
X_i	3	2	4	-2	1	4
Y_i	2	0	0	1	-5	2

א. $\sum_{i=1}^4 y_i$

ב. $\sum_{i=1}^6 a$

ג. $\sum_{i=1}^6 x_i y_i$

ד. $\sum_{i=1}^6 (x_i + y_i)$

ה. $\sum_{i=1}^6 x_i + a$

- (3) קבעו לכל זהות האם היא נכונה:

א. $\sum_{i=1}^n bX_i = b \cdot \sum_{i=1}^n X_i$

ב. $\sum_{i=1}^n a = a \cdot n$

ג. $\left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n X_i^2$

- (4) נתון: $\sum_{i=1}^{10} X_i = 80$, $\sum_{i=1}^{10} X_i^2 = 1640$

חשבו: $\sum_{i=1}^{10} (X_i - 4)^2$

תשובות סופיות:

- | | | | | |
|---|-----------|-----------|--------------|---------|
| 1 | א. 7. | ב. 9. | ג. 11. | ד. 121. |
| | ה. 14. | ו. 27. | ז. 126. | |
| 2 | א. 3. | ב. 12. | ג. 7. | |
| | ד. 12. | ה. 14. | | |
| 3 | א. נכונה. | ב. נכונה. | ג. לא נכונה. | |
| 4 | 1160. | | | |

סטטיסטיקה

פרק 4 - סטטיסטיקה תיאורית- מדדי מרכז

תוכן העניינים

1. כללי (ללא ספר)

סטטיסטיקה

פרק 5 - סטטיסטיקה תיאורית- מדדי פיזור

תוכן העניינים

1. כללי (ללא ספר)

סטטיסטיקה

פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני

תוכן העניינים

1. טווח בינרבעוני.....18

סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רבעוני:

רקע:

הטווח הבין-רבעוני (יש הקוראים לו התחום הבין-רבעוני) נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות. הרעיון ליצור מדד פיזורי שלא רגיש לתצפיות החריגות ביותר. כדי לחשב את הטווח הבין-רבעוני יש למצוא את הרבעון התחתון והעליון של התפלגות התצפיות.

רבעון תחתון – ערך שמחלק את ההתפלגות לשניים.
25% מהמקרים נמוכים ממנו או שווים לו ו 75% מהמקרים גבוהים או שווים לו.
סימון: Q_1 .

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.
75% מהמקרים נמוכים ממנו או שווים לו ו 25% מהמקרים גבוהים או שווים לו.
סימון: Q_3 .

הטווח הבין רבעוני הוא הפער בין שני הרבעונים: $IQR = Q_3 - Q_1$.

שלבים במציאת טווח בין-רבעוני בטבלת שכיחויות:

שלב א: נמצא את הרבעון התחתון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 25%.

שלב ב: נמצא את הרבעון העליון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 75%.

שלב ג: נמצא את הטווח הבין-רבעוני: נחסר את הרבעונים: $IQR = Q_3 - Q_1$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו. מהו הטווח הבין-רבעוני של מספר תוכניות החיסכון בסניף?

# תוכניות החיסכון	$f(x)$	שכיחות מצטברת	שכיחות יחסית מצטברת
0	100		
1	75		
2	25		
3	25		
4	25		

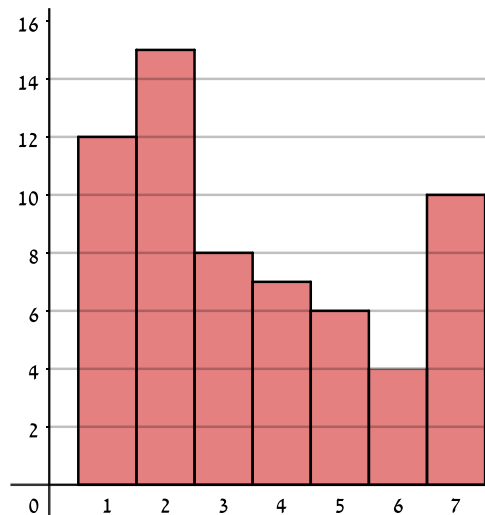
שאלות:

1) להלן התפלגות מספר המכוניות למשפחה ב"הגורן":

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

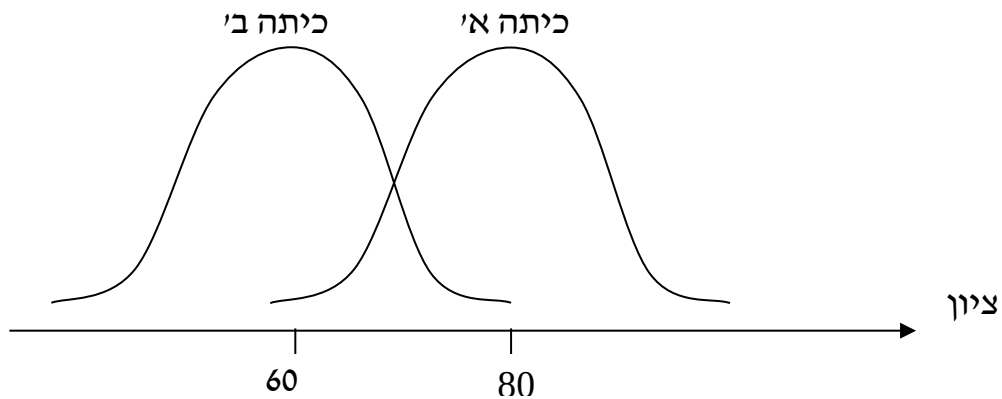
מהו הטווח הבין-רבעוני של מספר המכוניות למשפחה ב"הגורן"?

2) בסקר שנעשה בדקו את מספר ימי המחלה השנתיים של מורים בארץ.



- מה מייצגים הערכים בציר האופקי?
- מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים?
- אם נוסף 25 מורים אשר הצהירו שמספר ימי המחלה השנתיים שלהם הוא 4 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.
- אם מסתבר שחלק מהמורים בסקר הצהירו שהם חלו 7 ימים בשנה אבל בפועל הם חלו 8 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

3) לפניך שתי עקומות המתארות את התפלגות הציונים בכל כיתה. באיזו כיתה הטווח הבין-רבעוני גדול יותר?



- א. כיתה א.
- ב. כיתה ב'.
- ג. לשתיהן אותו טווח בין-רבעוני.
- ד. לא ניתן לדעת, אין מספיק נתונים.

4) הוספת גודל קבוע לכל תצפיות סדרת נתונים :

- א. תגדיל את הטווח הבין-רבעוני.
- ב. תקטין את הטווח הבין-רבעוני.
- ג. לא תשנה הטווח הבין-רבעוני.
- ד. לא ניתן לדעת מה יקרה לטווח הבין-רבעוני.

5) 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.

- (6) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן:
- 50 אנשים נכנסו 0 פעמים לסניף.
 - 20 אנשים נכנסו פעם אחת לסניף.
 - 5 אנשים נכנסו פעמיים לסניף.
 - 5 אנשים נכנסו יותר מפעמיים.
- מהו הטווח הבין-רבעוני?
- א. 60.
 - ב. 2.
 - ג. 50.
 - ד. 1.

- (7) התפלגות הציונים במבחן ווקסלר היא סימטרית לכן:
- א. טווח הציונים הוא אפס.
 - ב. הטווח הבין רבעוני של הציונים אפס.
 - ג. סעיפים א ו-ב הם נכונים.
 - ד. אף סעיף אינו נכון.

תשובות סופיות:

- (1) 2.
- (2) א. מספר ימי המחלה השנתיים. ב. 3. ג. יקטן. ד. לא ישתנה.
- (3) ג'.
- (4) ג'.
- (5) א'.
- (6) ד'.
- (7) ד'.

סטטיסטיקה

פרק 7 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

1. כללי 22

סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

רקע:

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

ציון תקן:

הנוסחה לציון תקן של תצפית היא: $Z = \frac{X - \bar{X}}{S}$.

ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע. כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע:

- ציון תקן חיובי אומר שהתצפית מעל הממוצע.
- ציון תקן שלילי אומר שהתצפית מתחת לממוצע.
- ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

דוגמה (פתרון בהקלטה):

במקום עבודה מסוים, ממוצע המשכורות הוא 8 אלף ₪, עם סטית תקן של אלפיים ₪. באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים, עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר, באופן יחסי, משכיל או משתכר?

שאלות:

- (1) תלמידי כיתה ח' ניגשו 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) במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. להלן טבלה המסכמת את המידע שנאסף על שני המשתנים:

ממוצע	תפוקה	מספר פועלים
ממוצע	48	15
סטיית תקן	10	2

- 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?
א. התפוקה.
ב. כמות הפועלים.
ג. חריגים באותה מידה.
ד. חסרים נתונים כדי לדעת זאת.

- (3) הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן של 10 סנטימטר. המשקל הממוצע הוא 66 ק"ג עם סטיית תקן של 8 ק"ג. ערך התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.
א. במה ערך חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?
ב. כמה ערך אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

תשובות סופיות:

- (1) א. לשון. ב. 72.
(2) ב'.
(3) א. משקל. ב. 70.

סטטיסטיקה

פרק 8 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 24

סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחוזונים בטבלה בדידה:

רקע:

האחוזון (המאון) ה- p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת, שעד אליו (כולל) יש $p\%$ מהנתונים. מסמנים את האחוזון ה- p ב- X_p .

חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:

האחוזון הוא הערך שבו בפעם הראשונה השכיחות היחסית המצטברת (באחוזים) גדולה או שווה ל- $p\%$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו:

שכיחות יחסית מצטברת	שכיחות מצטברת	$F(x)$	# תוכניות החיסכון
		100	0
		75	1
		25	2
		25	3
		25	4

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

שאלות:

(1) להלן התפלגות של משתנה כלשהו:

$F(x)$	X
10	0
40	1
30	2
15	3
5	4

מצאו להתפלגות את:

- א. האחוזון ה-60.
- ב. המאון ה-40.
- ג. העשירון העליון.
- ד. הטווח בין הרבעונים.

(2) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

חשבו את:

- א. העשירון התחתון.
- ב. האחוזון ה-30.
- ג. הערך ש-20% מהתצפית גדולות ממנו.
- ד. רבעון עליון.

תשובות סופיות:

- (1) א. 2. ב. 1. ג. 3. ד. 1.
- (2) א. 1. ב. 2. ג. 4. ד. 4.

סטטיסטיקה

פרק 9 - סטטיסטיקה תיאורית- תרשים קופסא

תוכן העניינים

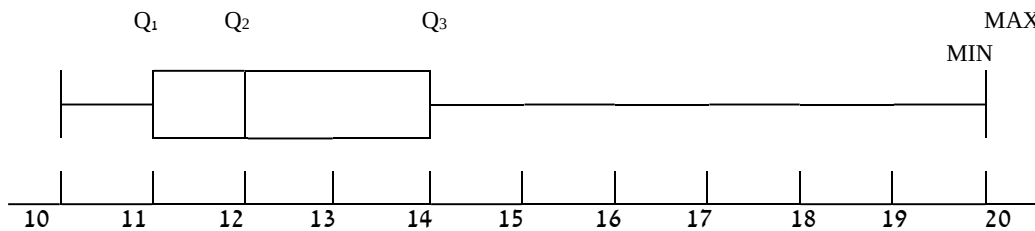
1. כללי.....26

סטטיסטיקה תיאורית – תרשים קופסא (Boxplot):

רקע:

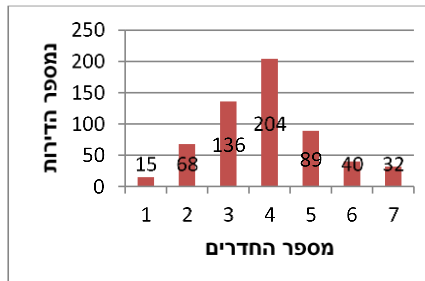
תרשים קופסא הינו תרשים שבעזרתו ניתן לבחון:

- (1) את המרכז של ההתפלגות על ידי החציון (Q_2).
- (2) את הפיזור של הנתונים (הטווח והטווח הבין רבעוני).
- (3) את צורת ההתפלגות (סימטרית \אסימטרית ימנית או אסימטרית שמאלית).



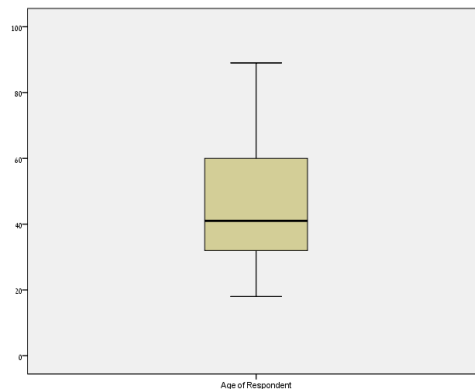
שאלות:

(1) להלן התפלגות מספר החדרים לדירות שנבנו בשנת 2009 בעיר אשדוד:



- א. מצאו את החציון, הרבעון התחתון והרבעון העליון של ההתפלגות.
- ב. שרטטו דיאגרמת קופסא להתפלגות.
- ג. מה ניתן לומר על צורת ההתפלגות?

(2) להלן דיאגרמת קופסא המתארת את התפלגות הגיל (בשנים) באוכלוסייה מסוימת:



- א. מה הגיל החציוני?
- ב. מה בערך טווח הגילאים?
- ג. מה ניתן להגיד על צורת ההתפלגות?

תשובות סופיות:

- (1) א. חציון: 4, רבעון תחתון: 3, רבעון עליון: 5.
 ב. ראה גרף מלא בסרטון וידאו. ג. כמעט סימטרית.
- (2) א. חציון: 40. ב. טווח: 70. ג. התפלגות אסימטרית ימנית.

סטטיסטיקה

פרק 10 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

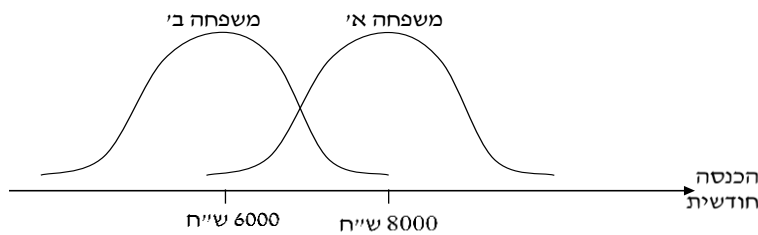
1. כללי 28

סטטיסטיקה תיאורית – שאלות אמריקאיות:

שאלות:

שאלות 1-3 מתייחסות לקטע הבא:

להלן שתי עקומות המתארות את התפלגות ההכנסות החודשיות של שתי משפחות שנבחרו באקראי:



- (1) לאיזו משפחה הכנסה שכיחה גבוהה יותר?
 - א. משפחה א'.
 - ב. משפחה ב'.
 - ג. לשתיהן אותה הכנסה שכיחה.
 - ד. לא ניתן לדעת – אין מספיק נתונים.
- (2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
 - א. משפחה א'.
 - ב. משפחה ב'.
 - ג. בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
 - ד. לא ניתן לדעת – אין מספיק נתונים.
- (3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
 - א. משפחה א'.
 - ב. משפחה ב'.
 - ג. לשתיהן אותה סטית תקן.
 - ד. לא ניתן לדעת – אין מספיק נתונים.

הנתונים הבאים מתייחסים לשאלות 4-6:

להלן נתונים חלקיים של טבלת שכיחויות:
כמו כן, נתון כי הממוצע הוא 1.66.

$F(x)$	x
?	0
10	1
6	2
15	3
?	4
50	סה"כ

(4) השכיח של הנתונים הוא:

- א. 0.
- ב. 15.
- ג. ישנם שני שכיחים: 0 ו-3.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של השכיח.

(5) חציון הנתונים הוא:

- א. 2.
- ב. 1.5.
- ג. 25.5.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(6) הטווח של הנתונים:

- א. 11.
- ב. 3.
- ג. 4.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(7) בהתפלגות אסימטרית ימנית של משתנה כמותי רציף, הערך המתאים למאון ה-30, ציון התקן שלו הוא בהכרח:

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

- 8) סדרת נתונים סטטיסטיים מונה 10 תצפיות. נתון כי סדרת הנתונים סימטרית סביב הממוצע. ממוצע הסדרה-40 ושונות הסדרה-100. בשלב מאוחר יותר נוספו שתי תצפיות נוספות לסדרה : 50 ו-30. השונות של 12 התצפיות :
- א. תקטן.
 - ב. תגדל.
 - ג. לא תשתנה.
 - ד. לא ניתן לחשב את השונות ללא ידיעת התצפיות.

הנתונים הבאים מתייחסים לשאלות 9-10:

בחברת "טיק" המשכורת הממוצעת היא 4,600 ₪ וסטיית התקן של משכורת זו הינה 200 ₪. לאחר מו"מ עם ועד עובדי ההנהלה סוכם כי המשכורת תוכפל פי 1.5.

- 9) מהי המשכורת הממוצעת החדשה (ב-₪)?

- א. 2,300.
- ב. 6,900.
- ג. 4,650.
- ד. 4,600.
- ה. חסרים נתונים כדי לדעת.

- 10) מהי סטיית התקן של המשכורת לאחר יישום המו"מ לגבי השכר (ב-₪)?

- א. 200.
- ב. 300.
- ג. 675.
- ד. לא ניתן לדעת.

- 11) 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 12-13:

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה:

שם התלמיד	ציון	ממוצע הציונים במועד בו נבחן	סטיית התקן של הציונים במועד בו נבחן
צבי	50	50	12
סטף	82	80	5
שרית	65	60	15
לובה	60	63	1.5
מיטב	70	70	10

12) התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא:

- א. מיטב.
- ב. צבי.
- ג. לובה.
- ד. שרית.
- ה. סטף.

13) פנינה נבחנה עם סטף וציון התקן שלה שווה לציון התקן של שרית לכן ציונה הוא:

- א. 80.55.
- ב. 65.
- ג. 80.
- ד. 81.66.

הנתונים הבאים מתייחסים לשאלות 14-16:

בבדיקת פתע של משרד הבריאות במפעל שוקולד, נמצא ש:

שוקולד פגום	0	1	2	3	4	5	6	7
מס' קופסאות	35	63	48	12	13	11	10	8

14) מהו החציון של מספר הפגומים בקופסא:

- א. 1.
- ב. 2.
- ג. 4.
- ד. לא ניתן לדעת.

15) מהו הרבעון התחתון של מספר הפגומים בקופסא?

- א. 1.
- ב. 2.
- ג. 3.
- ד. 4.
- ה. לא ניתן לדעת.

16) השכיח של מספר הפגומים בקופסא :

- א. 63.
- ב. 1.
- ג. 200.
- ד. לא ניתן לדעת.

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

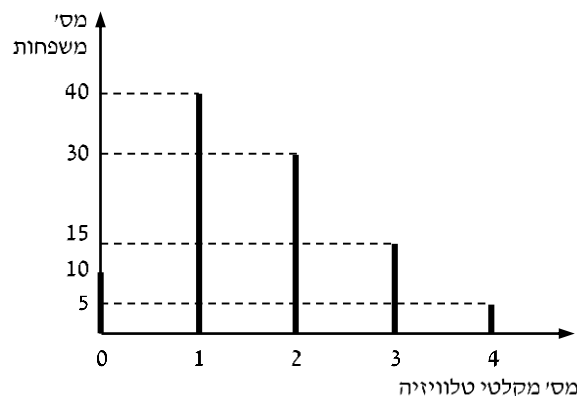
- א. בערכים הגבוהים.
- ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
- ג. בערכים הנמוכים.
- ד. לא ניתן לדעת.
- ה. אף לא תשובה מהנ"ל נכונה.

18) בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת.

- החציון והממוצע בשתיהן שווה 8.
- איזה מהטענות הבאות היא הנכונה והמלאה ביותר :
- א. השכיחות ב-2 החברות זהה אך שונה מ-8.
- ב. השכיח ב-2 החברות זהה אך לא ניתן לדעת מהו.
- ג. השכיח בשתי חברות הינו בהכרח 8.
- ד. שכיח בחברה אחת שונה מ-8 ובשנייה הוא 8.
- ה. אף תשובה אינה נכונה.

הנתונים הבאים מתייחסים לשאלות 19 עד 22 :

נערך סקר על מספר מקלטי הטלוויזיה הנמצאים בבית.
תוצאות הסקר נתונות בדיאגרמת מקלות הבאה :



19) הטווח של ההתפלגות הוא :

- א. 35.
- ב. 4.
- ג. 3.
- ד. 2.

20) ממוצע מספר מקלטי הטלויזיה למשפחה הוא :

- א. 1.65.
- ב. 1.5.
- ג. 1.
- ד. 2.

21) השכיח של התפלגות זו היא :

- א. 40.
- ב. 1.5.
- ג. 1.
- ד. 2.

22) מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלטי טלויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפיע על סטיית התקן?

- א. יקטין אותו.
- ב. יגדיל אותו.
- ג. לא ישנה אותו.
- ד. אין לדעת.

תשובות סופיות:

1) א'	2) ג'	3) ג'	4) ג'	5) ב'
6) ג'	7) א'	8) ג'	9) ב'	10) ב'
11) ג'	12) ה'	13) ד'	14) ב'	15) א'
16) ב'	17) ג'	18) ה'	19) ב'	20) א'
21) ג'	22) ב'			

סטטיסטיקה

פרק 11 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

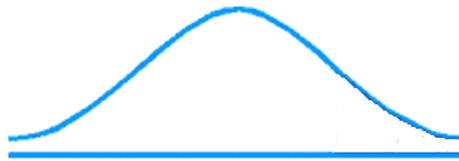
תוכן העניינים

1. כללי 34

התפלגויות רציפות מיוחדות – התפלגות נורמלית:

רקע:

התפלגות נורמלית הינה התפלגות של משתנה רציף. ישנם משתנים רציפים מסוימים שנהוג להתייחס אליהם כנורמליים כגון: זמן ייצור, משקל תינוק ביום היוולדו ועוד. פונקציית הצפיפות של ההתפלגות הנורמלית נראית כמו פעמון:



לעקומה זו קוראים גם עקומת גאוס ועקומה אחת נבדלת מהשנייה באמצעות הממוצע וסטיית התקן שלה.

אלה הם הפרמטרים שמאפיינים את ההתפלגות: $X \sim N(\mu, \sigma^2)$.

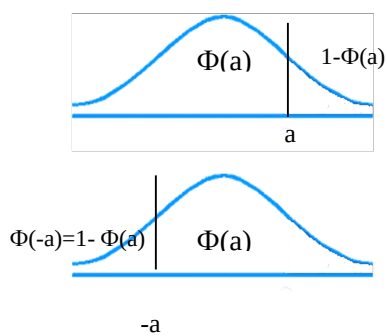
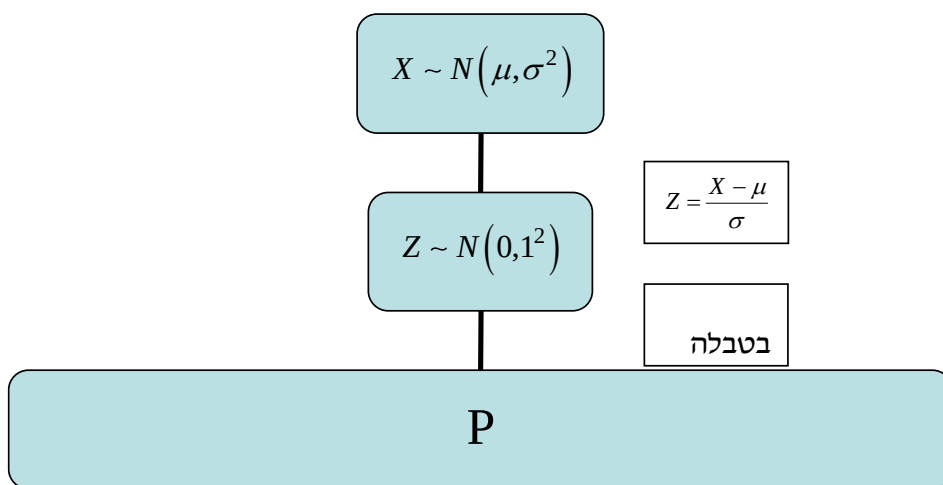
$$\text{נוסחת פונקציית הצפיפות: } f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

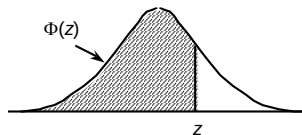
כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z : $Z \sim N(0, 1^2)$.

$$\text{תהליך התקנון מבוצע על ידי הנוסחה הבאה: } Z = \frac{X - \mu}{\sigma}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן. ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.

לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי, ובאופן כללי נתאר את הסכמה הבאה:



טבלת ההתפלגות המצטברת הנורמלית סטנדרטית – ערכי $\Phi(z)$:


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417
$\Phi(z)$	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995

דוגמה (הפתרון בהקלטה):

משקל חפיסות שוקולד המיוצרות בחברה מתפלג נורמלית עם ממוצע 100 גרם בסטיית תקן של 8 גרם.

- (1) מה אחוז חפיסות השוקולד ששוקלות מתחת ל-110 גרם?
- (2) מה אחוז חפיסות השוקולד ששוקלות מעל 110 גרם?
- (3) מה אחוז חפיסות השוקולד ששוקלות מתחת ל-92 גרם?
- (4) מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

שאלות:

- (1) הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ.
- מה אחוז האנשים שגובהם מתחת ל-182.4 ס"מ?
 - מה אחוז האנשים שגובהם מעל 190 ס"מ?
 - מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
 - מה אחוז האנשים שגובהם מתחת ל-170 ס"מ?
 - מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?
- (2) נתון שהזמן שלוקח לתרופה מסוימת להשפיע מתפלג נורמלית עם ממוצע של 30 דקות ושונות של 9 דקות רבועות.
- מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?
 - מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?
 - מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?
 - מה שיעור המקרים שבהן ההשפעה של התרופה תסטה מ-30 דקות בפחות מ-3 דקות?
- (3) המשקל של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 60 ק"ג וסטיית תקן של 8 ק"ג.
- מה אחוז האנשים שמשקלם נמוך מ-55 ק"ג?
 - מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
 - מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל-70 ק"ג?
 - לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ-4 ק"ג?
 - מה הסיכוי שאדם אקראי ישקול מתחת ל-140 ק"ג?
- (4) משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטית תקן 400 גרם.
- מצאו את העשירון העליון.
 - מצאו את האחוזון ה-95.
 - מצאו את העשירון התחתון.

- (5) ציוני מבחן אינטליגנציה מתפלגים נורמלית עם ממוצע 100 ושונות 225.
- מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - מה העשירון התחתון של ההתפלגות?
 - מהו הציון ש-20% מהנבחנים מקבלים מעליו?
 - מהו האחוזון ה-20?
 - מהו הציון ש-5% מהנבחנים מקבלים מתחתיו?
- (6) נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ"ל, ונתון ש-33% מהבקבוקים בעלי נפח שעולה על 508.8 מ"ל.
- מה ממוצע נפח משקה בבקבוק?
 - 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?
 - 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?
- (7) 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.
- מהו ממוצע אורך חיי מכשיר?
 - מהי סטית בתקן של אורך חיי מכשיר?
 - מה הסיכוי שמכשיר אקראי יחיה פחות מ-460 שעות?
 - מהו המאיון העליון של אורך חיי מכשיר?
 - 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?
- (8) להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל ביניהן.
- לאיזו התפלגות הממוצע הגבוה ביותר?
 - במה מבין המדדים הבאים התפלגות 1 ו-2 זהות?
 - בעשירון העליון.
 - בממוצע.
 - בשונות.
 - לאיזו התפלגות סטיית התקן הקטנה ביותר?
 - 1.
 - 2.
 - 3.
 - אין לדעת.



- (9) 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?
- ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משלושת רבעי השעה. מה ההסתברות שזמן הנסיעה הכולל יהיה פחות מ-50 דקות?
- ד. מה הסיכוי שבשבוע (חמישה ימי עבודה) בדיוק פעם אחת יהיה זמן הנסיעה לפחות שלושת רבעי השעה?

תשובות סופיות:

- | | | | | | |
|---|-----------|------------|-----------|-----------|---------|
| 1 | א. 89.25% | ב. 2.28% | ג. 0 | ד. 50% | ה. 50% |
| 2 | א. 0% | ב. 3.76% | ג. 0% | ד. 68.26% | |
| 3 | א. 26.43% | ב. 89.44% | ג. 39.44% | ד. 0.383 | |
| | ה. 100% | | | | |
| 4 | א. 3812.8 | ב. 3958 | ג. 2787.2 | | |
| 5 | א. 119.2 | ב. 80.8 | ג. 112.6 | ד. 87.4 | ה. 75.3 |
| 6 | א. 500 | ב. 532.9 | ג. 453.48 | | |
| 7 | א. 500 | ב. 100 | ג. 0.3446 | ד. 733 | |
| | ה. 267 | | | | |
| 8 | א. 3 | ב. בממוצע. | ג. 1 | | |
| 9 | א. 0.1587 | ב. 0.0228 | ג. 0.8563 | ד. 0.3975 | |

סטטיסטיקה

פרק 12 - התפלגויות בדידות מיוחדות - התפלגות בינומית

תוכן העניינים

1. כללי 41

התפלגויות בדידות מיוחדות – התפלגות בינומית:

רקע:

נגדיר את המושג ניסוי ברנולי :

ניסוי ברנולי הנו ניסוי שיש לו שתי תוצאות אפשריות: "הצלחה" ו"כישלון". למשל מוצר פגום או תקין, אדם עובד או מובטל, עץ או פלי בהטלת מטבע וכדומה. בהתפלגות בינומית חוזרים על אותו ניסוי ברנולי n פעמים באופן בלתי תלוי זה בזה. מגדירים את X להיות מספר ההצלחות שהתקבלו בסך הכול. נסמן ב- P את הסיכוי להצלחה בניסוי בודד, וב- Q את הסיכוי לכישלון בניסוי בודד. ואז נגיד ש: $X \sim B(n, p)$.

פונקציית ההסתברות של X :

$$P(X = K) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, 2, \dots, n$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}; \quad n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1; \quad 0! = 1$$

לגודל: $\binom{n}{k}$ ניתן לחשב באמצעות המחשבון.

$$E(X) = np \quad \text{תוחלת:}$$

$$V(X) = npq \quad \text{שונות:}$$

שימו לב, כדי לזהות שמדובר בהתפלגות בינומית צריכים להתקיים כל התנאים הבאים:

- (1) חוזרים על אותו ניסוי ברנולי באופן בלתי תלוי זה בזה.
- (2) חוזרים על הניסוי n פעמים.
- (3) X – מוגדר כמספר ההצלחות המתקבלות בסך הכול.

דוגמה (פתרון בהקלטה):

במדינה מסוימת ל-80% מהתושבים יש רישיון נהיגה. נבחרו 10 תושבים אקראיים מהמדינה.

- א. מה ההסתברות שבדיוק ל-9 מהם יש רישיון נהיגה?
- ב. מה ההסתברות שלפחות ל-9 מהם יש רישיון נהיגה?
- ג. מהי התוחלת ומהי סטיית התקן של מספר התושבים שנדגמו ושיש להם רישיון נהיגה?

שאלות:

- (1) במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסייה. נגדיר את X להיות מספר המובטלים שהתקבלו במדגם.
- מהי ההתפלגות של X ?
 - מה ההסתברות שיהיה בדיוק מובטל אחד?
 - מה ההסתברות שכולם יעבדו במדגם?
 - מה ההסתברות ששלושה יעבדו במדגם?
 - מה ההסתברות שלפחות אחד יהיה מובטל?
 - מה תוחלת ומהי השונות של מספר המובטלים במדגם?
- (2) על פי נתוני משרד התקשורת ל-70% מהאוכלוסייה יש סמארטפון. נבחרו 10 אנשים באקראי. נגדיר את X כמספר האנשים שנדגמו עם סמארטפון.
- מהי ההתפלגות של X ? הסבירו.
 - מה ההסתברות שבמדגם ל-8 אנשים יש סמארט-פון?
 - מה ההסתברות שבמדגם לפחות ל-9 יהיו סמארט-פון?
 - מה התוחלת ומה סטיית התקן של מספר האנשים שנדגמו ולהם סמארט-פון?
- (3) במדינה מסוימת התפלגות ההשכלה בקרב האוכלוסייה מעל גיל 30 היא כזו:
- | השכלה | נמוכה | תיכונית | תואר I | תואר II ומעלה |
|-----------|-------|---------|--------|---------------|
| פרופורציה | 0.1 | 0.6 | 0.2 | 0.1 |
- נבחרו 20 אנשים אקראיים מעל גיל 30.
- מה ההסתברות ש-5 מהם אקדמאים?
 - מה התוחלת של מס' בעלי ההשכלה הנמוכה?
- (4) במכללה מסוימת 20% מהסטודנטים גרים בת"א. מבין הסטודנטים שגרים בת"א 30% מגיעים ברכבם, ומבין הסטודנטים שלא גרים בת"א 50% מגיעים ברכבם למכללה.
- השומר בשער המכללה בודק לכל סטודנט את תיקו בהיכנסו למכללה. מה ההסתברות שבקרב 5 סטודנטים שנבדקו ע"י השומר רק 1 מתוכם הגיע למכללה ברכבו?
 - בהמשך לסעיף הקודם מה ההסתברות שרוב הסטודנטים בקרב ה-5 הגיעו למכללה ברכבם?

תשובות סופיות:

(1) א. $X \sim B(n=5, p=0.1)$ ב. 0.32805 ג. 0.59049

ד. 0.0729 ה. 0.40954 ו. תוחלת: 0.5, שונות: 0.45

(2) א. 0.2335 ב. 0.1493 ד. תוחלת: 7, סטיית תקן: 1.449

(3) א. 0.1789 ב. 2

(4) א. 0.1956 ב. 0.4253

סטטיסטיקה

פרק 13 - הסקה סטטיסטית - הקדמה

תוכן העניינים

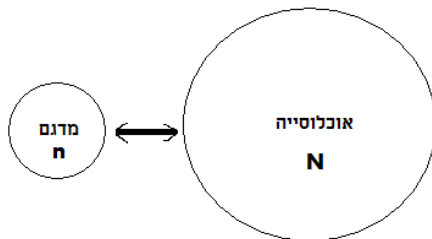
1. כללי 44

הסקה סטטיסטית – הקדמה:

רקע:

אוכלוסייה:

קבוצה שאליה מפנים שאלה מחקרית. למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.



מדגם:

חלק מתוך האוכלוסייה. למשל, אם נדגום באקראי 10 אנשים מתוך חולי הסוכרת אז זהו מדגם מתוך אוכלוסיית חולי הסוכרת.

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה. הדגימה בקורס תהיה דגימה מקרית - הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

סטטיסטי:

גודל המחושב על המדגם.

פרמטר:

גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים:

סטטיסטי (מדגם)	פרמטר (אוכלוסייה)	ממוצע
$\bar{X}$	μ	
$\hat{p}$	P	פרופורציה (שכיחות יחסית)

פרמטר הוא גודל קבוע גם אם אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן יש לו התפלגות הנקראת התפלגות הדגימה.

דוגמה (פתרון בהקלטה):

25% מאזרחי המדינה תומכים בהצעת החוק של חבר כנסת מסוים. הוחלט לדגום 200 אזרחים ומתוכם לבדוק מהו אחוז התומכים בהצעת החוק.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
- (2) להלן התפלגות מספר מקלטי הטלויזיה למשפחה ב"שוב "העוגן". נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלויזיה במדגם.
- מי היא האוכלוסייה ומהו המשתנה הנחקר?
 - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

מספר המשפחות	מספר מקלטים
50	0
250	1
350	2
300	3
50	4
סך הכול $N = 1000$	

- (3) נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.
- מהי האוכלוסייה?
 - מה המשתנה באוכלוסייה?
 - מהם הפרמטרים?
 - מהו הסטטיסטי?

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון. ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
- (2) א. האוכלוסייה: 1000 משפחות ב"שוב העוגן, המשתנה הנחקר: מס' מקלטים. ב. $\bar{X}$ = ממוצע מדגם.
- (3) א. השכירים במדינה. ב. השכלה: אקדמאי, לא אקדמאי. ג. שיעור ההצלחות באוכלוסייה: 0.2. ג. מס' האקדמאים במדגם.

סטטיסטיקה

פרק 14 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי 47
2. התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי 53
3. התפלגות מספר ההצלחות במדגם - קירוב נורמלי להתפלגות הבינומית 55
4. התפלגות פרופורציית ההצלחות במדגם 60

התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

רקע:

בפרק זה נדון בהתפלגות של ממוצע המדגם: $\bar{x} = \frac{\sum x_i}{n}$.

מכיוון שממדגם למדגם אנו יכולים לקבל ממוצע מדגם שונה, אזי ממוצע המדגם הוא משתנה מקרי ויש לו התפלגות.

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים.

להלן רשימה של פרמטרים החשובים לפרק זה:

ממוצע האוכלוסייה נסמן ב- μ (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- σ^2 .

סטיית תקן של אוכלוסייה: σ .

תכונות התפלגות:

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה: $E(\bar{x}) = \mu_{\bar{x}} = \mu$.

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- n .

תכונה זו נכונה רק במדגם מקרי: $V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$.

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקראת גם

טעות תקן: $\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.

דוגמה (פתרון בהקלטה):

השכר הממוצע במשק הינו 9000 ₪ עם סטיית תקן של 4000. דגמו באקראי 25 עובדים.

א. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

דגימה מהתפלגות נורמאלית:

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2 .

$$\text{ממוצע המדגם גם יתפלג נורמאלית: } Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}, \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

דוגמה (פתרון בהקלטה):

משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

משפט הגבול המרכזי:

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק

$$\text{גדול } (n \geq 30) \text{ ממוצע המדגם מתפלג בקירוב נורמאלי: } \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

דוגמה (פתרון בהקלטה):

משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם וסטיית תקן של 4 גרם.
דגמו מקו הייצור 36 חפיסות שוקולד אקראיות.
מה ההסתברות שהמשקל הממוצע של חפיסות השוקולד שנדגמו יהיה מתחת ל-102 גרם?

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מיהי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
 - מהו תוחלת ממוצע המדגם?
 - מהי טעות התקן?
- (2) משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם? נתון כי ביום מסוים נולדו 4 תינוקות.
 - מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
 - הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?
- (3) הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.
- מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יסטה מתוחלת הגבהים בפחות מ-5 ס"מ?
 - מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

- (4) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך הפתרון 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 גבוה ממנו?
- ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
- ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?
- (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?
- הסבר את ההבדל בתשובות של שני הסעיפים.
- (7) 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 ₪?

8) 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. היעזרו במשפט הגבול המרכזי.

9) נתון ש- $X \sim N(\mu, \sigma^2)$. דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע

המדגם $\bar{X}$. לכן: $P(\bar{X} > \mu)$ יהיה (בחרו בתשובה הנכונה):

א. 0.

ב. 0.5.

ג. 1.

ד. לא ניתן לדעת.

10) נתון ש- X מתפלג כלשהו עם תוחלת μ ושונות σ^2 .

החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים (בחרו בתשובה הנכונה):

א. $X \sim N\left(\mu, \frac{\sigma^2}{200}\right)$.

ב. $\mu \sim N\left(\mu, \frac{\sigma^2}{200}\right)$.

ג. $\bar{X} \sim N(\mu, \sigma^2)$.

ד. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{200}\right)$.

11) נתון ש- $X \sim N(\mu, \sigma^2)$. אם נדגום n תצפיות מתוך ההתפלגות ונגדיר: $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$,

אזי (בחרו בתשובה הנכונה):

א. μ ו- $\bar{X}$ יהיו משתנים מקריים.

ב. μ יהיה משתנה מקרי ו- $\bar{X}$ קבוע.

ג. $\bar{X}$ יהיה משתנה מקרי ו- μ קבוע.

ד. μ ו- $\bar{X}$ יהיו קבועים.

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון.
ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
ה. 78. ו. 10.6.
- (2) א. 0.8413. ב. 0.0013. ג. 0. ד. 0.1974.
- (3) א. 0. ב. 0. ג. 0.9544. ד. 178.205.
- (4) א. 0.0465. ב. 27.71. ג. 0.2628. ד. התשובה הייתה קטנה.
- (5) א. 0. ב. 0.1587. ג. 0.1587. ד. 0.5.
- (6) א. 0.5468. ב. 0.6826.
- (7) 0.0475.
- (8) א. 0.9772. ב. 0.0228. ג. 271.
- (9) ב'.
- (10) ד'.
- (11) ג'.

התפלגות סכום תצפיות בלתי תלויות ומשפט הגבול המרכזי:

רקע:

כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם: $T = \sum_{i=1}^n X_i$.
 כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה, כלומר, היו: $X_1, \dots, X_n$ -
 משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2 אזי:
 התוחלת והשונות של סכום התצפיות: $E(T) = n\mu$, $V(T) = n\sigma^2$.

דגימה מתוך התפלגות נורמלית:

אם: $X \sim N(\mu, \sigma^2)$, אזי: $Z = \frac{T - n\mu}{\sqrt{n\sigma^2}}$, $T \sim N(n\mu, n\sigma^2)$.

משפט הגבול המרכזי:

אם X מתפלג כלשהו וידוע כי: $E(X) = \mu$, $V(X) = \sigma^2$, אזי עבור מדגם מספיק גדול (לפחות 30): $T \rightsquigarrow N(n\mu, n\sigma^2)$.

דוגמה (פתרון בהקלטה):

- בעיר מסוימת המשכורת הממוצעת של עובד הינה 8000 ₪. עם סטיית תקן של 2000 ₪. נדגמו 100 עובדים מהעיר שמפקידים את משכורותיהם לסניף בנק.
- מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?
 - מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע"י אותם עובדים?

שאלות:

- (1) המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק"ג וסטיית תקן של 10 ק"ג.
- א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק"ג?
- ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק"ג?
- ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק"ג?
- (2) נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ"ל וסטיית תקן של 20 מ"ל. אדם קנה מארז של 4 בקבוקי יין.
- א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?
- ב. את היין שבמארז האדם מזג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?
- ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א' הייתה משתנה? האם התשובה לסעיף ב' הייתה משתנה?
- (3) בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב-4 דקות עם סטיית תקן של 1 דקות.
- א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?
- ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?

תשובות סופיות:

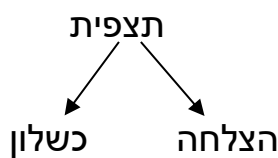
- (1) א. 0.6915 ב. 0.8413 ג. 0.5
- (2) א. תוחלת 3000 מ"ל, סטיית תקן 40 מ"ל. ב. 0.0062
- ג. סעיף א'- לא משתנה, סעיף ב'- לא פתיר, התבסס על התפלגות נורמלית.
- (3) א. 0.0571 ב. 2036.8

התפלגות מספר ההצלחות במדגם – קירוב נורמלי להתפלגות הבינומית:

רקע:

תזכורת על התפלגות בינומית:

בפרק זה נדון בהתפלגות מספר ההצלחות במדגם אקראי (תצפיות בלתי תלויות זו בזו). את מספר ההצלחות במדגם נסמן ב- Y . מחלקים כל תצפית במדגם להצלחה או כישלון.



כעת מה שמשתנה מתצפית לתצפית הוא משתנה דיכוטומי (משתנה שיש לו שני ערכים). הסיכוי להצלחה יסומן עם הפרמטר p וכישלון יסומן ע"י הפרמטר $q = 1 - p$. מבצעים מדגם אקראי בגודל n : $Y \sim B(n, p)$.

פונקציית ההסתברות של ההתפלגות הבינומית היא: $p(y = k) = \binom{n}{k} p^k q^{n-k}$.

תוחלת: $E(y) = np$.

שונות: $V(y) = npq$.

קירוב נורמלי עבור התפלגות בינומית:

אם לפנינו התפלגות בינומית: $Y \sim B(n, p)$, ומתקיים ש:

$$1. \quad n \cdot p \geq 5$$

$$2. \quad n \cdot (1 - p) \geq 5$$

$$y \sim N(np, npq)$$

$$\text{אז: } Z_y = \frac{y - np}{\sqrt{npq}}$$

תיקון רציפות:

כאשר משתמשים בקירוב הנורמלי להתפלגות הבינומית יש לבצע תיקון רציפות. הסיבה שעוברים כאן מהתפלגות בדידה להתפלגות נורמלית שהיא התפלגות רציפה. על פי הכללים הבאים:

$$1. \quad p(Y = a) \cong p\left(a - \frac{1}{2} \leq Y \leq a + \frac{1}{2}\right)$$

$$2. \quad P(Y \leq a) \cong P(Y \leq a + 0.5)$$

$$3. \quad P(Y \geq a) \cong P(Y \geq a - 0.5)$$

הערות:

- התנאים למעבר מבינומי לנורמלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:

$$1. \quad n \cdot p \geq 5$$

$$2. \quad n \cdot (1 - p) \geq 5$$

- ישנם מרצים שנותנים את התנאי המחמיר הבא:

$$1. \quad n \cdot p \geq 10$$

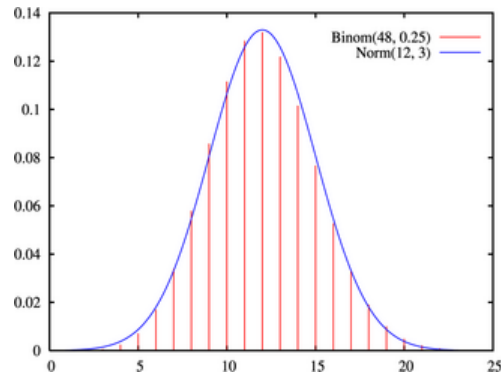
$$2. \quad n \cdot (1 - p) \geq 10$$

- וישנם מרצים שהתנאי שהם נותנים הוא: $(n \geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור מהתפלגות בינומית לנורמלית.
- הערה נוספת היא לגבי תיקון רציפות. ישנם מרצים שלא מחייבים לבצע תיקון רציפות שהמדגמים גדולים (בדרך כלל מעל 100 תצפיות) בפתרונות שאציג תמיד אבצע תיקון רציפות במעבר מבינומי לנורמלי כיוון שכך הפתרון יהיה יותר מדויק (בכל מקרה שהמדגמים גדולים העניין זניח).

דוגמה (הפתרון בהקלטה):

נתון שבקרב אוכלוסיית הנוער 25% זקוקים למשקפיים. נדגמו באקראי 48 בני נוער.

1. מה הסיכוי שבדיוק 14 מתוכם יהיו זקוקים למשקפיים?
2. מה הסיכוי שלכל היותר 13 מתוכם זקוקים למשקפיים?



שאלות:

- (1) נתון ש-20% מאוכלוסייה מסוימת אקדמאית. נבחרו באקראי 10 אנשים באותה אוכלוסייה.
 - א. מה ההסתברות ששלושה מהם אקדמאים?
 - ב. מה ההסתברות שלכל היותר אחד מהם אקדמאי?
 - ג. מה התוחלת ומהי סטיית התקן של מספר האקדמאים במדגם?
- (2) במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.
 - א. מה ההסתברות שנדגמו לפחות 6 מוצרים פגומים?
 - ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?
- (3) ציוני פסיכומטרי בקרב הנרשמים למוסד מסוים מתפלגים נורמאלית עם ממוצע 500 וסטיית תקן 100. למוסד מסוים הוחלט לקבל אך ורק סטודנטים שקיבלו מעל 600 בפסיכומטרי. 100 סטודנטים אקראיים נרשמו למוסד. מה ההסתברות שלפחות 20 יתקבלו?
- (4) מטילים מטבע 50 פעמים.
 - א. מה ההסתברות לקבל לכל היותר 30 עצים?
 - ב. מה ההסתברות לקבל 28 עצים לפי התפלגות הבינומית ולפי הקירוב הנורמאלי?
- (5) במטוס מקום ל-400 נוסעים. נרשמו לטיסה 430 אנשים (overbooking). מנתונים סטטיסטיים ידוע שהסיכוי שאדם שנרשם לטיסה אכן יגיע הוא 0.9.
 - א. מה ההסתברות שלא יהיו מקומות ישיבה לכל האנשים שהגיעו לטיסה?
 - ב. מה צריך להיות גודל המטוס כדי שבסיכוי שלפחות 95% המטוס יספיק לכמות הנרשמים?
- (6) מפעל לייצור ארטיקים טוען שהסיכוי שארטיק שהוא מייצר יהיה פגום הוא 0.01. מוכר הזמין 1000 ארטיקים מהמפעל. מה ההסתברות שהמוכר יקבל לפחות 980 ארטיקים תקינים אם טענת המפעל מוצדקת?
- (7) מהמר מטיל קובייה הוגנת 100 פעמים. בכל הטלה, אם מתקבל תוצאה זוגית בקובייה המהמר זוכה בשקל. אחרת, המהמר משלם שקל. המהמר הטיל את הקובייה 100 פעמים מה הסיכוי שהרווח של המהמר יהיה לכל היותר 10?

תשובות סופיות:

- (1) א. 0.201 ב. 0.3758 ג. התוחלת: 2, סטיית התקן: 1.2649.
- (2) א. 0.9332 ב. 0.6915
- (3) 0.1611
- (4) א. 0.9406 ב. בינומית - 0.0788, קירוב לנורמלית - 0.0778.
- (5) א. 0.015 ב. 0.398
- (6) 0.9996
- (7) 0.8643

התפלגות פרופורציית ההצלחות במדגם:

רקע:

בפרק זה נדון בהתפלגות הדגימה של פרופורציית המדגם.
 Y - מספר ההצלחות במדגם (למשל, מספר המובטלים במדגם).

$$\hat{p} = \frac{Y}{n} \text{ - פרופורציית ההצלחות במדגם.}$$

למשל, שיעור המובטלים במדגם - $n = 200$:

מספר המובטלים : $Y = 20$.

$$\hat{p} = \frac{20}{200} = 0.1 \text{ : פרופורציית המובטלים במדגם}$$

נסמן ב- p את שיעור ההצלחה באוכלוסייה וב- q את שיעור הכישלונות באוכלוסייה.
 נבצע מדגם מקרי (הנחה שהתצפיות בלתי תלויות זו בזו) ונתבונן בהתפלגות של פרופורציית המדגם.

התוחלת, השונות וסטיית התקן של פרופורציית המדגם:

$$E(\hat{p}) = p, \quad V(\hat{p}) = \frac{pq}{n}$$

משפט הגבול המרכזי עבור הפרופורציה המדגמית:

$$\text{אם: } np \geq 5 \text{ \& } nq \geq 5, \text{ אזי: } \hat{p} \sim N\left(p, \frac{pq}{n}\right) \quad Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

הערות:

- התנאים לקרוב הנורמאלי הם נזילים, כלומר משתנים ממרצה אחד לשני.
 התנאי שהצגתי כאן הוא הפופולרי ביותר:
 1. $n \cdot p \geq 5$
 2. $n \cdot (1 - p) \geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
 1. $n \cdot p \geq 10$
 2. $n \cdot (1 - p) \geq 10$
- וישנם מרצים המשתמשים בתנאי: $(n \geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור לנורמלית.

- כיוון שפרופורציה אינה חייבת להיות מספר שלם בהכרח לא נהוג לבצע כאן תיקון רציפות.

דוגמה (פתרון בהקלטה):

לפי נתוני משרד החינוך בעיר ירושלים ל-60% מתלמידי התיכון זכאים לתעודת בגרות. נדגמו 200 תלמידי תיכון.

- מה ההסתברות שהשכיחות היחסית ($\hat{p}$) של הזכאים לבגרות במדגם תעלה על 60%?
- מה ההסתברות שפרופורציית הזכאים לבגרות במדגם תעלה על 70%?

שאלות:

- (1) במדינה מסוימת 10% מכלל האוכלוסייה הינם מובטלים. נדגמו באקראי 140 אנשים מהמדינה.
 - א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
 - ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
 - ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?
- (2) נניח כי 30% מהאוכלוסייה תומכים בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
 - א. לפחות 35% יתמכו בהצעת החוק במדגם.
 - ב. לכל היותר 25% יתמכו בהצעת החוק במדגם.
 - ג. יותר מ-27% יתמכו בהצעת החוק במדגם.
- (3) לפי נתוני משרד התקשורת 40% מהאוכלוסייה מחזיקים בטלפון נייד מסוג "סמארטפון". נדגמו 400 אנשים מהאוכלוסייה.
 - א. מה ההסתברות שבמדגם לכל היותר ל-40% יש סמארטפון?
 - ב. מה ההסתברות שבמדגם לרוב יש סמארטפון?
 - ג. מה ההסתברות שפרופורציית בעלי הסמארטפון במדגם תסטה מהפרופורציה באוכלוסייה בלא יותר מ-4%?
 - ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם?
- (4) נתון כי 80% מבתי האב מחוברים לאינטרנט. נדגמו 400 בתי אב אקראיים.
 - א. מה ההסתברות שלפחות 340 מהם מחוברים לאינטרנט?
 - ב. מה ההסתברות שפרופורציית המחוברים לאינטרנט במדגם תסטה מהפרופורציה האמיתית ביותר מ-4%?
 - ג. כמה בתי אב יש לדגום כדי שהסטייה בין הפרופורציה המדגמית לפרופורציה האמיתית לא תעלה על 3% בהסתברות של 90%?
 - ד. מהו העשירון התחתון של התפלגות פרופורציית המדגם?
- (5) נתון שציוני פסיכומטרי מתפלגים נורמלית עם תוחלת 500 וסטיית תקן 100. ב"מועדון ה-700" נכללים נבחנים שמקבלים ציון מעל 700 בפסיכומטרי. מה הסיכוי שבמועד בו נבחנו 2000 נבחנים אקראיים יהיו לפחות 3% המשתייכים למועדון?

(6) נתון ש- $X \sim B(n, p)$, ונגדיר את המשתנה הבא: $\hat{P} = \frac{X}{n}$.

א. הוכיחו ש: $E(\hat{P}) = p$, $V(\hat{P}) = \frac{p(1-p)}{n}$.

ב. מה p המביא את $V(\hat{P})$ להיות במקסימום?

תשובות סופיות:

- (1) א. התוחלת: 0.1, השונות: 0.00064. ב. 0.5. ג. 0.3446.
- (2) א. 0.0618. ב. 0.0618. ג. 0.8238.
- (3) א. 0.5. ב. 0. ג. 0.8968. ד. גדלה.
- (4) א. 0.0062. ב. 0.0456. ג. 0.481. ד. 0.77436.
- (5) 0.0154.
- (6) א. שאלת הוכחה. ב. 0.5.

סטטיסטיקה

פרק 15 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 64

מושגי יסוד באמידה:

רקע:

כזכור מהמפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת. כמו ממוצע הגבהים בקרב מתגייסים לצה"ל - μ . כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה - p . בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

- נסמן באופן כללי פרמטר באות θ ואומד ב- $\hat{\theta}$. הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את θ .
- שגיאת אמידה: $|\hat{\theta} - \theta|$ - ההפרש בין האומד לאמת (הפרמטר).

דוגמה (פתרון בהקלטה):

בכנסת ה-19 קיבלה מפלגת העבודה 15 מנדטים. בערוץ 10 ברגע סגירת הקלפיות העריכו את מספר המנדטים של המפלגה להיות 17 מנדטים וזאת על סמך תוצאות מדגם של הערוץ.

- א. מה הפרמטר בדוגמה זו?
 - ב. מהי טעות האמידה של ערוץ 10?
- $\hat{\theta}$ יהיה אומד חסר הטיה ל- θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל- θ : $E(\hat{\theta}) = \theta$.
 - טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma(\hat{\theta}) = S.E$

פרמטרים מרכזיים והאומדים שלהם:**ממוצע האוכלוסייה μ :**

האומד הנקודתי שלו יהיה: ממוצע המדגם: $\bar{x} = \frac{\sum x}{n}$.

$E(\bar{x}) = \mu$ לכן $\bar{x}$ הינו אומד חסר הטיה ל- μ . כמו כן, טעות תקן: $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE$.

פרופורציה באוכלוסייה p :

האומד הנקודתי שלו יהיה: פרופורציה במדגם: $\hat{p} = \frac{y}{n}$.

$E(\hat{p}) = p$, לכן $\hat{p}$ הינו אומד חסר הטיה ל- p . כמו כן טעות התקן: $\sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$.

שונות האוכלוסייה σ^2 :

האומד הנקודתי שלו יהיה: $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$.

$E(S^2) = \sigma^2$ ולכן S^2 הינו אומד חסר הטיה ל- σ^2 : $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

הערה: אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

דוגמה (פתרון בהקלטה):

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה.
להלן התוצאות שהתקבלו: 2, 1, 3, 2, 1, 4, 5, 2, 1, 3.
אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

1. ממוצע מספר הילדים למשפחה בתל אביב.
2. שונות מספר הילדים למשפחה בתל אביב.
3. פרופורציית המשפחות בנות שני ילדים.

שאלות:

- (1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן כשהמדגם בגודל 500?
 - מהו האומדן לפרמטר?
 - מהי טעות האמידה?
- (2) לפי נתוני היצרן, מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה.
- במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן?
 - מהו האומדן לפרמטר?
 - מהי טעות האמידה?
- (3) נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו: 168, 184, 192, 171, 180, 177, 187, 168, 177 ו-175.
- מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
- (4) נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים.
- להלן התוצאות שהתקבלו: $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X_i^2 = 1502.2$.
- אמדו את השכר הממוצע של השכירים במשק.
 - אמדו את סטיית התקן של שכר השכירים במשק.
- (5) במטרה לאמוד את ממוצע האוכלוסייה, דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
- סטיית התקן של האוכלוסייה.
 - סטיית התקן של ממוצע האוכלוסייה.
 - סטיית התקן של המדגם.
 - סטיית התקן של ממוצע המדגם.

6) משקל הממוצע של אוכלוסייה מסוימת הוא 75 ק"ג עם שונות של 25. אם יבחרו כל המדגמים האפשריים בגודל 10 מאוכלוסייה זו סטיית התקן של ממוצעי המדגמים תהייה:

- א. 3.
- ב. 2.5.
- ג. 1.581.
- ד. אין מספיק נתונים לדעת.

7) במדגם מקרי, מתי סכום ריבועי הסטיות מהממוצע, $\sum_{i=1}^n (x_i - \bar{x})^2$, מחולק ב- $n-1$?

- א. כאשר n קטן.
- ב. כאשר תצפיות המדגם אינן בלתי תלויות.
- ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.
- ד. כאשר מעוניינים באומד חסר הטיה לשונות האוכלוסייה ממנה הוצא המדגם.
- ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

8) $X_1, X_2, \dots, X_{16}$ מדגם מקרי מתוך אוכלוסייה בעלת ממוצע μ לא ידוע ושונות: $\sigma^2 = 64$. טעות התקן של האומד ל- μ היא:

- א. 16.
- ב. 8.
- ג. 4.
- ד. 2.

9) מהו אומד חסר הטיה?

- א. אומד שערכו שווה לממוצע התפלגות הדגימה שלו.
- ב. אומד שערכו שווה לערך הפרמטר באוכלוסייה.
- ג. אומד שממוצע התפלגות הדגימה שלו שווה לערך הפרמטר באוכלוסייה.
- ד. אומד שהסיכוי שערכו יהיה גבוה מערך הפרמטר באוכלוסייה שווה לסיכוי שיהיה נמוך ממנו.

- | | | | |
|---|----------|---------|---------|
| א. 0.25 | ב. 0.019 | ג. 0.24 | ד. 0.01 |
| א. אוכלוסייה : מקררים של יצרן, תוחלת : 2400, סטיית תקן : 500. | | | |
| ב. 100 | | ג. 2342 | ד. 58 |
| א. 177.9 | ב. 64.1 | ג. 0.4 | |
| א. 8.1 | ב. 3.16 | | |
| ד' | | | |
| ג' | | | |
| ד' | | | |
| ד' | | | |
| ג' | | | |

סטטיסטיקה

פרק 16 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

1. קביעת גודל מדגם 69
2. רווח סמך כששונות האוכלוסיה לא ידועה 71

קביעת גודל מדגם:

רקע:

אם מעוניינים לאמוד את ממוצע האוכלוסייה כאשר סטיית התקן של האוכלוסייה ידועה: σ ברמת סמך של $1-\alpha$ ושגיאת אמידה שלא תעלה על ε מסוים, נציב

$$.n \geq \left(\frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{\varepsilon} \right)^2 \quad \text{בנוסחה הבאה:}$$

כדי להציב בנוסחה צריך שהמשתנה הנחקר יתפלג נורמלית או שהמדגם ייצא בגודל של לפחות 30 תצפיות.

דוגמה (פתרון בהקלטה):

חברת תעופה מעוניינת לאמוד את תוחלת משקל המטען של נוסע. נניח שמשקל מטען של נוסע מתפלג נורמאלית עם סטיית תקן של 2 ק"ג. כמה נוסעים יש לדגום אם מעוניינים שבביטחון של 98% הסטייה המרבית בין ממוצע המדגם לממוצע האמיתי לא יעלה על 0.5 ק"ג? (תשובה: 87).

שאלות:

- (1) משתנה מקרי מתפלג נורמאלית עם סטיית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רווח סמך ברמת סמך של 98% שאורכו לא יעלה על 2?
- (2) מעוניינים לאמוד את הדופק הממוצע של מתגייסים לצבא. מעוניינים שבביטחון של 95% שגיאת האמידה המרבית תהיה 0.5. נניח שהדופק מתפלג נורמאלית על סטיית תקן של 3 פעימות לדקה.
- א. כמה מתגייסים יש לדגום?
- ב. אם ניקח מדגם הגדול פי 4 מהמדגם של סעיף א ונאמוד את הממוצע באותה רמת סמך כיצד הדבר ישפיע על שגיאת האמידה?
- (3) יהי X משתנה מקרי עם ממוצע μ וסטיית תקן σ . חוקר רוצה לבנות רווח בר סמך ל- μ ברמת ביטחון של 0.95, כך שהאורך של הרווח יהיה 0.5σ . מהו גודל המדגם הנדרש?

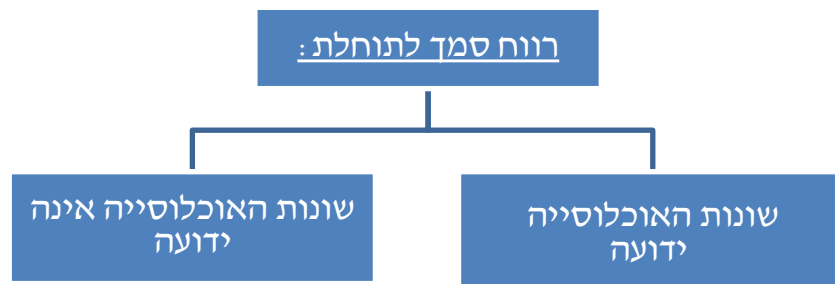
תשובות סופיות:

- (1) 780.
- (2) א. 139. ב. הדבר יקטין את ε פי 2.
- (3) $n = 62$.

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:

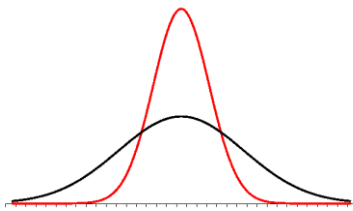


בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו.

מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול.

רווח סמך: $\bar{X} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \cdot \frac{S}{\sqrt{n}}$



האומד לשונות: $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$

התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה

להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים.

התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן: $df = n-1$.

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה.

כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית.

במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות

שהתקבלו בדקות: 4.7, 5.2, 4.6, 5.3.

בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

שאלות:

- (1) מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב. ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 84, 88, 84, 79, 89. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
 - א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
 - ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
 - ג. בהמשך לסעיף א', אם היינו בונים את רווח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רווח הסמך?
- (2) במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן: $S = 13$ ס"מ. בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?
- (3) אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 30, 40, 32, 34, 27.
 - א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
 - ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?
- (4) ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
 - א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
 - ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.
 - ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.
- (5) נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$.
 - א. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

- (6) נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו: $\bar{x} = 13.8$, $S = 2$. בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.
- (7) שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר μ . לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10. שניהם קבעו אותה רמת סמך. סטטיסטיקאי א': הניח $\sigma = 20$. סטטיסטיקאי ב': חישב לפי המדגם וקיבל $S = 20$. למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר?
- א. סטטיסטיקאי א'.
 ב. סטטיסטיקאי ב'.
 ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.
 ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.
- (8) נתון ש: $X \sim N(\mu, \sigma^2)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח הסמך שהתקבל הוא: 8.765. מהי רמת הביטחון של רווח הסמך?

תשובות סופיות:

- (1) א. $79.88 < \mu < 89.72$ ב. כן. ג. הוא היה גדל.
- (2) ראה בסרטון.
- (3) א. צריך להניח שהמשתנה מתפלג נורמלית. ב. לא ניתן לדעת.
- (4) א. $96.63 < \mu < 107.37$ ב. $96.90 < \mu < 107.10$ ג. ראה בסרטון.
- (5) $3.149 < \mu < 3.351$
- (6) $13.42 < \mu < 14.18$
- (7) ב'.
- (8) 90%.

סטטיסטיקה

פרק 17 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

74	1. הקדמה
78	2. סוגי טעויות

הקדמה:

רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה. בבדיקת השערות על פרמטרים נעבוד לפי השלבים הבאים:

שלב א: נוהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.

השערת האפס המסומנות ב- H_0 .

בדרך כלל השערת האפס מסמלת את אשר היה מקובל עד עכשיו, את השגרה, הנורמה.

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

ההשערה האלטרנטיבית מסמלת את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה היא שאלת המחקר.

שלב ג: נבדוק האם התנאים לביצוע התהליך מתקיימים ונניח הנחות במידת הצורך.

שלב ד: נרשום את כלל ההכרעה. בתהליך של בדיקת השערות יוצרים כלל שנקרא כלל הכרעה. הכלל יוצר אזורי שנקראים:

1. **אזור דחייה:**

דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.

2. **אזור קבלה:**

קבלה של השערת האפס ודחייה של האלטרנטיבה. כלל ההכרעה מתבסס על איזשהו סטטיסטי. אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שנקרא רמת מובהקות ומסומן ב- α .

שלב ה: בתהליך יש ללכת לתוצאות המדגם ולחשב את הסטטיסטי המתאים ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה.

שלב ו: להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה (פתרון בהקלטה):

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום לידתם בישראל 3300 גרם. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20, \bar{X} = 3120, S = 280$$

- א. מהי אוכלוסיית המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

שאלות:

בשאלות הבאות, ענו על הסעיפים הבאים:

- א. מהי אוכלוסיית המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

- (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.
- (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.
- (3) במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.
- (4) בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

תשובות סופיות:

- (1) א. נבחנים בבגרות באנגלית.
 ב. ציון.
 ג. ממוצע הציונים בשיטת לימוד חדשה.
 ד. $H_0: \mu = 72$
 $H_1: \mu > 72$
- (2) א. משקאות בבקבוק של חברה מסוימת.
 ב. נפח משקה בסמ"ק.
 ג. ממוצע נפח המשקה בבקבוק.
 ד. $H_0: \mu = 500$
 $H_1: \mu < 500$
- (3) א. מועמדים לפקולטה למשפטים.
 ב. משתנה דיכוטומי (התקבל, לא התקבל).
 ג. אחוז הקבלה.
 ד. $H_0: p = 0.25$
 $H_1: p < 0.25$
- (4) א. אזרחים בוגרים במשק.
 ב. משתנה דיכוטומי (מובטל, עובד).
 ג. אחוז האבטלה כיום.
 ד. $H_0: p = 0.08$
 $H_1: p \neq 0.08$

סוגי טעויות:

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה.
הכלל יוצר אזורים שנקראים:

1. אזור דחייה – דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור קבלה – קבלה של השערת האפס ודחייה של האלטרנטיבה.

כלל ההכרעה מתבסס על איזשהו סטטיסטי.
בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת. לכן יתכנו טעויות במסקנות שלנו:

	הכרעה		
		H_0	H_1
	מצאות	H_0	H_1
		אין טעות	טעות מסוג 1
		טעות מסוג 2	אין טעות

הגדרת הטעויות:

טעות מסוג ראשון: להכריע לדחות את H_0 למרות שבמצאות H_0 נכונה.

טעות מסוג שני: להכריע לקבל את H_0 למרות שבמצאות H_1 נכונה.

דוגמה (פתרון בהקלטה):

אדם חשוד בביצוע עבירה ונתבע בבית המשפט.
אילו סוגי טעויות אפשריות בהכרעת הדין?

שאלות:

- (1) 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. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
- רשמו את השערות המחקר.
 - מה מסקנת המחקר?
 - איזו סוג טעות יתכן וביצעו במחקר?
- (2) במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
- האם ניתן לדעת אם בוצע טעות במחקר?
 - מה סוג הטעות האפשרית?
- (3) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
- מהי אוכלוסיית המחקר?
 - מה המשתנה הנחקר?
 - מה הפרמטר הנחקר?
 - מה השערות המחקר?
 - מה מסקנת המחקר?
 - מהי סוג הטעות האפשרית במחקר?

תשובות סופיות:

- (1) א. $H_0: \mu = 500$
 ב. לא דחינו את H_0 .
 ג. טעות מסוג שני.
- (2) א. לא ניתן לדעת.
 ב. טעות מסוג ראשון.
- (3) א. משפחות כיום.
 ב. מס' הילדים.
 ג. תוחלת מספר הילדים למשפחה כיום.
 ד. $H_0: \mu = 2.3$
 $H_1: \mu < 2.3$
 ה. לא לדחות את H_0 . ו. טעות מסוג שני.

סטטיסטיקה

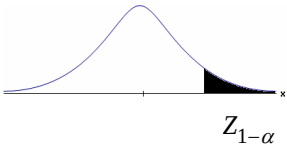
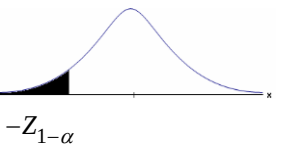
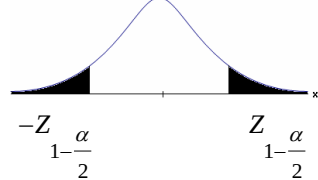
פרק 18 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה. 80
2. סיכוי לטעויות ועוצמה (ששונות האוכלוסיה ידועה). 84
3. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה) 90
4. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה. 95
5. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה לא ידועה). 99
6. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע) 102

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$Z_{\bar{x}} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}} < -Z_{\frac{\alpha}{2}}$ או $Z_{\bar{x}} > Z_{\frac{\alpha}{2}}$  $-Z_{\frac{\alpha}{2}}$ $Z_{\frac{\alpha}{2}}$ דוחים את H_0

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

סטטיסטי המבחן:

חלופה אחרת לכלל הכרעה:

$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	נדחה H_0 אם מתקיים:
--	--	--	-----------------------

דוגמה:

יבול העגבניות מתפלג נורמלית עם תוחלת של 10 טון לדונם וסטיית תקן של 2.5 טון לדונם בעונה. משערים ששיטת זיבול חדשה תעלה את תוחלת היבול לעונה מבלי לשנות את סטיית התקן. נדגמו 4 חלקות שזוכלו בשיטה החדשה. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדקו את ההשערה ברמת מובהקות של 1%.

פיתרון:

אוכלוסייה: עגבניות.

המשתנה: $X =$ יבול העגבניות בטון לעונה.

הפרמטר: $\mu =$ תוחלת היבול בשיטת הזיבול החדשה.

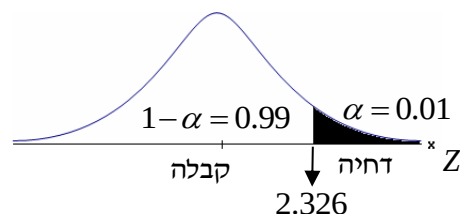
השערות:
 $H_0: \mu = 10$
 $H_1: \mu > 10$

תנאים:

1. $X \sim N$.

2. $\sigma = 2.5$.

כלל הכרעה:



נדחה את H_0 אם $Z_{\bar{X}} > 2.326$

תוצאות: $n = 4$, $\bar{x} = 12.5$

סטטיסטי המבחן: $Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

נציב: $Z_{\bar{X}} = \frac{12.5 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$

מסקנה:

לא נדחה H_0 (נקבל H_0).

ברמת מובהקות של 1% לא נוכל לקבל את הטענה ששיטת הזיבול החדשה מעלה את תוחלת היבול של העגבניות.

שאלות:

- (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 תהיה 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 מכיילת (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%?
- (4) המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
- (5) לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ. קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.
- (6) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחרו בתשובה הנכונה.
 - א. הגדלת רמת המובהקות לא תשנה את מסקנת המחקר.
 - ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 - ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 - ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

(7) חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס.

אם החוקר היה עורך מבחן צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח:

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס הייתה לא נדחית.
- ג. לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

(8) שני סטטיסטיקאים בדקו השערות: $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$,

עבור שונות ידועה ובאותה רמת מובהקות.

שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.

- א. אם חוקר א' החליט לדחות את H_0 , מה יחליט חוקר ב'? נמקו.
- ב. אם חוקר א' יחליט לא לדחות את H_0 , מה יחליט חוקר ב'? נמקו.

תשובות סופיות:

(1) נקבל H_0 , בר"מ של 5% לא נקבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.

(2) א. נדחה H_0 , בר"מ של 2.5% נקבל את תלונת אגודת הצרכנים בדבר הפחתת נפח המשקה בבקבוק.

ב. הגדלנו את רמת המובהקות לכן אנחנו נשארים בדחייה של H_0 והמסקנה לא תשתנה.

(3) נדחה H_0 , בר"מ של 5% נקבע שהמכונה לא מאופסת.

(4) נדחה H_0 , בר"מ של 0.1 נקבל את הטענה שהדיאטה יעילה ומפחיתה את המשקל הממוצע.

(5) נקבל H_0 , בר"מ של 0.05 נכריע שתוחלת עובי הבורג מתים למפרט.

(6) א'.

(7) ג'.

(8) א. לדחות. ב. לא ניתן לדעת.

סיכוי לטעויות ועוצמה (ששונות האוכלוסייה ידועה):

רקע:

מציאות	הכרעה		
		H_0	H_1
	H_0	אין טעות	טעות מסוג 1
	H_1	טעות מסוג 2	אין טעות

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$$\alpha = P(H_0 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ לדחות})$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל את } H_0 \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון:

$$1 - \alpha = P(H_0 \text{ לקבל את } H_0 \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה:

$$1 - \beta = \pi = P(H_1 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ לדחות})$$

התהליך לחישוב סיכוי לטעות מסוג שני:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
חישוב β :	$P_{\mu_1} \left(\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	$P_{\mu_1} \left(\bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	$P_{\mu_1} \left(\mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$

התפלגות ממוצע המדגם: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

התקנון:

דוגמה:

בתחילת השנה חשבון הטלפון הסלולארי הממוצע לאדם היה 200 ₪ עם סטיית תקן של 80 ₪ לחודש. בעקבות כניסתן של חברות טלפון סלולארית חדשות מעוניינים לבדוק האם כיום ממוצע חשבון הטלפון הסלולארי פחת. לצורך בדיקה דגמו באקראי 36 אנשים וחשבון הטלפון הסלולארי שלהם היה 150 ₪ בממוצע לחודש.

א. רשמו את השערות המחקר ובנו כלל הכרעה במונחי חשבון ממוצע מדגמי ברמת מובהקות של 5%.

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

ג. נניח שבמציאות כיום החשבון הממוצע הוא 160 ₪. מה הסיכוי לבצע טעות מסוג שני?

ד. אם נקטין את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

פתרון:

א. אוכלוסייה: משלמי חשבון טלפון סלולאר כיום.

המשתנה: X = חשבון הטלפון החודשי בשקלים.

הפרמטר: μ .

השערות:
 $H_0: \mu = 200$
 $H_1: \mu < 200$

תנאים:

1. $\mu = 200$.

2. $n = 36$.

$$\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, K = \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\alpha = 0.05$$

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$

$$\text{נציב: שקלים } K = 200 - 1.645 \cdot \frac{80}{\sqrt{36}} = 178.07$$

כלל ההכרעה: דחה את H_0 אם שקלים $\bar{X} < 178.07$.

ב. ברמת מובהקות של 5% נכריע שאכן ממוצע חשבון הטלפון הסלולרי פחת מתחילת השנה.

ג. השערות:
 $H_0: \mu_0 = 200$
 $H_1: \mu < 200$

כלל ההכרעה: נדחה את H_0 אם $\bar{X} < 178.07$.

$$H_1: \bar{X} \sim N\left(160, \frac{80^2}{36}\right)$$

$$Z = \frac{178.07 - 160}{\frac{80}{\sqrt{36}}} = 1.36$$

$$\beta = P_{H_1} \left(\overset{\text{לקבל את}}{H_0} \right) = P_{H_1} (\bar{X} > 178.07) = 1 - \phi(136) = 1 - 0.9131 = 0.0869$$

ד. הקטנת α מגדילה את β .

שאלות:

(1) נתון ש: $X \sim N(\mu, \sigma^2 = 1)$.

- להלן השערות של חוקר לגבי הפרמטר μ : $H_0: \mu = 5$, $H_1: \mu = 7$.
 מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.
 א. עבור אילו ערכים של X שידגם נדחית השערת H_0 ?
 ב. מה הסיכוי לבצע טעות מסוג שני?
 ג. אם במדגם התקבל ש- $X = 6.9$ מה תהיה המסקנה ומה הטעות האפשרית?

- (2) 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 העצמה של הכלל מסעיף א'?

- (3) להלן נתונים על תהליך של בדיקת השערות על תוחלת:

$$H_0: \mu = 200, H_1: \mu \neq 200, \sigma = 30, n = 225$$

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

- (4) 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. רשמו את ההשערות ואת כלל 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. הסבירו ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.
 ד. הסבירו ללא חישוב כיצד התשובה לסעיף ב תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

- (5) להלן השערות של מחקר: $H_0: \mu = 50$, $H_1: \mu = 58$. מעוניינים לדגום 100 תצפיות. ידוע שסטיית התקן של ההתפלגות הינה 20.
- א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%. מהי רמת המובהקות?
- ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?
- i. סטיית התקן הייתה יותר גדולה.
- ii. הסיכוי לטעות מסוג שני גדול יותר.

השאלות שלהלן הן שאלות רב-ברירה, בחרו בתשובה הנכונה ביותר:

- (6) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:
- א. הסיכוי לטעות מסוג ראשון גדל.
- ב. העוצמה של המבחן גדלה.
- ג. הסיכוי לטעות מסוג שני גדל.
- ד. תשובות א' ו-ב' נכונות.
- (7) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:
- א. השערת האפס נכונה.
- ב. השערת האפס נדחתה.
- ג. השערת האפס לא נדחתה.
- ד. אף אחת מהתשובות לא נכונה בהכרח.

- (8) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

$1 - \beta$	α
גדולה	א. גדולה
קטנה	ב. גדולה
גדולה	ג. קטנה
קטנה	ד. קטנה

- (9) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך:
- א. הן α , והן $1 - \beta$, יקטנו.
- ב. α יישאר ללא שינוי ואילו $1 - \beta$ יגדל.
- ג. α יגדל ואילו $1 - \beta$ יקטן.
- ד. הן α והן $1 - \beta$ יגדלו.

- (10) ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?
- טעות מסוג ראשון.
 - טעות מסוג שני.
 - טעות מסוג שלישי.
 - אין טעות במסקנתו.

תשובות סופיות:

- (1) א. מעל 6.645. ב. 0.3594.
ג. דחינו את H_0 , תתכן טעות מסוג ראשון.
- (2) א. נדחה H_0 אם $\bar{X} < 2.24$. ב. נדחה H_0 ג. 1.
- (3) א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$. ב. 0.8051. ג. תקטן.
- (4) א. נדחה H_0 אם $\bar{X} < 48.9$. ב. 0.0885. ג. תקטן. ד. תקטן.
- (5) א. 0.0033. ב. i. רמת המובהקות הייתה קטנה.
ב. ii. רמת המובהקות הייתה גדלה.
- (6) ד'.
- (7) ג'.
- (8) ג'.
- (9) א'.
- (10) ב'.

מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה ידועה):

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.

המסקנה של המחקר תקבע לפי העיקרון הבא: אם $p_v \leq \alpha$, דוחים את H_0 . מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני) $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני) $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

שערת האפס: השערה אלטרנטיבה:		$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	
1. σ ידועה		תנאים:	
2. $X \sim N$ או מדגם מספיק גדול		p-value	
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \Leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$	

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}, \quad \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad \text{כאשר בהנחת השערת האפס:}$$

דוגמה:

המשקל הממוצע של מתגייסים לצבא לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שמשקל המתגייסים מתפלג נורמאלי עם סטיית תקן של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

א. מהי מובהקות התוצאה?

ב. מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא 1%?

פתרון:

א. אוכלוסייה: המתגייסים לצבא כיום.

משתנה: X = משקל בק"ג.

פרמטר: μ .

השערות:
 $H_0: \mu = 65$
 $H_1: \mu > 65$

תנאים:

1. $X \sim N$.

2. $\sigma = 12$.

תוצאות מדגם:

$$n = 16$$

$$\bar{X} = 71$$

$$P_V = P_{H_0} \left(\begin{array}{c} \text{לתוצאות} \\ \text{המדגם} \\ \text{וקיצוני} \end{array} \right) = P_{H_0} (\bar{X} \geq 71) = 1 - \phi(2) = 1 - 0.9772 = 0.0228$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{71 - 65}{\frac{12}{\sqrt{16}}} = 2$$

$$\alpha_{\min} = 0.0228$$

שאלות:

- (1) להלן השערות של מחקר: $H_0: \mu = 70$, $H_1: \mu > 70$. המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות: $n = 100$, $\bar{x} = 74$. מהי מובהקות התוצאה?
- (2) השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?
- (3) אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ-100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטיית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן להגיע להחלטה. לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.
 - א. רשמו את השערות המחקר.
 - ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?
 - ג. מהי רמת המובהקות המקסימלית עבורה נקבל את השערת האפס?
 - ד. מה המסקנה ברמת מובהקות של 5?
- (4) מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג נורמלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.
 - א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?
 - ב. אם נוסיף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?
 - ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.
- (5) אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון/לא נכון? נמק.

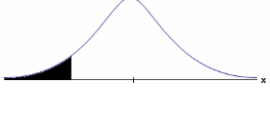
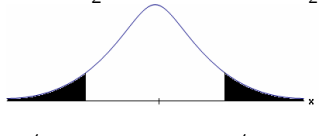
- (6) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$.
 מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?
 בחרו בתשובה הנכונה.
- א. יקבל את השערת האפס בכל מקרה.
 - ב. ידחה את השערת האפס מקרה.
 - ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
 - ד. לא ניתן לדעת כי אין מספיק נתונים.
- (7) מובהקות התוצאה (PV) היא גם (בחרו בתשובה הנכונה):
- א. רמת המובהקות המינימאלית לדחות השערת האפס.
 - ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
 - ג. רמת המובהקות שנקבעת מראש על ידי החוקר שטרם קיבל את תוצאות המחקר.
 - ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.
- (8) בבדיקת השערות מסוימת התקבל: $p\text{ value} = 0.0254$ לכן (בחרו בתשובה הנכונה):
- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
 - ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
 - ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
 - ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

תשובות סופיות:

- (1) 0.0228.
- (2) עבור כל רמת מובהקות סבירה.
- (3) א. $H_0: \mu = 100$
 $H_1: \mu < 100$
 ב. 0.1056 ג. 0.1056.
- ד. נכריע שיש עמידה בהתחייבות של החברה.
- (4) א. 0.0006 ב. יקטן. ג. נכריע שאין כיול.
- (5) נכון.
- (6) א'.
- (7) א'.
- (8) ג'.

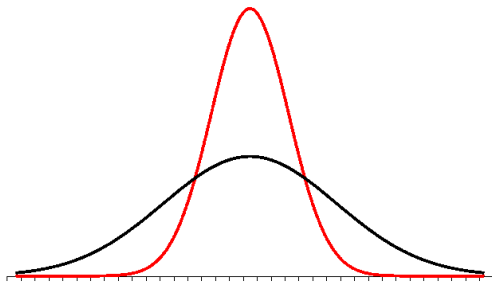
בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה לא ידועה:

רקע:

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	1. σ אינה ידועה 2. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{x}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את H_0
חלופה לכלל ההכרעה: נדחה H_0 אם מתקיים:	$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ או $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$

$$. t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} : \text{סטטיסטי המבחן}$$

$$. S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$



התפלגות T:

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש.

דרגות החופש הן: $df = n - 1$.

ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

דוגמה (פתרון בהקלטה):

מפעל קיבל הזמנה לייצור משטחים בעובי של 0.1 ס"מ. כדי לבדוק האם המפעל עומד בדרישה נדגמו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיית תקן 0.002 ס"מ.

א. מהן השערות המחקר?

ב. מה ההנחה הדרושה לצורך פתרון?

ג. בדוק ברמת מובהקות של 5%.

שאלות:

(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) משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היוולדם בישראל 3300 גר'. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

מה מסקנתכם ברמת מובהקות של 5% מה יש להניח לצורך פתרון?

(3) ציוני מבחן אינטליגנציה מתפלגים נורמלית. בארה"ב ממוצע הציונים הוא 100. במדגם שנעשה על 23 נבחנים ישראלים, התקבל ממוצע ציונים 104.5 וסטיית התקן המדגמית 16. האם בישראל ממוצע הציונים שונה מארה"ב? הסיקו ברמת מובהקות של 5%.

(4) באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{10} X_i = 750$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית.

בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80.

- (5) ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם. ליאור השתמש בטבלה של התפלגות Z . רוני השתמשה בטבלה של התפלגות t . מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.
- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
 - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
 - שני החוקרים בהכרח יגיעו לאותה מסקנה.
 - לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

- (6) נתון ש: $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$. $H_1: \mu < \mu_0$. חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות. בחר בתשובה הנכונה:
- א. כעת בברור הוא ידחה את השערת האפס.
 - ב. כעת הוא דווקא יקבל את השערת האפס.
 - ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

תשובות סופיות:

- (1) נדחה H_0 .
- (2) נדחה H_0 .
- (3) נקבל H_0 .
- (4) נקבל H_0 .
- (5) ב'.
- (6) ג'.

מובהקות תוצאה – אלפא מינימלית (ששונות האוכלוסייה לא ידועה):

רקע:

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם $p_v \leq \alpha$ דוחים את H_0 . מובהקות התוצאה היא הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

(לקבל את תוצאות המדגם וקיצוני) $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

(לקבל את תוצאות המדגם וקיצוני) $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	השערת האפס: השערה אלטרנטיבה:
1. σ אינה ידועה או 2. מדגם מספיק גדול $X \sim N$			תנאים:
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \Leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$	p-value

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

$$d.f = n-1$$

דוגמה:

ממוצע זמן הנסיעה של אדם לעבודה הינו 40 דקות. הוא מעוניין לבדוק דרך חלופית שאמורה להיות יותר מהירה. לצורך כך הוא דוגם 5 ימים שבהם הוא נוסע בדרך החלופית. זמני הנסיעה שקיבל בדקות הם: 34, 40, 30, 32, 27. הניחו שזמן הנסיעה מתפלג נורמלי.

- א. רשמו את השערות המחקר.
- ב. מצאו חסמים למובהקות התוצאה.
- ג. מה המסקנה ברמת מובהקות של 5%?

פתרון:

אוכלוסייה: כלל הנסיעות לעבודה בדרך החלופית.

משתנה: X = זמן נסיעה בדקות.

תנאים: $X \sim N$.

פרמטר: μ .

א. השערות:
 $H_0: \mu = 40$
 $H_1: \mu < 40$

ב. תוצאות המדגם:

$$n = 5, \bar{X} = \frac{\sum X_i}{n} = \frac{34 + 40 + \dots}{5} = 32.6$$

$$S^2 = \frac{\sum X_i^2 - n \cdot \bar{X}^2}{n-1} = \frac{34^2 + 40^2 - \dots - 5 \cdot 32.6^2}{5-1} = 23.4$$

$$S = \sqrt{23.4}$$

$$t_{\bar{X}} = \frac{\frac{\bar{X} - \mu_0}{S}}{\frac{1}{\sqrt{n}}} = \frac{32.6 - 40}{\frac{4.88}{\sqrt{5}}} = -3.39$$

$$P_V = P_{H_0} = (\bar{X} \leq 32.6) = P(t \leq -3.39)$$

$$d.f = 5 - 1 = 4$$

$$1\% < P_V < 2.5\%$$

$P_V < \alpha = 0.05$, לכן דוחים את H_0 .

מסקנה: בר"מ של 5% נכריע שהדרך החלופית מהירה יותר.

שאלות:

(1) קו ייצור אריזות סוכר נארזות כך שהמשקל הממוצע של אריזות הסוכר צריך להיות אחד קילוגרם. בכל יום דוגמים מקו הייצור 5 אריזות במטרה לבדוק האם קו הייצור תקין. בבדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים: 1024, 1008, 1005, 996, 997.

א. רשמו את השערות המחקר.

ב. מהי מובהקות התוצאה? הצג חסמים.

ג. מה המסקנה ברמת מובהקות של 5%?

(2) חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות. מהי ה- α המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר?

(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%?

תשובות סופיות:

$$(1) \quad \begin{array}{ll} \text{א.} & H_0: \mu = 1000 \\ & H_1: \mu \neq 1000 \\ \text{ב.} & 20\% \leq P_v \leq 50\% \end{array}$$

ג. ברמת מובהקות של 5% לא נוכל לקבוע שקו הייצור אינו תקין.

(2) 10%.

(3) 1.01, נקבל את H_0 .

הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע):

רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על μ :

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

על ידי בניית רווח סמך ברמת סמך של $1-\alpha$ ל- μ :

אם μ_0 נופל ברווח $\leftarrow$ נקבל את H_0 .

אם μ_0 לא נופל ברווח $\leftarrow$ נדחה את H_0 .

דוגמה:

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו :

$$H_0: \mu = 80, H_1: \mu \neq 80, \alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90% וקיבל: $79 < \mu < 84$.
האם אפשר לדעת מה מסקנתו, ואם כן מהי?

פתרון (פתרון מלא בהקלטה):

רווח הסמך ברמת סמך של 90% מכיל "80".

ברמת סמך של 95% רווח הסמך יגדל ויכיל "80".

לכן, ברמת מובהקות של 5% נקבל H_0 .

שאלות:

- (1) 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.
- (2) 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 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.
א. בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.
ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבירו.
- (3) יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצלין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.
א. בנו רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
ב. בדקו ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

תשובות סופיות:

- (1) נקבל השערת.
(2) א. $112.87 \leq \mu \leq 118.13$.
ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.
(3) א. $191.8 \leq \mu \leq 200.2$. ב. נכריע שיש אמת בפרסום.

סטטיסטיקה

פרק 19 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששונניות האוכלוסיה ידועות.....104
2. כששונניות האוכלוסיה לא ידועות ובהנחת שוויון שונניות.....106
3. כששונניות האוכלוסיה לא ידועות והמדגמים גדולים.....108

כששונויות האוכלוסייה ידועות:

רקע:

המטרה היא לאמוד את פער התוחלות: $\mu_1 - \mu_2$, כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$.

התנאים לבניית רווח הסמך:

1. σ_1^2, σ_2^2 ידועות.

2. $X_1, X_2 \sim N$ או $n_1, n_2 > 30$.

3. שני מדגמים בלתי תלויים.

רווח סמך: $(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של $1-\alpha$, לא קיים הבדל בין התוחלות.

דוגמה (פתרון בהקלטה):

נדגמו 100 תושבים מאזור A והמשכורת הממוצעת הייתה ש"ח 9200.
כמו כן נדגמו 120 תושבים מאזור B וממוצע המשכורות שהתקבל ש"ח 8700.
לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ש"ח. אמדו ברמת סמך של 90% את הפרש השכר הממוצע בין אזור A לאזור B.

שאלות:

- (1) מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?
- (2) ציוני IQ מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99. א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה"ב בממוצע הציונים במבחן ה-IQ. ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?
- (3) חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25. עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

תשובות סופיות:

- (1) $(-20, 90)$.
- (2) א. $-3.99 < \mu_1 - \mu_2 < 13.99$.
 ב. לא נוכל לטעון בביטחון של 95% שקיים הבדל בין ישראל לארה"ב.
- (3) רמות בטחון הגבוהות מ-0.9476.

כששונויות האוכלוסייה לא ידועות ובהנחת שוויון שונויות:

רקע:

המטרה היא לאמוד את פער התוחלות: $\mu_1 - \mu_2$, כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי: $\bar{X}_1 - \bar{X}_2$.

התנאים לבניית רווח הסמך:

$$1. \sigma_1^2 = \sigma_2^2.$$

$$2. X_1, X_2 \sim N.$$

3. מדגמים בלתי תלויים.

השונויות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונויות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{הנוסחה הבאה:}$$

$$d.f = n_1 + n_2 - 2 \quad \text{דרגות החופש:}$$

$$\text{רווח סמך: } (\bar{X}_1 - \bar{X}_2) \pm t_{\frac{1-\alpha}{2}, n_1+n_2-2} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של $1 - \alpha$, לא קיים הבדל בין התוחלות.

דוגמה (פתרון בהקלטה):

מחקר מעוניין לבדוק האם קיים הבדל בין תל אביב לבאר שבע מבחינת ההכנסה הממוצעת של אקדמאים. להלן תוצאות המדגם שנעשה:

מספר האקדמאים	תל אביב	באר שבע
10	20	
ממוצע הכנסות של אקדמאים	11,000	9500
סטיית התקן של הכנסות אקדמאים	200	250

בנו רווח סמך ברמת ביטחון של 90% להפרש תוחלות ההכנסה בשני האזורים. הניחו שהשכר מתפלג נורמלית עם אותה שונות בכל אחד מהאזורים.

שאלות:

- (1) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

המדינה	ישראל	ארה"ב
גודל המדגם	15	15
סכום הציונים	1560	1470
סכום ריבועי הציונים	165,390	147,560

מצאו רווח סמך ברמת סמך של 95% לסטייה בין ממוצע הציונים בישראל לממוצע הציונים בארה"ב. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (2) להלן 4 תצפיות על משתנה X שמתפלג: $N(\mu_x, \sigma^2)$, ומשתנה Y שמתפלג: $N(\mu_y, \sigma^2)$.

X	22	20	21	25
Y	18	25	17	12

חשבו רווח סמך ל- $\mu_y - \mu_x$ ברמת הסמך 90%, בהנחה ששני המדגמים בלתי תלויים.

תשובות סופיות:

- (1) הנחות:
- השונות שווה.
 - שהציונים מתפלגים נורמלית.
 - המדגמים אינם תלויים זה בזה.
- $$-5.52 < \mu_1 - \mu_2 < 17.52$$
- (2) $-9.6 < \mu_y - \mu_x < 1.6$

כששונויות האוכלוסייה לא ידועות והמדגמים גדולים:

רקע:

המטרה היא לאמוד את פער התוחלות: $\mu_1 - \mu_2$, כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$.

התנאים לבניית רווח הסמך:

1. מדגמים גדולים.

2. מדגמים בלתי תלויים.

$$\text{רווח סמך: } (\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של $1-\alpha$, לא קיים הבדל בין התוחלות.

דוגמה (פתרון בהקלטה):

יבואן רכב מעוניין להשוות בין רכבים יפנים לעומת רכבים אמריקאים מבחינת תצרוכת הדלק שלהם. נדגמו 40 רכבים מכל סוג. לכל רכב נבדק כמות הק"מ שהמכונית נסעה עבור 1 ליטר דלק במהירות של 100 קמ"ש. התוצאות שהתקבלו במדגם הם: ברכבים היפנים ממוצע 12.5 עם סטיית תקן 1.5 וברכבים האמריקאים הממוצע 11.3 עם סטיית תקן של 1.4.

א. בנו רווח סמך ברמת סמך של 95% לפער הממוצעים של צריכת הדלק של מכוניות יפניות לעומת אמריקאיות.

ב. האם על סמך רווח הסמך ניתן לקבוע שקיים הבדל בין שני סוגי הרכבים מבחינת צריכת הדלק?

שאלות:

- (1) במדגם של 45 תלמידים הלומדים משפטים התקבל שבשבוע הם לומדים בממוצע 4 שעות עם סטיית תקן 2 שעות מעבר לשעות הלימוד בכיתה. במדגם של 55 תלמידים הלומדים הנדסה התקבל שבשבוע הם לומדים בממוצע 10 שעות עם סטיית תקן של 3 שעות.
- א. אמדו את הפרש ממוצעי שעות הלמידה של סטודנטים למשפטים לעומת סטודנטים להנדסה ברמת בטחון של 90%.
- ב. האם קיים הבדל בין סטודנטים למשפטים ולסטודנטים להנדסה מבחינת ההשקעה שלהם מעבר לשעות הלימוד בכיתה?
- (2) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלו נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות וסטיית תקן של 0.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות וסטיית תקן של 1.1 שעות. מצאו רווח סמך לפער בין ממוצע שעות הצפייה בטלוויזיה בין שני האזורים בביטחון של 95%.

תשובות סופיות:

- (1) א. $-5.17 < \mu_1 - \mu_2 < -6.83$. ב. כן.
- (2) כן, בביטחון של 90% סטודנטים להנדסה לומדים יותר מסטודנטים למשפטים.

סטטיסטיקה

פרק 20 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 110

רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים:

רקע:

מדגם מזווג: מדגם אחד שבו יש n צמדים. כל תצפית במדגם תנפק זוג ערכים: X ו- Y .

ניצור משתנה חדש: $D = x - y$.

הפרמטר שנרצה לאמוד: μ_D .

התנאים לבניית רווח הסמך:

$$1. \quad x, y \sim N$$

2. המדגם מזווג.

$$\text{נוסחת רווח הסמך: } \bar{D} \pm t_{1-\frac{\alpha}{2}}^{n-1} \frac{S_D}{\sqrt{n}}$$

כאשר דרגות החופש: $df = n - 1$.

דוגמה (פתרון בהקלטה):

מעוניינים לבדוק האם יש הבדל בין מהירות הריצות של שתי תוכנות מחשב. לקחו 5 קבצים אקראיים והריצו אותם בשתי התוכנות:

הקובץ	1	2	3	4	5
הזמן בתוכנה הראשונה	25	48	49	46	38
הזמן בתוכנה השנייה	27	46	42	40	48

הניחו כי זמני הריצות מתפלגים נורמלית.

מצאו רווח סמך של 95% להפרש תוחלת הזמן בין שתי התוכנות.

שאלות:

- (1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים בסמסטר א' ו-ב' :

82	75	90	68	74	סמסטר א'
100	76	87	84	80	סמסטר ב'

נניח שהציונים מתפלגים נורמאלית.

- א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.
- ב. האם על סמך רווח הסמך קיים הבדל בין הסמסטרים מבחינת תוחלת הציונים?
- ג. מה צריך לשנות בנתונים כדי שהמדגמים יהיו בלתי תלויים?
- (2) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות :

חברה/ מדינה	ארה"ב	קנדה	הולנד	פולין	מצרים	סין	יפן
בזק - X	1.5	2.1	2.2	3	3.5	3.2	4.2
קווי זהב - Y	1.4	2	1.9	3.1	3.3	3.2	4.2

בהנחה והמחירים מתפלגים נורמלית עבור כל חברה, בנו רווח סמך ברמת סמך של 90% לתוחלת הפרש המחירים של שתי החברות.

תשובות סופיות:

- (1) א. $-19 < \mu_0 < 38$. ב. בביטחון של 95% לא קיים הבדל.
- ג. ראה הסבר בסרטון.
- (2) $-0.013 < \mu < 0.185$.

סטטיסטיקה

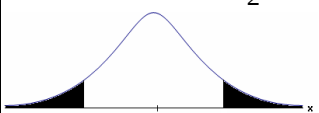
פרק 21 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסייה ידועות..... 112
2. כששונויות האוכלוסייה אינן ידועות והמדגמים גדולים..... 116
3. כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות..... 119

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונות של האוכלוסייה ידועות – רקע

$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$	השערת האפס: השערה אלטרנטיבית:
מדגמים בלתי תלויים σ_1, σ_2 ידועות $X_1, X_2 \sim N$ או מדגמים מספיק גדולים			תנאים:
$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0 : $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$

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

חלופה אחרת לכלל הכרעה:

נדחה H_0 אם מתקיים:	
$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ או $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	

התפלגות הפרש הממוצעים: $\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{התקנון:}$$

דוגמה (פתרון בהקלטה):

בשנת 2004 הפער בין השכר הממוצע של הגברים לנשים היה 3000 ₪ לטובת הגברים. מעוניינים לבדוק האם כיום הצטמצם הפער בין הגברים לנשים מבחינת השכר הממוצע. נדגמו 100 עובדים גברים. שכרם הממוצע היה 9,072 ₪. נדגמו 80 עובדות, שכרן הממוצע היה 7809 ₪. לצורך פתרון נניח שסטיות התקן של השכר ידועות ושוות ל-2000 ₪ באוכלוסיית הנשים ו-3000 ₪ באוכלוסיית הגברים. מה המסקנה ברמת מבוהקות של 5%?

שאלות

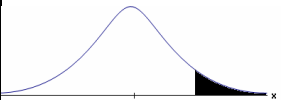
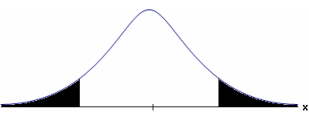
- (1) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלה נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות. לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.
- (2) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.
- (3) במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.
 - א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.
 - ב. מה תהיה המסקנה ברמת מובהקות של 8%?
- (4) במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע. מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו-40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים ו-12 ס"מ אצל הגברים.
 - א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?
 - ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

תשובות סופיות

- (1) נדחה H_0 .
- (2) לא נדחה את H_0 .
- (3) א. 0 ב. נדחה את H_0 .
- (4) א. נדחה את H_0 , אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס"מ.
ב. 0.6331

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונות של האוכלוסייה לא ידועות והמדגמים גדולים – רקע

השערת האפס: השערה אלטרנטיבית:	השערת האפס: השערה אלטרנטיבית:	השערת האפס: השערה אלטרנטיבית:	השערת האפס: השערה אלטרנטיבית:
$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$	תנאים: 1. מדגמים בלתי תלויים 2. σ_1, σ_2 לא ידועות 3. מדגמים מספיק גדולים
$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}$ דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0 :

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

חלופה אחרת לכלל הכרעה:

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

דוגמה (פתרון בהקלטה):

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

שאלות

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

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

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

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

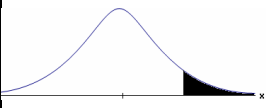
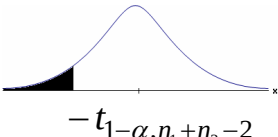
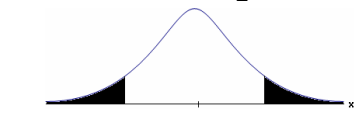
תשובות סופיות

(1) נדחה את H_0 .

(2) א. 0 ב. נדחה את H_0 .

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

השערת האפס: השערה אלטרנטיבית: תנאים:	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$
1. מדגמים בלתי תלויים 2. σ_1, σ_2 לא ידועות אך שוות 3. המשתנים בכל אוכלוסייה מתפלגים נורמלית	$t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ $t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ דוחים את H_0

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \quad \text{סטטיסטי המבחן:}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{השונות המשוקללת:}$$

חלופה אחרת לכלל הכרעה:

נדחה H_0 אם מתקיים:	
$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ או $\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	

דוגמה (פתרון בהקלטה):

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשתמשים בה כיום לבניית בניינים. לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות מהסוג החדש. להלן תוצאות המדגם:

טמפרטורת ההתכה הממוצעת במתכת הישנה 1170 מעלות עם אומד חסר הטיה לשונות $S^2 = 200$.

טמפרטורת ההתכה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות $S^2 = 260$.

נניח לצורך פתרון שטמפרטורת ההתכה מתפלגת נורמאלית עם אותה שונות במתכות השונות. בדקו ברמת מובהקות של 5%.

שאלות

- (1) להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר):

120	94	90	130	95	112	120	2012
	69	74	105	91	82	100	2013

בדקו שבשנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012 עבור רמת מובהקות של 5%.
הניחו ששטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שונות.

- (2) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

המדינה	ישראל	ארה"ב
גודל המדגם	15	15
סכום הציונים	1560	1470
סכום ריבועי הציונים	165,390	147,560

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים לאמריקאים מבחינת ממוצע הציונים במבחן ה-IQ לטובת ישראל.
רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

- (3) להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

הקבוצה	60W	100W
$\bar{X}$	1007	956
S	80	72
n	13	15

- א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.
ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג 100?
ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

תשובות סופיות

- (1) נדחה את H_0 .
- (2) הנחות:
1. סטיות התקן שוות.
 2. המשתנים מתפלגים נורמלית.
- נקבל את H_0 .
- (3) א. נדחה את H_0 . ב. רמת מובהקות של לפחות 5%.
- ג. לא נדחה את H_0 .

סטטיסטיקה

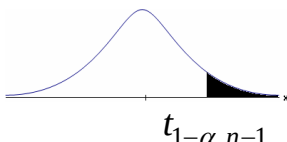
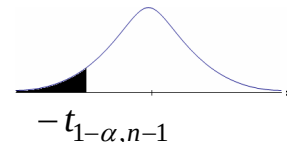
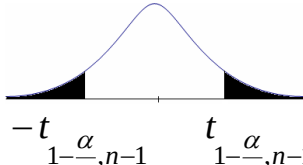
פרק 22 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. בדיקת השערות למדגמים מזווגים.....123

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

בדיקת השערות למדגמים מזווגים – רקע

$H_0: \mu_D = C$ $H_1: \mu_D > C$	$H_0: \mu_D = C$ $H_1: \mu_D < C$	$H_0: \mu_D = C$ $H_1: \mu_D \neq C$	השערת האפס: השערה אלטרנטיבית:
1. σ_D אינה ידועה 2. $D \sim N$ או מדגם מספיק גדול			תנאים:
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את H_0	$t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ ו $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}, \quad t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

סטטיסטי המבחן:

דוגמה (פתרון בהקלטה):

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחירה נמוכים מהמחירים מרשת השיווק "שופרסל". לצורך הבדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים:

המוצר / רשת	מגה בעיר	שופרסל
שמפו	17	18
ג'ל כביסה	48	57
עוגת גבינה	35	35
לחם	12	10
קפה נמס	49	47
בקבוק יין	113	142
גבינה בולגרית	20	26

בהנחה והמחירים מתפלגים נורמאלית, בדקו ברמת מובהקות של 5% את טענת רשת "מגה בעיר".

שאלות

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

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

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

- (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 מתפלגים נורמלית.

- (3) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים שלהם בסמסטר א' ו-ב':

סטטיסטיקה א'	74	68	90	75	82
סטטיסטיקה ב'	80	84	87	76	100

פורסם שתלמידים שמסיימים את סמסטר ב' משפרים במוצא את הציונים ב-5 נקודות לעומת סמסטר א'. הניחו שהציונים מתפלגים נורמלית.

- א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ-5 נקודות?
 ב. על סמך הסעיף הקודם, מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ-5 נקודות?
 ג. לאור זאת, מה המסקנה ברמת מובהקות של 10%?

- (4) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמלית ללא ידיעת השונות האמתית. המבחן שיש לבצע כאן הוא:

- א. מבחן Z למדגם יחיד.
 ב. מבחן T למדגם יחיד.
 ג. מבחן T למדגמים בלתי תלויים.
 ד. מבחן T למדגמים מזווגים.

- (5) בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

משקל במכשיר 1	2.5	0.7	9.6	4.5
משקל במכשיר 2	0.5	1.7	6.9	3.5

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.
- (6) כדי להשוות בין שני אצנים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים. המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.

תשובות סופיות

- לא נדחה H_0 .
- לא נדחה H_0 .
- א. $0.25 \leq p \leq 0.5$ ב. 0.5 ג. לא נדחה H_0 .
- ד'.
- ד'.
- ג'.

סטטיסטיקה

פרק 23 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות 127

הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α על $\mu_1 - \mu_2$:

$$H_0: \mu_1 - \mu_2 = C, \quad H_1: \mu_1 - \mu_2 \neq C$$

על ידי בניית רווח סמך ברמת סמך של $1 - \alpha$ ל- $\mu_1 - \mu_2$:

אם C נופל ברווח $\leftarrow$ נקבל את H_0 .

אם C לא נופל ברווח $\leftarrow$ נדחה את H_0 .

דוגמה (פתרון בהקלטה) :

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג.

$$H_0: \mu_D = 80, \quad H_1: \mu_D \neq 80, \quad \alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90%, $78 < \mu_D < 83$.

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

שאלות

(1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'.

להלן ציוניהם בסמסטר א' ו- ב' :

סמסטר א'	סמסטר ב'
74	80
68	84
90	87
75	76
82	100

א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.

ב. פורסם שתלמידים שמסיימים את סמסטר ב' משפרים בממוצע את הציונים ב-5 נק' לעומת סמסטר א'. האם יש אמת בפרסום?

(2) הוחלט להשוות הציונים אצל מרצה X ואצל מרצה Y. נבחרו באקראי 6

סטודנטים, 3 סטודנטים של מרצה X ו-3 סטודנטים של מרצה Y, עבורם התקבלו הציונים הבאים :

מרצה X	82	90	68
מרצה Y	68	81	64

א. חשבו רווח סמך ברמת סמך 90% להפרש בין התוחלות של הציונים אצל שני המרצים.

ב. האם ברמת מובהקות של 10% נכריע שיש הבדל בין תוחלות הציונים אצל שני המרצים?

שאלות רב-ברירה :

(3) סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.

הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$.

אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות :

$H_0 : \mu_1 - \mu_2 = 0$; $H_1 : \mu_1 - \mu_2 \neq 0$ ברמת מובהקות 0.05, מסקנתו תהיה :

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.

ד. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

- (4) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו : $-0.0293 < \mu_D < 0.2145$, רווח הסמך הוא ברמת סמך של 95%.
לכן מסקנת המחקר היא :
- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
 - ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
 - לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D .

תשובות סופיות

- (1) א. $-3.8 \leq \mu_D \leq 19$ ב. נכריע שיש אמת בפרסום.
- (2) א. $-8.5 \leq \mu_X - \mu_Y \leq 26.5$ ב. נכריע שאין הבדל.
- (3) ג'.
- (4) א'.

סטטיסטיקה

פרק 24 - רווח סמך לפרופורציה

תוכן העניינים

- 1. רווח הסמך לפרופורציה 130
- 2. קביעת גודל מדגם 133

רווח הסמך לפרופורציה:

רקע:

המטרה היא לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי:

$$\hat{p} = \frac{Y}{n} \quad (Y - \text{מספר ההצלחות שבמדגם}).$$

$$\text{רווח הסמך ל- } p : \hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

תנאי לבניית רווח הסמך:

מדגם של לפחות 30 תצפיות (לעיתים נותנים תנאי של מספר הצלחות ומספר כשלונות לפחות 5 או לפחות 10).

$$\text{האומד לטעות התקן: } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{מתקיים ש: } \hat{p} = \frac{A+B}{2}, \quad L = 2\varepsilon$$

דוגמה (פתרון בהקלטה):

במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים, מתוכם התקבל ש-24 היו מובטלים.

א. בנו רווח סמך לאחוז המובטלים באוכלוסייה ברמת סמך של 95%.

ב. מהו האומד לטעות התקן?

שאלות:

- (1) נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.
 א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.
 ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?
 ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95% למספר הדירות בחיפה עם ממ"ד בפועל.
- (2) במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.
 א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי הייטק).
 ב. כיצד רווח הסמך של סעיף א' היה משתנה אם היינו מקטינים את רמת הסמך?
 ג. כיצד רווח הסמך היה משתנה אם היינו מגדילים את גודל המדגם?
- (3) במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים:
 $0.08 < p < 0.18$
 א. כמה נהגים במדגם היו נהגים חדשים?
 ב. מהי רמת הסמך של רווח הסמך שנבנה?
- (4) במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו. 510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של $\pm 3\%$ מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?
- (5) במדגם של 300 נשים בגילאי 35-40 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.
 א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.
 ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?
- (6) ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95%. אורכו הינו 8.3156%. מהו גודל המדגם שנלקח?

תשובות סופיות:

- (1) א. $18.1\% < p < 29.9\%$
 ב. בביטחון של 95% שגיאת האמידה היא לכל היותר 0.059.
 ג. $14,480 < \mu < 23,920$
- (2) א. $0.545 \leq p \leq 0.655$ ב. האורך שלו היה קטן. ג. לא ניתן לדעת.
- (3) א. 52. ב. 0.997.
- (4) 0.925
- (5) א. $30.9\% > p > 22.5\%$ ב. $60.72\% > p > 45.91\%$
- (6) 200.

קביעת גודל מדגם:

רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת: החוקר קובע מראש את רמת הסמך הרצויה: $1-\alpha$. החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: ε (או את אורך רווח הסמך).

$L = 2\varepsilon$ - אורך רווח הסמך.

ε - טעות אמידה מרבית: המרחק המקסימאלי (הסטייה) בין הפרמטר (p) לאומד ($\hat{p}$).

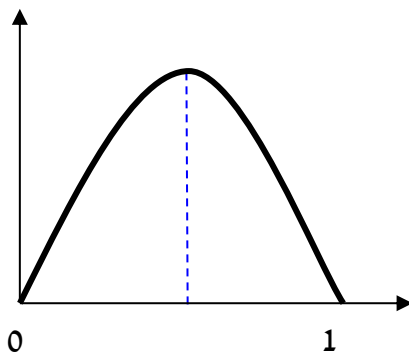
$$\varepsilon = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

$$n \geq \left(\frac{2 \cdot z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2 \quad \text{נקבל ש:}$$

הבעיה שאין אנו יודעים את $\hat{p}$.

נתבונן בביטוי: $\hat{p}(1-\hat{p})$.



כיוון שאין לנו ידע מוקדם על $\hat{p}$ נציב את המקרה השמרני ביותר שממקסם את הביטוי עבור: $\hat{p} = 0.5$.

$$n \geq \left(\frac{2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{0.5 \cdot 0.5}}{L} \right) \Rightarrow n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{L} \right)$$

אך אם תהיה לנו אינפורמציה מוקדמת על הפרופורציה נציב את הערך הקרוב ביותר ל-0.5 האפשרי.

דוגמה (פתרון בהקלטה):

מעוניינים לאמוד את שיעור האבטלה במשק. האמידה צריכה להתבצע ברמת סמך של 90% ועם שגיאת אמידה שלא תעלה על 4%.

א. מהו גודל המדגם המינימאלי שיש לקחת?

ב. חזור לסעיף א' אם ידוע שהאבטלה לא אמורה לעלות על 20%.

שאלות:

- (1) הממשלה אומדת מדי חודש את אחוז התמיכה בה. מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמיתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?
- (2) משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.
 - א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?
 - ב. חזרו על סעיף א' אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.
- (3) ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים. המטרה שבביטחון של 95% הסטייה המרבית בין האומד לרייטינג האמיתי לא תעלה על 4%.
 - א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?
 - ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?
- (4) ענו על הסעיפים הבאים:
 - א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?
 - ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.
 - ג. על סמך סעיף ב', האם תקבלו את הטענה שמיעוט האוכלוסייה תומך הממשלה?
- (5) משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.
 - א. כמה מחוסנים יש לדגום?
 - ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבתם בסעיף הקודם וקיבל ש-15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.
 - ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98% מדוע הוא קטן מ-3%?

תשובות סופיות:

- (1) .1068
- (2) א. .423 ב. .271
- (3) א. .601 ב. 108,000 ₪.
- (4) א. .335 ב. $0.367 < p < 0.473$.
- ג. בביטחון של 0.95 ניתן להגיד שמיעוט באוכלוסייה תומך בממשלה.
- (5) א. .1509 ב. 0.15 ± 0.02 ג. ראה סרטון.

סטטיסטיקה

פרק 25 - רווח סמך להפרש פרופורציות

תוכן העניינים

1. רווח סמך להפרש פרופורציות 136

רווח סמך להפרש פרופורציות:

רקע:

המטרה: לאמוד את $p_1 - p_2$: הפרש פרופורציות בין שתי אוכלוסיות שונות.

האומד הנקודתי: $\hat{p}_1 - \hat{p}_2$.

התנאי לבניית רווח הסמך: כל מדגם מעל 30 או לבדוק שמספר ההצלחות ומספר הכישלונות בכל מדגם לפחות 5 בכל מדגם (יש כאלה שבדקים לפחות 10).

$$\text{רווח סמך: } (\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

רק שאפס נופל בתחומי רווח הסמך להפרש הפרופורציה נאמר שלא ניתן לקבוע שקיים הבדל מובהק בין הפרופורציות באוכלוסיות.

דוגמה (פתרון בהקלטה):

במטרה להשוות בין שתי תרופות נדגמו 200 איש שלקחו תרופה X , מתוכם 180 טענו שהתרופה עזרה להם. כמו כן, נלקחו 300 איש שלקחו את תרופה Y . מתוכם 150 טענו שהתרופה עזרה להם. בנו רווח סמך להפרש אחוזי ההצלחה של התרופות ברמת סמך של 95%. מה ניתן לומר על סמך רווח הסמך על ההבדלים בין התרופות?

שאלות:

- (1) מתוך 150 נשים שנדגמו באקראי 30% תמכו בהצעת חוק מסוימת. מתוך 200 גברים שנדגמו באקראי 25% תמכו בהצעת החוק.
א. בנו רווח סמך לפער בין אחוזי התמיכה של הנשים לעומת הגברים ברמת סמך של 96%.
ב. בנו רווח סמך ברמת סמך של 95% לאחוז התמיכה בהצעת החוק.
- (2) במחקר רפואי השתתפו 200 אנשים הסובלים מכאבים כרוניים. הם חולקו באקראי ל-2 קבוצות שוות בגודלן. קבוצה 1 קיבלה את תרופה A וקבוצה שנייה קיבלה את תרופה B. בקרב לוקחי תרופה A 90 טענו שמצבם השתפר. בקרב לוקחי תרופה B 70 טענו שמצבם השתפר.
א. בנו רווח סמך ברמת סמך של 95% להפרש בין שיעורי ההצלחה של שתי התרופות.
ב. האם על סמך סעיף א' ניתן לקבוע שקיים הבדל בין התרופות מבחינת שיעורי ההצלחה?
- (3) נדגמו 200 משפחות מגוש דן. ל-70% מתוכן מכשיר DVD בבית. נדגמו 300 משפחות מאזור הצפון ל-65% מתוכן מכשיר DVD בבית.
א. בנו רווח סמך ברמת סמך של 98% לפרופורציות המשפחות בגוש דן עם DVD בבית.
ב. בנו רווח סמך ברמת סמך של 95% להפרש בין פרופורציות המשפחות בגוש דן עם DVD לבין פרופורציות המשפחות בצפון עם DVD.

תשובות סופיות:

- (1) א. $-4.9\% < P_F - P_M < 14.9\%$. ב. $22.5\% < p < 31.8\%$.
- (2) א. $0.093 < P_A - P_B < 0.307$. ב. כן.
- (3) $0.625 < p < 0.7754$

סטטיסטיקה

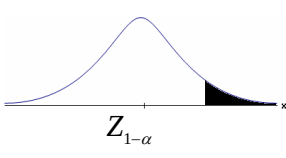
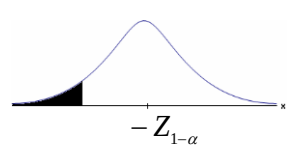
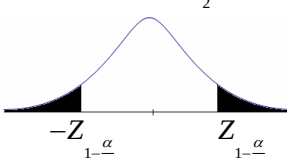
פרק 26 - בדיקת השערות על פרופורציה

תוכן העניינים

138	1. התהליך
141	2. סיכוי לטעויות ועוצמה
145	3. מובהקות התוצאה - אלפא מינימלית

התהליך:

רקע:

השערת האפס: השערה אלטרנטיבית:	$H_0: p = p_0$ $H_1: p > p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p \neq p_0$
תנאים:	$np_0 \geq 5 \text{ \& } n(1-p_0) \geq 5$		
כלל ההכרעה: אזור הדחייה של H_0 :	$Z_{\hat{p}} > Z_{1-\alpha}$  -דוחים את H_0	$Z_{\hat{p}} < -Z_{1-\alpha}$  -דוחים את H_0	או $Z_{\hat{p}} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  -דוחים את H_0

סטטיסטי המבחן:
$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

חלופה אחרת לכלל הכרעה:

כלל ההכרעה – אזור הדחייה של H_0 :		
$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ $\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$

דוגמה (פתרון בהקלטה):

בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים.
בדקו ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

שאלות:

- (1) במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. ברמת מובהקות של 5% האם השנה הקשו על תנאי הקבלה?
- (2) במדגם של 300 אזרחים 57% מתנגדים להצעת חוק מסוימת. לאור נתונים אלה האם רוב האזרחים מתנגדים להצעת החוק? בדקו ברמת מובהקות של 10%.
- (3) הטילו מטבע 50 פעמים וקיבלו 28 פעמים עץ. האם המטבע הוגן ברמת מובהקות של 5%?
- (4) קפיטריה במכללה מסוימת מעריכה כי אחוז הסטודנטים שקונים קפה בקפיטריה הינו 20%. נערך סקר אשר כלל 200 סטודנטים. התברר כי 33 מהם רוכשים קפה בקפיטריה. מטרת הסקר הייתה לבדוק את אמיתות הערכה של הקפיטריה.
 - א. רשמו את ההשערות.
 - ב. בדקו את ההשערות ברמת מובהקות של 10%.
 - ג. מה תהיה המסקנה אם נקטין את רמת המובהקות?
- (5) חבר כנסת רוצה להעביר חוק. לצורך כך הוא דוגם 400 אזרחים במטרה לבדוק האם רוב האזרחים תומכים בחוק. במדגם התקבל ש-276 אזרחים תומכים בחוק.
 - א. מה מסקנתכם ברמת מובהקות של 5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה אם רמת המובהקות תהיה גדולה יותר? הסבירו.
- (6) שני חוקרים בדקו את ההשערות הבאות: $H_0: p = p_0$, $H_1: p > p_0$. חוקר א' השתמש ברמת מובהקות α_1 וחוקר ב' ברמת מובהקות α_2 החוקר הראשון דחה את H_0 ואילו החוקר השני קיבל את H_0 . שניהם התבססו על אותם תוצאות של מדגם. בחר בתשובה הנכונה:
 - א. $\alpha_1 = \alpha_2$.
 - ב. $\alpha_1 > \alpha_2$.
 - ג. $\alpha_1 < \alpha_2$.
 - ד. המצב המתואר לא אפשרי.

תשובות סופיות:

- (1) נדחה H_0 .
- (2) נדחה H_0 .
- (3) נקבל H_0 .
- (4) א. $H_0: p = 0.2$
 ב. $H_1: p \neq 0.2$
 ג. המסקנה לא תשתנה.
- (5) א. נדחה H_0 .
 ב. המסקנה לא תשתנה.
- (6) ג'.

סיכוי לטעויות ועוצמה:

רקע:

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$$\alpha = P(H_0 \text{ לדחות את } H_0 \mid H_0 \text{ נכונה}) = P_{H_0}(H_0 \text{ לדחות})$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל את } H_1 \mid H_1 \text{ נכונה}) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון:

$$(1-\alpha) = P(H_0 \text{ לקבל את } H_0 \mid H_0 \text{ נכונה}) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה:

$$\pi = (1-\beta) = P(H_0 \text{ לדחות את } H_1 \mid H_1 \text{ נכונה}) = P_{H_1}(H_0 \text{ לדחות})$$

מציאות	הכרעה		
		H_0	H_1
	H_0	טעות מסוג 1	אין טעות
	H_1	טעות מסוג 2	אין טעות

התהליך לחישוב סיכוי לטעות מסוג שני:

השערת האפס: השערה אלטרנטיבית:	$H_0: p = p_0$ $H_1: p \neq p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$
תנאים:	$np_0 \geq 5 \& n(1-p_0) \geq 5$		
כלל ההכרעה: אזור הדחייה של H_0 :	$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ או $\hat{p} < p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$

חישוב β :
$P_{H_1} \left(\hat{p} < p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \right)$
$P_{H_1} \left(p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \right)$
$P_{H_1} \left(\hat{p} > p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \right)$

כאשר: $\hat{p} \sim N \left(p, \frac{p(1-p)}{n} \right)$

והתקנון: $Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

דוגמה (פתרון בהקלטה):

רופאי שיניים טוענים שיותר ממחצית האוכלוסייה הבוגרת בארץ אינם מבקרים אצל רופא שיניים באופן קבוע, כנדרש. כדי לבדוק טענה זו, נערך סקר בקרב 150 אנשים בוגרים.

- רשמו את ההשערות וכלל הכרעה ברמת מובהקות של 10%.
- מהי עוצמת המבחן אם מסתבר ש 60% מהאוכלוסייה אינם מבקרים אצל רופא שיניים באופן קבוע.

שאלות:

- (1) משרד הבריאות פרסם ש-10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.
 - א. רשמו את השערות המחקר, וצרו מבחן ברמת מובהקות של 5% לבדיקתן.
 - ב. מהי עצמת המבחן של סעיף א' בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?
 - ג. כיצד תשנה התשובה לסעיף ב' אם מסתבר שבחיפה 18% סובלים מאסטמה?
 - ד. בהמשך לסעיף א' האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?
- (2) אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
 - נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10%, מהי עצמת המבחן עבור רמת מובהקות של 5%?
- (3) בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.
 - א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.
 - ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?
- (4) מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת. חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?
- (5) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן (בחרו בתשובה הנכונה):
 - א. השערת האפס נכונה.
 - ב. השערת האפס נדחתה.
 - ג. השערת האפס לא נדחתה.
 - ד. אף אחת מהתשובות לא נכונה בהכרח.
- (6) קבעו אם הטענה הבאה נכונה: בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני.

תשובות סופיות:

- (1) א. $H_0 : p = 0.1$
 $H_1 : p > 0.1$
 ב. 0.9015 ג. תגדל. ד. טענה לא נכונה.
- (2) 0.4404
- (3) א. 0.1446 ב. תקטן.
- (4) חוקר א'.
- (5) ג'.
- (6) נכונה.

מובהקות התוצאה – אלפא מינימלית:

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה: באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות. המסקנה של המחקר תקבע לפי העיקרון הבא:

אם $p_v \leq \alpha$ דוחים את H_0 .

מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

לקבל את תוצאות המדגם וקיצוני $p_v = P_{H_0}$.

אם ההשערה היא דו צדדית:

לקבל את תוצאות המדגם וקיצוני $p_v = 2P_{H_0}$.

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: השערה אלטרנטיבית:	$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$
תנאים:	$np_0 \geq 5 \& n(1-p_0) \geq 5$		
p-value	$P_{H_0}(\hat{p} \geq \hat{p})$	$P_{H_0}(\hat{p} \leq \hat{p})$	אם $2 \cdot P_{H_0}(\hat{p} \geq \hat{p}) \Leftarrow \hat{p} > p_0$ אם $2 \cdot P_{H_0}(\hat{p} \leq \hat{p}) \Leftarrow \hat{p} < p_0$

כאשר בהנחת השערת האפס: $\hat{P} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

התקנון: $Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

דוגמה (פתרון בהקלטה):

ישנה טענה שיש הבדל בין אחוז הבנים ואחוז הבנות הפונים ללמוד להנדסאי מחשבים. לשם כך נלקח מדגם מקרי של 200 תלמידים הלומדים מחשבים והתברר כי 112 מהם בנים.

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

שאלות:

- (1) במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. רוצים לבדוק האם השנה הקשו על תנאי הקבלה.
א. מהי מובהקות התוצאה?
ב. מה תהיה המסקנה ברמת מובהקות של 1% וברמת מובהקות של 5%?
- (2) נהוג לחשוב ש-60% מהילדים בגיל שלוש קמים מהמיטה במהלך הלילה לפחות פעם אחת. ישנה טענה שללא שנת צהריים פחות מ-60% מהילדים בגיל זה יקומו לפחות פעם אחת במהלך הלילה. נדגמו 80 ילדים בגיל 3 אשר אינם ישנים בצהריים מתוכם התקבל ש-41 קמו במהלך הלילה.
א. מהי רמת המובהקות המינימלית עבורה תתקבל הטענה במחקר?
ב. מהי רמת המובהקות המקסימלית עבורה לא תתקבל טענת המחקר?
ג. עבור אילו רמות מובהקות נקבל את טענת המחקר?
ד. מה תהיה מסקנת המחקר ברמת מובהקות של 6%?
- (3) במטרה לבדוק האם מטבע הוא הוגן מטילים אותו 80 פעמים. התקבל ש-60 מההטלות הראו עץ. רשמו את השערות המחקר, חשבו את מובהקות התוצאה והסיקו מסקנה ברמת מובהקות של 5%.
- (4) בבדיקת השערות על פרופורציה התקבל שה- $p\text{-value} = 0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 5%:
(בחרו בתשובה הנכונה)
א. יקבל את השערת האפס
ב. ידחה את השערת האפס.
ג. לא ניתן לדעת כי אין מספיק נתונים.
- (5) קבעו אם הטענה הבאה נכונה:
"במבחן לבדיקת השערות חד-צדדי התקבל ערך $p\text{-value}$ של 3%,
לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי),
היינו מקבלים ערך $p\text{-value}$ של 6%".
- (6) במפעל 10% מהעובדים נפגעים לפחות פעם אחת בשנה מתאונות עבודה. לאור זאת, המפעל החליט לצאת בתוכנית לצמצום שיעור הנפגעים. תכנית זו נוסתה על 100 עובדים. מתוכם 12 נפגעו בתאונות עבודה במשך השנה. מהי רמת המובהקות הקטנה ביותר עבורה יוחלט שהתכנית יעילה?

תשובות סופיות:

- (1) א. 0.0455.
 ב. ברמת מובהקות של 1% : לא דוחים את H_0 .
 ברמת מובהקות של 5% : נדחה את H_0 .
- (2) א. 0.0548. ב. 0.0548. ג. מעל 0.0548.
 ד. נכריע לטובת טענת המחקר.
- (3) $p_v = 0$, נדחה את H_0 .
- (4) ב'.
- (5) הטענה נכונה.
- (6) 0.7486.

סטטיסטיקה

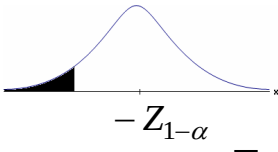
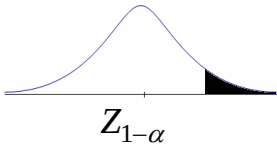
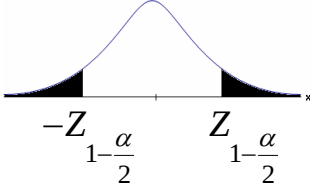
פרק 27 - בדיקת השערות על הפרש פרופורציות

תוכן העניינים

1. כללי 149

בדיקת השערות על הפרש פרופורציות

רקע

השערת האפס : השערה אלטרנטיבית:	1. מדגמים בלתי תלויים	2. מדגמים גדולים	
$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 > 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 < 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 \neq 0$	
תנאים:	1. מדגמים בלתי תלויים	2. מדגמים גדולים	כלל ההכרעה: אזור הדחייה של:
	$Z_{\hat{p}_1 - \hat{p}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ - דוחים את H_0	$Z_{\hat{p}_1 - \hat{p}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ - דוחים את H_0	
	$Z_{\hat{p}_1 - \hat{p}_2} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\hat{p}_1 - \hat{p}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ - דוחים את H_0		

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} \quad \text{סטטיסטי המבחן:}$$

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} \quad \text{כאשר הפרופורציה המשוקללת:}$$

חלופה אחרת לכלל הכרעה:

כלל ההכרעה: אזור הדחייה של H_0	
$\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$ או $\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$
$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	

$$\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}) \quad : \hat{p}_1 - \hat{p}_2 \text{ התפלגות של}$$

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

תקנון:

$$Z_{\hat{p}_1 - \hat{p}_2} \stackrel{H_0}{=} \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

דוגמה (פתרון בהקלטה):

נדגמו 80 סטודנטים שנבחנו במיקרו-כלכלה. מתוכם 60 עברו את הבחינה.
 נדגמו 100 סטודנטים שנבחנו בסטטיסטיקה א'. מתוכם 82 עברו את הבחינה.
 האם שיעור העוברים את הבחינה בסטטיסטיקה גבוה מאשר מהבחינה במיקרו כלכלה? בדקו ברמת מבוהקות של 10%.

שאלות

- (1) במדגם של 200 גברים, 8% היו מובטלים. במדגם של 180 נשים, 10% מהן היו מובטלות.
האם קיים הבדל מובהק בין פרופורציית המובטלים לפרופורציית המובטלות?
בדקו ברמת מובהקות של 5%.
- (2) אחוז בעלי רישיון נהיגה בקרב האוכלוסייה הבוגרת הינו 60%. במדגם של 300 בוגרים מתל אביב 204 היו בעלי רישיון נהיגה.
במדגם של 220 בוגרים מירושלים 100 היו בעלי רישיון נהיגה.
א. ברמת מובהקות של 5% האם תקבלו את הטענה שאחוז בעלי הרישיון בתל אביב גבוה מהאחוז הארצי?
ב. ברמת מובהקות של 10% האם תקבלו את הטענה שאחוז בעלי הרישיון נהיגה בתל אביב גבוה מאחוז בעלי רישיון הנהיגה בירושלים?
- (3) נדגמו 500 בוגרים מתוכם 200 גברים והיתר נשים. במדגם התקבל: מתוך הגברים ל-48% תעודת בגרות. מתוך הנשים ל-58% תעודת בגרות. מטרת המחקר היא לבדוק האם שיעור הזכאיות לבגרות גבוה משיעור הזכאים.
א. מהי מובהקות התוצאה?
ב. מה תהיה המסקנה ברמת מובהקות של 8%?
- (4) במדגם שנערך על 100 פרות מחוות בדרום הארץ התקבל כי 20 פרות נושאות וירוס מסוים. במדגם שנערך על 200 פרות מחוות בצפון הארץ התקבל כי 10 מתוכן נושאות וירוס גם כן.
א. בנו מבחן ברמת מובהקות של 5% לבדיקת הטענה כי הווירוס תקף את פרות הדרום באופן משמעותי יותר מאשר את הפרות בצפון הארץ.
ב. מהי המסקנה לבדיקת הטענה של סעיף א ומהי הטעות האפשרית במסקנה?
ג. מהי עוצמת המבחן אם שיעור הפרות בדרום עם הווירוס גבוה ב-10% משיעור הפרות בצפון עם הווירוס?
ד. כיצד העוצמה תשתנה אם נגדיל את רמת המובהקות?

תשובות סופיות

- | | |
|-----|---|
| (1) | לא נדחה את H_0 . |
| (2) | א. נדחה H_0 .
ב. נדחה H_0 . |
| (3) | א. 0.0139
ב. נדחה H_0 . |
| (4) | א. ראה סרטון.
ב. נדחה H_0 .
ג. 0.8238
ד. תגדל. |

סטטיסטיקה

פרק 28 - נמבחן הבינום

תוכן העניינים

1. מבחן הבינום 153
2. חישוב רמת מובהקות ועוצמת מבחן במבחן הבינום 158

מבחנים אפרמטריים למדגם יחיד

מבחן הבינום – רקע

מבחן הבינום הינו מבחן סטטיסטי על הפרמטר p . הפרמטר p מייצג את פרופורציית ההצלחות באוכלוסייה כלומר, הסיכוי בניסוי בודד להצלחה. ההשערות האפשריות על הפרמטר הן:

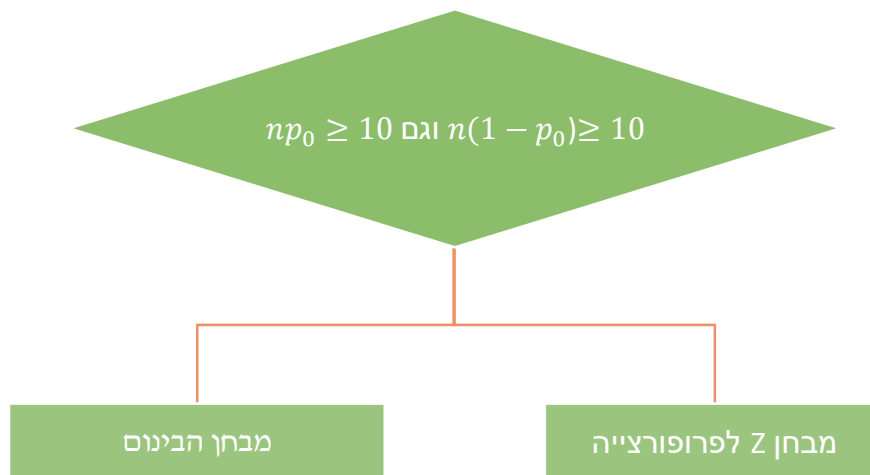
$H_0 : p = p_0$	$H_0 : p = p_0$	$H_0 : p = p_0$	השערת האפס:
$H_1 : p > p_0$	$H_1 : p < p_0$	$H_1 : p \neq p_0$	השערה אלטרנטיבית:

דוגמה:

במדינה אירופאית התנהל לפני 4 שנים משאל עם בדבר לגליזציה של הקנאביס. במשאל העם 40% מהאזרחים היו בעד לגליזציה של הקנאביס ובשל כך חוק הלגליזציה לא עבר באותה המדינה. במדגם שנעשה כיום בו השתתפו 15 אזרחים מהמדינה 10 ענו שהם תומכים בלגליזציה של הקנאביס במדינה. פרלמנטר מהמדינה חוקר האם כיום אחוז התומכים בלגליזציה של הקנאביס במדינה עלה יחסית לאחוז שהתקבל במשאל העם במדינה לפני 4 שנים. רשמו את השערות המחקר.

במבחן הבינום מבצעים מדגם אקראי בגודל n ומתבוננים במספר ההצלחות שהתקבלו במדגם, אותן נסמן ב- Y . בהנחה והתצפיות במדגם בלתי תלויות זו בזו אנו אומרים ש: $Y \sim B(n, p)$.

מבחן הבינום נכנס לקטגוריה של מבחנים אפרמטרים והוא בא כחלופה למבחן הפרמטרי על פרופורציה אחת כאשר התנאים לקירוב הנורמלי אינם מתקיימים.



דוגמה:

מהו המבחן הסטטיסטי המתאים? נמקו.

הטכניקה הנוחה ביותר למבחן הבינום היא לחשב את מובהקות התוצאה ולדחות את השערת האפס אם $\alpha \geq PV$. מובהקות התוצאה היא הסיכוי לתוצאות של המדגם וקיצוני יותר בהנחת השערת האפס. כזכור, פונקציית ההסתברות של ההתפלגות הבינומית היא:

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

כאשר התוחלת של ההתפלגות הבינומית היא: $E(Y) = np$. בעמוד הבא מצורפת טבלה של התפלגות בינומית מצטברת. זו טבלה שיכולה לעזור בתהליך החישוב.

דוגמה:

חשבו את מובהקות התוצאה. מה תהיה מסקנת המחקר ברמת מובהקות של 5%?

טבלת התפלגות בינומית מצטברת

n	x	p									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

שאלות

- (1) אחוז האנשים שאוכלים גלילות בחודשי החורף הינו 30%, קיים חשש בקרב מיצרי הגלילות כי השנה פחת אחוז אוכלי הגלילות בחודשי החורף. לשם כך נדגמו 12 אנשים אשר מתוכם 2 טענו שהם אוכלים גלילה בחודשי החורף.
- א. רשמו את השערות המוצגות בשאלה זו וציינו מהו המבחן הסטטיסטי המתאים.
- ב. מה תהיה המסקנה ברמת מובהקות של 10%?
- ג. מה תהיה המסקנה ברמת מובהקות גדולה יותר מ- 10%?
- (2) מטבע הוטל 15 פעמים במטרה לבדוק האם המטבע סימטרי.
- בסך הכול התקבלו בהטלות 3 פעמים עץ.
- א. בדקו ברמת מובהקות של 5% האם המטבע הוא סימטרי.
- ב. עבור אילו רמות מובהקות נוכל להסיק שהמטבע אינו סימטרי?
- (3) אחוז המובטלים במשק לפני 5 שנים היה 10%. מעוניינים לבדוק האם כיום אחוז המובטלים קטן לעומת זה שהיה לפני 5 שנים. במדגם של 20 אנשים התקבל מובטל אחד ויחיד.
- א. מהי רמת המובהקות הקטנה ביותר עבורה יוסק שכיום אחוז האבטלה נמוך מאשר לפני 5 שנים?
- ב. מה תהיה המסקנה אם רמת המובהקות תהיה 5%?
- (4) נניח ש 40% מהאוכלוסייה מכירה את המוצרים של חברת "רמקס". החברה מתכננת לצאת בקמפיין פרסום שמטרתו להעלות את המודעות לקיומם של המוצרים של החברה באוכלוסייה. בזמן הקמפיין נדגמו 20 אנשים אקראיים מתוכם 70% טענו שהם מכירים את המוצרים של חברת "רמקס".
- א. מהי מובהקות התוצאה?
- ב. כיצד מובהקות התוצאה הייתה משתנה אם במדגם 80% היו טוענים שהם מכירים את המוצרים של חברת "רמקס"?

- (5) חברה לתוספי מזון דיווחה שנטילת מולטי ויטמין מקטינה את הסיכוי לחלות במחלות חורף במהלך החורף. לפי משרד הבריאות 70% מהאוכלוסייה חולים במחלות חורף במהלך החורף. במחקר השתתפו 20 אנשים אשר נטלו במהלך שנה מולטי ויטמין. במהלך החורף נמצא ש 12 מתוכם חלו במחלות חורף במהלך החורף. את המחקר יש לבצע ברמת מובהקות של 5%. נסמן ב: PV את מובהקות התוצאה של המחקר.
- א. האם המשפט הבא נכון?
 "אם כלל האוכלוסייה הייתה נוטלת מולטי ויטמין אז בסיכוי של 5% ניתן להגיד ש-70% מהאוכלוסייה הייתה חולה במחלת חורף".
- ב. האם המשפט הבא נכון?
 "אם כלל האוכלוסייה הייתה נוטלת מולטי ויטמין אז בסיכוי של PV 70% מהאוכלוסייה הייתה חולה במחלת חורף".
- ג. חשבו את PV.

תשובות סופיות

- (1) א. מבחן הבינום.
 $H_0 : p = 0.3$
 $H_1 : p < 0.3$
- (2) א. נדחה את H_0 .
 ב. לפחות 0.0352
- (3) א. 0.3917
 ב. לא נדחה את H_0 .
- (4) א. 0.0065
 ב. תקטן.
- (5) א. לא נכון.
 ב. לא נכון.
 ג. 0.2277.

חישוב הסתברויות לטעויות ועוצמת המבחן במבחן הבינום – רקע

מבחן הבינום הינו מבחן סטטיסטי על הפרמטר p . הפרמטר p מייצג את פרופורציית ההצלחות באוכלוסייה כלומר, הסיכוי בניסוי בודד להצלחה. ההשערות האפשריות על הפרמטר הן:

$H_0 : p = p_0$	$H_0 : p = p_0$	$H_0 : p = p_0$	השערת האפס:
$H_1 : p > p_0$	$H_1 : p < p_0$	$H_1 : p \neq p_0$	השערה אלטרנטיבית:

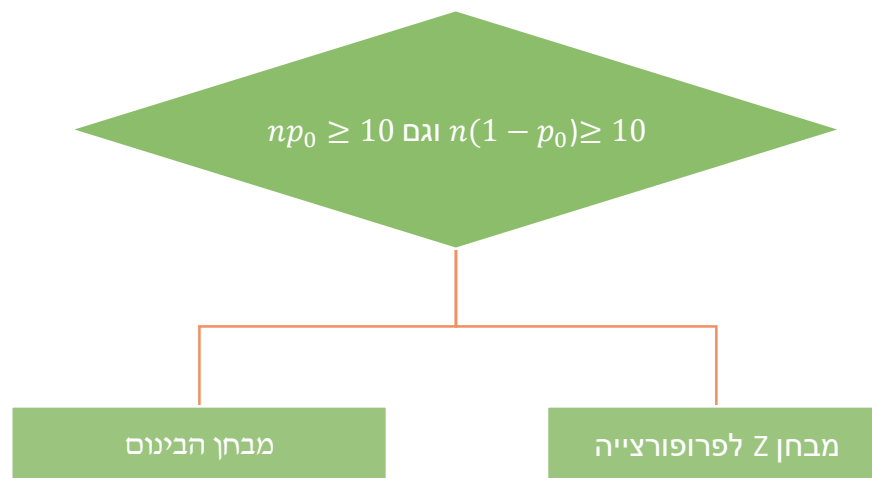
דוגמה:

בחברת פז 40% מעובדי החברה מאחרים לעבודה. מנכ"ל החברה מעוניין לצמצם את אחוז העובדים המאחרים בחברה. לצורך כך, בוחר המנכ"ל מחלקה אקראית במטרה לבדוק עליה שיטת תמריצים שאמורה לצמצם את אחוז המאחרים. במחלקה שנבחרה 10 עובדים והופעלה עליה שיטת התמריצים הנבדקת. הוסכם שאם מספר המאחרים במחלקה יהיה לכל היותר עובד אחד לאחר הפעלת שיטת התמריצים, תופעל שיטת התמריצים על כלל החברה.

• מהן השערות המחקר?

במבחן הבינום מבצעים מדגם אקראי בגודל n ומתבוננים במספר ההצלחות שהתקבלו במדגם, אותן נסמן ב- Y . בהנחה והתצפיות במדגם בלתי תלויות זו בזו אנו אומרים ש: $Y \sim B(n, p)$.

מבחן הבינום נכנס לקטגוריה של מבחנים אפרמטריים והוא בא כחלופה למבחן הפרמטרי על פרופורציה אחת כאשר התנאים לקירוב הנורמלי אינם מתקיימים.



דוגמה:

- מהו המבחן הסטטיסטי המתאים? נמקו.

כזכור, פונקציית ההסתברות של ההתפלגות הבינומית היא:

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

בעמוד הבא מצורפת טבלה של התפלגות בינומית מצטברת. זו טבלה שיכולה לעזור בתהליך החישוב.

נזכיר את המצבים האפשריים בבדיקת השערות:

מציאות	הכרעה		
		H0	H1
	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

$$\alpha = P(H_0 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ לדחות})$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל} \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון:

$$1-\alpha = P(H_0 \text{ לקבל} \mid \text{לדחות את } H_0) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה:

$$\pi = 1-\beta = P(H_1 \text{ נכונה} \mid \text{לדחות את } H_0) = P_{H_1}(H_0 \text{ לדחות})$$

דוגמה:

- מהי רמת המובהקות של המחקר המוצע?
- אם בשיטת התמריצים, הסיכוי לאחר הוא 10%, מה ההסתברות שהמחקר יגלה זאת?

טבלת הסתברות בינומית (מצטברת) עבור n - ים שונים

n	x	p									
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

שאלות

- (1) במטרה לבדוק האם במוסד לימוד מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 תלמידים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.
- א. מהי רמת המובהקות במבחן המוצע?
 - ב. מהי העוצמה במבחן המוצע בהנחה ובמוסד 30% בנים?
 - ג. כיצד התשובות לסעיפים הקודמים ישתנו אם ישנו את המבחן כך שיוחלט שאם מספר הבנים במדגם יהיה לכל היותר 3 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות? הסבירו ללא חישוב מחדש.
- (2) הסיכוי לזכות במשחק מסוים הינו 70%. מעוניינים לבדוק האם שיטת משחק מסוימת מעלה את סיכויי הזכייה לצורך כך מחליטים לשחק את המשחק 20 פעמים תוך שימוש בשיטה הנבדקת. מחליטים שאם מספר הזכיות במדגם 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 במבחן שהוצע?
- (3) במטרה לבדוק האם מטבע הוא סימטרי 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 להחליט שהמטבע אינו סימטרי למרות שבפועל המטבע הינו סימטרי?
 - ג. מהי ההסתברות שאם במטבע הסיכוי לעץ הוא 20%, אכן נגלה זאת?
 - ד. כיצד התשובה לסעיף הקודם תשתנה אם הסיכוי לעץ במטבע אף נמוך מ-20%?
- (4) נתון ש- $X \sim B(15, p)$ ההשערות של המחקר הן: $H_0: p = 0.3$ $H_1: p > 0.3$.
- כלל ההכרעה הוא: נדחה את השערת האפס אם $X \geq 8$.
- א. מהי רמת המובהקות במחקר זה?
 - ב. מה יקרה לרמת המובהקות אם כלל ההכרעה יהיה: נדחה את השערת האפס אם $X \geq 7$? הסבירו ללא חישוב.
 - ג. מה יקרה לרמת המובהקות אם השערות המחקר יהיו $H_0: p = 0.3$ $H_1: p \neq 0.3$ אבל כלל ההכרעה לא ישתנה? הסבירו ללא חישוב.

תשובות סופיות

- (1) א. 0.0547 ב. 0.3828 ג. תגדלנה.
- (2) $\beta = 0.3231$ $\alpha = 0.0355$
- (3) א. $H_0: p = 0.5$
 $H_1: p \neq 0.5$ ב. 0.0078 ג. 0.1678
- ד. העוצמה תגדל.
- (4) א. 0.05 ב. תגדל. ג. לא תשתנה.

סטטיסטיקה

פרק 29 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות פתוחות מסכמות.....163
2. שאלות רב ברירה (אמריקאיות).....166

שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות

- (1) שני חוקרים נתבקשו לבדוק את ההשערות הבאות: $H_0: \mu = 520$, $H_1: \mu > 520$. כל חוקר בדק מדגם של 225 נחקרים. ידוע ש- $\sigma = 20$. חוקר א' קבע את כלל ההכרעה לפי $\alpha = 0.05$. חוקר ב' מחליט לדחות H_0 אם $\bar{X} > 522$.

- למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?
- מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור $\mu_1 = 525$?
- הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור $\mu_1 = 525$, של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'?
- חוקר א' קיבל במדגם שלו $\bar{X} = 523$. מהי מסקנתו?

- (2) ידוע כי תוחלת מספר הלייקים היומי של דנה היא 12 עם סטיית תקן 5. דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר לייקים מדנה ביום. על-מנת לבדוק זאת ספר דני כמה לייקים הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב-49 ימים) וקיבל סך-הכול 637 לייקים. נניח כי סטיית התקן של מספר הלייקים שדני מקבל ביום זהה לסטיית התקן של דנה.
- מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולרי יותר בכך שהוא מקבל יותר לייקים מדנה ביום)?
 - אם דני משער שתוחלת מספר ה"לייקים" שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

מוצר / רשת	A	B
1	5	5
2	4	5
3	5	3
4	7	4

- (3) ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:
- הניחו כי המחירים מתפלגים נורמלית.
- אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.
- בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.
 - חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

- (4) במטרה לבדוק האם סטודנטים הלומדים במכללות משקיעים יותר זמן ללימודים מאשר סטודנטים באוניברסיטאות נדגמו 12 סטודנטים ובדקו לכל סטודנט את הזמן שהוא משקיע ביום ללימודים. הזמנים נמדדו בדקות:

סטודנטים באוניברסיטאות	176	156	189	171	140	180
סטודנטים במכללות	180	190	191	186	204	150

- א. נסחו את ההשערות ובדוק אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.
- ב. חשבו את p-value.
- ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדוק את הטענה כאשר רמת המובהקות הינה 5%.

- (5) במטרה להשוות בין אחוזי הצפייה של גברים ונשים בתוכנית טלוויזיה מסוימת בוצע סקר ובו התקבלו תוצאות הבאות:

צופים	לא צופים	
נשים	320	42
גברים	72	120

- א. האם יש הבדל בין אחוזי הצפייה של גברים ונשים ברמת מובהקות של 1%?
- ב. עבור רמת מובהקות של 5% בדוק טענה שמבין הצופים בתוכנית הטלוויזיה אחוז הנשים גדול פי 2 מאחוז הגברים.

- (6) בשנת 2000 ל-60% היה מדיח כלים בבית. מחקר רוצה לבדוק האם כיום פרופורציית המשפחות עם מדיח כלים עלה. הוחלט לבצע מדגם אקראי של 150 משפחות.

- א. רשמו את השערות המחקר.
- ב. מה היא מסקנת המחקר ברמת מובהקות של 5% אם במדגם ל-102 משפחות היה מדיח כלים.
- ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם. האם ניתן לדעת את הסתברותה?

- (7) נערך מחקר על הקשר בין עישון ויתר לחץ דם. נבדק מדגם מקרי של 200 מעשנים ונמצא כי 30 סבלו מיתר לחץ דם. ידוע שבאוכלוסייה 18% סובלים מיתר לחץ דם.

- א. בדקו ברמת מובהקות 0.1 את ההשערה כי אחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
- ב. מהי רמת המובהקות המינימלית לקבלת הטענה שאחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
- ג. מהי עצמת המבחן, אם אחוז הסובלים מיתר לחץ דם בקרב אוכלוסיית המעשנים היא בפועל 25%.

8) להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות. בדקו ברמת מובהקות של 5% :

4	3	2	1	0	מספר הנסיעות
12	20	26	102	84	מספר המשפחות

- א. באיטליה משפחות נוסעות בממוצע פעמיים בשנה לחופשה. האם בישראל משפחות נוסעות פחות מאשר באיטליה?
- ב. בהולנד 80% מהמשפחות נוסעות לפחות פעם אחת בשנה לחופשה, האם בישראל אחוז המשפחות שנוסעות לפחות פעם אחת בשנה לחופשה נמוך מאשר בהולנד?

9) נתון כי: $X \sim N(\mu, \sigma^2 = 10^2)$.

- מעוניינים לבדוק את ההשערות: $H_0: \mu = 40$, $H_1: \mu > 40$.
 דגמו 25 תצפיות מהאוכלוסייה והתקבל $\bar{X} = 45$.
 א. חשבו את p-value (מובהקות התוצאה).
 ב. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu < 40$.
 ג. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu \neq 40$.

תשובות סופיות

- 1) א. חוקר א' ב. 0.0122 ג. גדלה. ד. נדחה H_0 .
- 2) א. לפחות 0.0808 ב. 0.7995
- 3) א. לא נדחה H_0 ב. לא נדחה H_0 .
- 4) א. לא נדחה H_0 ב. בין 5% ל-10% ג. נדחה H_0 .
- 5) א. נדחה H_0 ב. נדחה H_0 .
- 6) א. $H_0: p = 0.6$ ב. נדחה H_0 ג. טעות מסוג ראשון בסיכוי של 0.05.
א. $H_1: p > 0.6$
- 7) א. לא נדחה H_0 ב. 0.8643 ג. 0.8749.
- 8) א. נדחה H_0 ב. נדחה H_0 .
- 9) א. 0.0062 ב. 0.9938 ג. 0.0124.

שאלות סיכום – שאלות רב ברירה על בדיקת השערות

(1) בבדיקת השערה חד-צדדית ימנית ברמת מובהקות $\alpha = 0.01$, נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות $\alpha = 0.05$?

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס לא הייתה נדחית.
- ג. ההשערה המחקרית הייתה נדחית.
- ד. בהעדר נתונים נוספים, לא ניתן לדעת.

(2) על מנת לבדוק האם ההסתברות ללידת בן הינה חצי, נבחר מדגם מקרי של 200 ילדים, ונמצא שישנם 120 בנים. מהן ההשערות האלטרנטיביות להשערת האפס?

א. $H_1: p = 0.5$

ב. $H_1: p = 0.6$

ג. $H_1: p > 0.5$

ד. $H_1: p \neq 0.5$

(3) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית. המבחן שיש לבצע כאן הוא:

- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.

(4) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות:

מהן ההשערות הנבדקות? (ההפרש חושב $X - Y$)

68	82	93	69	לפני הנישואין - X
71	84	88	80	לאחר הנישואין - Y

א. $H_1: \mu_d < 0, H_0: \mu_d = 0$

ב. $H_1: \mu_x - \mu_y < 0, H_0: \mu_x - \mu_y = 0$

ג. $H_1: \mu_x - \mu_y < 0, H_0: \mu_x - \mu_y = 0$

ד. $H_1: \mu_d > 0, H_0: \mu_d = 0$

(5) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

- השערת האפס נכונה.
- השערת האפס נדחתה.
- השערת האפס לא נדחתה.
- אף אחת מהתשובות לא נכונה בהכרח.

(6) ידוע כי ילד בגיל שנתיים ישן בממוצע 9 שעות בלילה. במדגם של 20 תינוקות

בני שנתיים המתגוררים בצפון נמצא, כי ממוצע שעות השינה בלילה הינו 10 עם סטיית תקן של 1.1 במדגם של 10 תינוקות בדרום נמצא, כי ממוצע שעות השינה בלילה הינו 7.9 עם סטיית תקן של 1.1. על מנת להשוות בין ממוצע שעות השינה של ילדים מהצפון לבין זה של כלל הילדים יש לערוך _____, ועל מנת להשוות בין ממוצע שעות השינה של ילדים מהדרום לזה של ילדים המתגוררים בצפון יש לערוך _____.

יש להניח שההנחות הדרושות מתקיימות.

- מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
- מבחן T למדגם יחיד ; מבחן T למדגמים תלויים.
- מבחן T למדגם יחיד ; מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים בלתי תלויים ; מבחן T לממוצע יחיד.

(7) מובהקות התוצאה (PV) היא גם :

- רמת המובהקות המינימאלית לדחות השערת האפס.
- רמת המובהקות המקסימאלית לדחיית השערת האפס.
- רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- רמת המובהקות המינימאלית לאי דחיית השערת האפס.

(8) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח

חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר

נישואיהם. הנה התוצאות :

68	82	93	69	לפני הנישואין
71	84	88	80	לאחר הנישואין

לבדיקת

באיזה התפלגות משתמשים

ההשערות, ובכמה דרגות חופש :

- ההתפלגות Z ללא דרגות חופש.
- ההתפלגות T ו-3 דרגות חופש.
- ההתפלגות T ו-6 דרגות חופש.
- ההתפלגות χ^2 ו-3 דרגות חופש.

- (9) שני סטטיסטיקאים בודקים השערות ברמת מובהקות $\alpha = 0.05$ על סמך אותו מדגם. סטטיסטיקאי א' בודק את ההשערה: $H_0: \mu = 20$ כנגד האלטרנטיבה $H_1: \mu \neq 20$ ומחליט לא לדחות את השערת האפס. סטטיסטיקאי ב' בודק את ההשערה $H_0: \mu \leq 20$ כנגד האלטרנטיבה $H_1: \mu > 20$ מה יחליט סטטיסטיקאי ב'?
- לדחות את השערת האפס.
 - לא לדחות את השערת האפס.
 - ללא נתונים נוספים אי אפשר לדעת מה יחליט.
- (10) חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?
- הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
 - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
 - הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.
 - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
- (11) רמת הכולסטרול בדמם של אנשים מתפלג נורמאלית עם תוחלת של 180 מ"ג (ל 100 סמ"ק דם). וסטיית תקן של 10 מ"ג. מעוניינים לבדוק את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר. נניח שסטיית התקן אצל צמחונים זהה לסטיית התקן של כלל האנשים. במדגם של 20 צמחונים התקבל ממוצע רמת כולסטרול 174.5 מ"ג. אם הוחלט לקבל את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר איזה סוג טעות אפשרית במסקנה?
- טעות מסוג ראשון.
 - טעות מסוג שני.
 - טעות מסוג שלישי.
 - לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחונים.

12 בסקר שנערך התקבל ש 60% מתוך 220 נשאלים מבקרים אצל השיננית לפחות פעם אחת בשנה. עבור אילו רמות מובהקות ניתן יהיה לקבוע שרוב האוכלוסייה מבקרת אצל השיננית לפחות פעם בשנה?

- א. רמת מובהקות הגדולה מ-5%.
- ב. רמת מובהקות הקטנה מ-5%.
- ג. רמת מובהקות הגדלה מ-0.0015.
- ד. רמת מובהקות הקטנה מ-0.0015.

13 שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה.

חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?

- א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.
- ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.
- ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.
- ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

14 ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?

- א. תוחלת ציוני העולים באוכלוסייה.
- ב. ממוצע ציוני העולים במדגם.
- ג. תוחלת ציוני האוכלוסייה מנתוני העבר.
- ד. ממוצע ציוני שאר האוכלוסייה במדגם.

15 חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?

- א. החוקר דחה את השערת H_0 כאשר היא הייתה נכונה.
- ב. החוקר דחה את השערת H_1 כאשר היא הייתה נכונה.
- ג. החוקר לא דחה את השערת H_0 כאשר היא הייתה לא נכונה.
- ד. המדגם של החוקר שייך בפועל להתפלגות הדגימה של H_1 .

- 16** חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?
- מבחן T למדגמים בלתי תלויים עם 38 דרגות חופש.
 - מבחן T למדגמים מזווגים, עם 39 דרגות חופש.
 - מבחן T למדגמים בלתי תלויים עם 39 דרגות חופש.
 - מבחן T למדגמים מזווגים עם 38 דרגות חופש.

- 17** בינואר השנה פורסם שהשכר הממוצע במשק הוא 8,900 ₪. במדגם שנעשה בחודש יוני על 60 עובדים נרשם עבור כל עובד במדגם האם השכר שלו נמוך או לא נמוך מהשכר הממוצע שפורסם בחודש ינואר. מהו המבחן המתאים כדי לבדוק שרוב העובדים בחודש יוני קיבלו שכר הנמוך מהשכר הממוצע שפורסם בחודש ינואר?
- מבחן Z על פרופורציה.
 - מבחן T על תוחלת אחת.
 - מבחן T על שתי תוחלות במדגמים בלתי תלויים.
 - מבחן T על שתי תוחלות במדגמים תלויים.

- 18** שלושה חוקרים רצו לבדוק את השפעתו של שידור פרסומות נגד תאונות דרכים על מהירות הנהיגה של נהגים בישראל (השונות של מהירות הנהיגה בישראל אינה ידועה). עידו השווה את מהירות הנהיגה של קבוצת נהגים אחת, חודש לפני שידור הפרסומות וחודש לאחר שידור הפרסומות. רון השווה את מהירות הנהיגה של קבוצת נהגים, שראו את הפרסומות, למהירות הנהיגה של קבוצת נהגים, שלא ראו את הפרסומות. יואב השווה את מהירות הנהיגה של קבוצת נהגים בחודש בו שודרו הפרסומות, למהירות הנהיגה הממוצעת בישראל על פי נתוני משרד התחבורה. המבחנים בהם צריכים החוקרים להשתמש הם:
- שלושתם במבחן T למדגמים בלתי תלויים.
 - עידו במבחן T למדגמים מזווגים, ורון ויואב במבחן T למדגמים בלתי תלויים.
 - עידו במבחן T למדגמים מזווגים, רון במבחן T למדגמים בלתי תלויים ויואב במבחן T למדגם יחיד.
 - עידו במבחן T למדגמים מזווגים, רון ויואב במבחן T למדגם יחיד.

- 19** במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%. מה תמיד נכון?
- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
 - הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות תשנה את מסקנת המחקר.
- 20** חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס. אם החוקר היה עורך מבחן חד צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח:
- השערת האפס הייתה נדחית.
 - השערת האפס הייתה לא נדחית.
 - לא ניתן לדעת מה תהיה מסקנתו במקרה זה.
- 21** ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם.
- ליאור השתמש בטבלה של התפלגות Z.
 - רוני השתמשה בטבלה של התפלגות T.
 - מה נוכל לומר בנוגע להחלטת המחקר שלהם?
 - אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
 - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
 - שני החוקרים בהכרח יגיעו לאותה מסקנה.
 - לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.
- 22** נתון ש $X \sim N(\mu, \sigma^2)$ כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$, $H_1: \mu < \mu_0$.
- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.
- כעת בברור הוא ידחה את השערת האפס.
 - כעת הוא דווקא יקבל את השערת האפס.
 - כעת לא ניתן לדעת מה תהיה מסקנתו.
- 23** אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:
- הסיכוי לטעות מסוג ראשון גדל.
 - העוצמה של המבחן גדלה.
 - הסיכוי לטעות מסוג שני גדל.
 - תשובות א ו-ב נכונות.

(24) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

- השערת האפס נכונה.
- השערת האפס נדחתה.
- השערת האפס לא נדחתה.
- אף אחת מהתשובות לא נכונה בהכרח.

(25) מה המצב הרצוי לחוקר המבצע בדיקת השערה :

- | | |
|-------------|----------|
| $1 - \beta$ | α |
| א. גדולה | א. גדולה |
| ב. קטנה | ב. גדולה |
| ג. גדולה | ג. קטנה |
| ד. קטנה | ד. קטנה |

(26) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך :

- הן α , והן $(1 - \beta)$, יקטנו.
- α יישאר ללא שינוי ואילו $(1 - \beta)$ יגדל.
- α יגדל ואילו $(1 - \beta)$ יקטן.
- הן α והן $(1 - \beta)$ יגדלו.

(27) ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?

- טעות מסוג ראשון.
- טעות מסוג שני.
- טעות מסוג שלישי.
- אין טעות במסקנתו.

(28) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$. מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה :

- יקבל את השערת האפס בכל מקרה.
- ידחה את השערת האפס מקרה.
- ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
- לא ניתן לדעת כי אין מספיק נתונים.

(29) מובהקות התוצאה (PV) היא גם :

- א. רמת המובהקות המינימאלית לדחות השערת האפס.
- ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
- ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

(30) בבדיקת השערות מסוימת התקבל $p \text{ value} = 0.0254$, לכן :

- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
- ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
- ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
- ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

(31) רמת המובהקות במחקר הייתה 2% לכן.

- א. בסיכוי של 2% נדחה את השערת האפס.
- ב. בסיכוי של 2% לא נדחה את השערת האפס.
- ג. בסיכוי של 2% השערת האפס לא נכונה.
- ד. אף תשובה לא נכונה.

(32) נתון ש: $X \sim N(\mu, \sigma^2)$. כמו כן נתונות ההשערות הבאות: $H_0: \mu = \mu_0$, $H_1: \mu < \mu_0$.

- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.
- σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5%. אם הוא היה מגדיל את רמת המובהקות ל-10% אזי :
- א. כעת בברור הוא ידחה את השערת האפס.
- ב. כעת הוא דווקא יקבל את השערת האפס.
- ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

(33) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים

- זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה.
- לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית. מספר דרגות החופש במבחן הוא :

א. 9

ב. 19

ג. 18

ד. 8

34) בתחנת טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

משקל במכשיר 1	2.5	0.7	9.6	4.5
משקל במכשיר 2	0.5	1.7	6.9	3.5

נניח שהמשקלים מתפלגים נורמלית.

המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים מזווגים.

35) כדי להשוות בין שני אצים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים. המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים מזווגים.

36) סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של תוצאה זו היא:

- לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
- תמיד לדחות את השערת האפס כאשר היא נכונה.
- לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- תמיד לדחות את השערת האפס כאשר היא לא נכונה.

37) סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים. הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$. אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות: $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 \neq 0$,

ברמת מובהקות 0.05 מסקנתו תהיה:

- לדחות את השערת האפס.
- לא לדחות את השערת האפס.
- שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
- שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

(38) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפללים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו: $-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%. לכן מסקנת המחקר היא:

- א. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ב. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- ג. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D .

(39) אם רמת מובהקות של מבחן סטטיסטי הינה 0, הכוונה היא:

- א. תמיד נדחה H_0 כאשר היא נכונה, אך לא תמיד נדחה אותה כאשר היא לא נכונה.
- ב. לא נדחה את H_0 אף פעם.
- ג. לא נדחה את H_0 כאשר היא נכונה אך יתכן ונדחה אותה כאשר היא לא נכונה.
- ד. כל התשובות לא נכונות.

(40) חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות: $H_0: \mu = 10$, $H_1: \mu \neq 10$. לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל: $t_x = -2.63$. לכן המסקנה היא:

- א. הוא ידחה H_0 ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
- ב. הוא ידחה H_0 ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
- ג. הוא ידחה H_0 ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
- ד. הוא לא ידחה H_0 ברמת מובהקות 0.1.

(41) האיגוד האמריקני לרפואת ילדים מפרסם הנחיות חדשות הקובעות כי יש ליטול תוספת יוד במהלך תקופת ההיריון וההנקה. מחסור במינרל זה עלול לגרום לפגיעה מוחית אצל העובר והתינוק. החלטה זו נקבעה על סמך מחקר בו השתתפו 1050 נשים שנטלו יוד במהלך תקופת ההיריון וההנקה. מתוך הנשים שהשתתפו במחקר, רק ל-21 נמצאו ילדים בעלי פגיעה מוחית לעומת 3% באוכלוסייה הכללית. בנוסף, פורסם שהאיגוד האמריקאי מגיע למסקנותיו על סמך רמת מובהקות של 0.5%. מה הסיכוי לבצע טעות מסוג ראשון במחקר?

- א. 0.005
- ב. 0.03
- ג. 0.0287
- ד. 0.05

42) חוקרת שיערה, כי משקלן של נשים כשנה לאחר החתונה גבוה ממשקלן בעת החתונה. החוקרת דגמה 15 נשים, ובדקה את משקלן בשתי נקודות הזמן (בעת החתונה, ושנה לאחריה), אך לא מצאה הבדל מובהק ברמת מובהקות 0.01. בהנחה, כי **במציאות** השערתה של החוקרת נכונה, סביר כי אם היא תגדיל את גודל המדגם, אזי:

- א. יקטן הסיכוי לטעות מסוג שני (β).
- ב. תגדל רמת הביטחון ($1 - \alpha$).
- ג. אף תשובה לא נכונה.
- ד. כל התשובות נכונות.

43) איזה מהמשפטים הבאים נכון תמיד?

- א. $POWER + \alpha + \beta = 1$
- ב. $POWER = 0.5 - \beta$
- ג. $POWER + \alpha = 1$
- ד. $\beta + \alpha = 1$
- ה. הכול לא נכון.

44) מה נכון לומר לגבי הנחת שיוויון השונויות במבחן T למדגמים בלתי תלויים?

- א. היא אומרת שהשונויות המדגמיות שוות.
- ב. בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.
- ג. היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.
- ד. אף תשובה אינה נכונה.

45) חוקר החליט לא לדחות השערה ברמת מובהקות של α . במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של 2α על סמך אותם נתונים, האם ההשערה תדחה?

- א. ההשערה תדחה.
- ב. ההשערה לא תדחה.
- ג. התשובה תלויה בעוצמת המבחן.
- ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

46) חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה:

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

(47) עוצמה שווה ל-1 פרושה :

- א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
- ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

(48) מה מהבאים נכון לגבי מבחן T מדגמים מזווגים?

- א. כל התצפיות במחקר אינן תלויות זו בזו.
- ב. כל התצפיות במחקר תלויות זו בזו.
- ג. כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
- ד. התצפיות בתוך כל צמד אינן תלויות זו בזו.

(49) לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית $H_0: \mu \geq 10$, נלקח מדגם והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה הדו צדדית $H_0: \mu = 10$, $H_1: \mu \neq 10$, אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05 :

- א. ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
- ב. מקבלים את השערת האפס.
- ג. דוחים את השערת האפס.
- ד. לא ניתן להכריע בין ההשערות שכן חסרים נתונים.

(50) לבדיקת ההשערה החד צדדית ימנית $H_0: \mu = 55$, $H_1: \mu = 65$,

נלקח מדגם מקרי בגודל n מאוכלוסייה בעלת התפלגות נורמלית ושונות σ^2 . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות :

- עבור מדגם בגודל n וברמת מובהקות 5% לבדיקת ההשערות : $H_0: \mu = 55$, $H_1: \mu = 60$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות : $H_0: \mu = 55$, $H_1: \mu = 65$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות : $H_0: \mu = 55$, $H_1: \mu = 65$ העוצמה תהיה קטנה מ-0.9.

- א. שלושת הטענות אינן נכונות.
- ב. טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.
- ג. טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.
- ד. טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.

תשובות סופיות:

שאלה	תשובה	שאלה	תשובה
1	א	26	א
2	ד	27	ב
3	ד	28	א
4	א	29	א
5	ג	30	ג
6	ג	31	ד
7	א	32	א
8	ב	33	א
9	ג	34	ד
10	א	35	ג
11	א	36	א
12	ג	37	ג
13	א	38	א
14	א	39	ג
15	א	40	א
16	א	41	א
17	א	42	א
18	ג	43	ה
19	א	44	ד
20	ג	45	ד
21	ב	46	א
22	ג	47	ד
23	ד	48	ג
24	ג	49	ב
25	ג	50	ד

סטטיסטיקה

פרק 30 - שאלות מסכמות על רווחי סמך

תוכן העניינים

1. שאלות מסכמות על רווחי סמך 179

שאלות מסכמות על רווחי סמך:

שאלות:

- (2) 200 אנשים נשאלו כמה פעמים ביום הם שותים כוס קפה. להלן התפלגות התשובות:

מספר פעמים	0	1	2	3	4	5
מספר אנשים	86	34	28	22	20	10

- א. תנו רווח סמך לממוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום. $\alpha = 0.05$.
- ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז "המכורים לקפה". $\alpha = 0.1$.
- (3) חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה. רווח הסמך שהתקבל הוא: $81 < p < 91$. רווח הסמך הנ"ל התבסס על מדגם של 500 איש.
- א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?
- ב. באיזו רמת סמך נבנה רווח הסמך?
- ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 96% על סמך תוצאות המדגם.
- (4) ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן ה-IQ התקבלו התוצאות הבאות:
- $$\sum_{i=1}^{20} x_i = 2040, \quad \sum_{i=1}^{20} x_i^2 = 210740$$
- א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה-IQ בישראל – מהי ההנחה הדרושה לפתרון?
- ב. על סמך רווח הסמך של סעיף א' האם תקבלו את הטענה שבישראל ממוצע הציונים שונה מארה"ב?
- ג. מה היה קורה לרווח הסמך אם היינו מגדילים את רמת הסמך שלו?

- (5) להלן תוצאות מדגם שבדק עבור כל משפחה האם יש לה בבית מכשיר טאבלט:

אזור מגורים	גוש דן	שאר הארץ
גודל המדגם	200	240
מספר משפחות בעלי טאבלט	160	168

- א. בנו רווח סמך להבדל בין אחוז המשפחות עם טאבלט בגוש דן ואחוז המשפחות בעלי טאבלט בשאר חלקי הארץ. ברמת סמך של 98%.
- ב. בנו רווח סמך לפרופורציות משפחות בעלות טאבלט בכלל הארץ ברמת סמך של 95%.

- (7) בנק מתלבט האם לפתוח סניף באזור A או באזור B. לצורך פתרון נניח שסטיית התקן של המשכורת באזור A היא 1200 ובאזור B 1500. הבנק דגם 50 אנשים מאזור A, המשכורת הממוצעת שהתקבלה במדגם היא 6,800 ₪. כמו כן, נדגמו 40 אנשים מאזור B, המשכורת הממוצעת שהתקבלה במדגם היא 6,600 ₪.

- א. בנו רווח סמך ברמת סמך של 95% להפרש הממוצעים של המשכורות בשני האזורים. האם על סמך רווח הסמך ניתן להמליץ לבנק היכן לפתוח את הסניף. אם כן, היכן?
- ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

- (8) להלן מדגם של שכר הדירה ב-₪ של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב:

7500	6500	7000	7500	8000	שנת 2012
7700	6800	7800	8200	8000	שנת 2013

- בנו רווח סמך ברמת סמך של 95% לתוחלת עליית שכר הדירה משנת 2012 לשנת 2013 בשכונת בבלי. ניתן להניח ששכר הדירה בשכונה מתפלג נורמלית.

תשובות סופיות:

- (2) א. $1.21 \leq \mu \leq 1.65$ ב. $10.85\% \leq p \leq 19.15\%$
- (3) א. 70 ב. 0.9988 ג. $83\% < p < 89\%$
- (4) א. $97.4 \leq \mu \leq 106.6$ ב. לא ג. יגדל.
- (5) א. $0.5\% \leq p_1 - p_2 \leq 19.5\%$ ב. $0.704 \leq p \leq 0.786$
- (7) א. $-372 \leq \mu_A - \mu_B \leq 772$, לא ב. $6467 \leq \mu_A \leq 7133$
- (8) $-21 \leq \mu_D \leq 821$

סטטיסטיקה

פרק 31 - יסודות ההסתברות

תוכן העניינים

1. כללי 182

הגדרות יסודיות:

רקע:

ניסוי מקרי: תהליך לו כמה תוצאות אפשריות. התוצאה המתקבלת נודעת רק לאחר ביצוע התהליך. למשל: תוצאה בהטלת קובייה, מזג האוויר בעוד שבועיים.

מרחב מדגם: כלל התוצאות האפשריות בניסוי המקרי. לדוגמה, בהטלת קובייה: $\{1, 2, 3, 4, 5, 6\}$, או: מזג האוויר בעוד שבועיים: $\{\text{נאה, שרבי, מושלג, גשום, מעונן, חלקית, אביד}\}$.

מאורע: תת קבוצה מתוך מרחב במדגם. מסומן באותיות: A, B, C . בהטלת קובייה למשל, המאורע 'לקבל לפחות 5' יסומן: $A = \{5, 6\}$. המאורע 'לקבל תוצאה זוגית' יסומן: $B = \{2, 4, 6\}$.

גודל מרחב המדגם: מספר התוצאות האפשריות במרחב המדגם. בהטלת קובייה למשל נקבל: $|\Omega| = 6$.

גודל המאורע: מספר התוצאות האפשריות במאורע עצמו. למשל, בהטלת הקובייה האירועים הקודמים יסומנו: $|A| = 2, |B| = 3$.

מאורע משלים: מאורע המכיל את כל התוצאות האפשריות במרחב המדגם פרט לתוצאות במאורע אותו הוא משלים. למשל, בהטלת הקובייה: $\bar{A} = \{1, 2, 3, 4\}$, $\bar{B} = \{1, 3, 5\}$.

מרחב מדגם אחיד (סימטרי): מרחב מדגם בו לכל התוצאות במרחב המדגם יש את אותה עדיפות, אותה סבירות למשל, קובייה הוגנת, אך לא כמו מזג האוויר בשבוע הבא.

הסתברות במרחב מדגם אחיד: במרחב מדגם אחיד הסיכוי למאורע יהיה: $P(A) = \frac{|A|}{|\Omega|}$.

דוגמה: מה הסיכוי בהטלת קובייה לקבל לפחות 5? $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6}$.

דוגמה: מה הסיכוי בהטלת קובייה לקבל תוצאה זוגית? $P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6}$.

הסתברות במרחב לא אחיד: תחושב לפי השכיחות היחסית: $\frac{f}{n}$.

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

הציון x	מספר התלמידים – השכיחות f
5	2
6	4
7	8
8	5
9	4
10	2

מה ההסתברות שתלמיד אקראי שניבחר בכיתה קיבל את הציון 8? $\frac{f}{n} = \frac{5}{25} = 0.2$.

מה ההסתברות שתלמיד אקראי שניבחר בכיתה יכשל? $\frac{f}{n} = \frac{2}{25} = 0.08$.

הסתברות למאורע משלים: הסתברות לקבוצת המשלים של המאורע ביחס למרחב המדגם: $P(\bar{A}) = 1 - P(A)$. למשל, בדוגמה הקודמת הסיכוי לעבור את הבחינה יכול

להיות מחושב לפי הסיכוי להיכשל: $P(A) = 1 - \frac{2}{25} = \frac{23}{25}$.

שאלות:

(1) מהאותיות E, F ו- G יש ליצור מילה בת 2 אותיות, לא בהכרח בת משמעות.

א. הרכיבו את כל המילים האפשריות.

ב. רשמו את המקרים למאורע:

i. במילה נמצאת האות E .

ii. במילה האותיות שונות.

ג. רשמו את המקרים למאורע $\bar{A}$.

(2) מטילים זוג קוביות.

א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם אחיד?

ב. רשמו את כל האפשרויות לאירועים הבאים:

i. סכום התוצאות 7.

ii. מכפלת התוצאות 12.

ג. חשבו את הסיכויים לאירועים שהוגדרו בסעיף ב'.

(3) נבחר באקראי ספרה מבין הספרות 0-9.

א. מה ההסתברות שהספרה שנבחרה גדולה מ-5?

ב. מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?

ג. מה ההסתברות שהספרה שנבחרה היא אי זוגית?

(4) להלן התפלגות מספר מקלטי הטלוויזיה עבור כל משפחה בישוב מסוים:

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

נבחרה משפחה באקראי מהישוב.

א. מה ההסתברות שאין מקלטים למשפחה?

ב. מה ההסתברות שיש מקלטים למשפחה?

ג. מה ההסתברות שיש לפחות 3 מקלטים למשפחה?

(5) להלן התפלגות מספר המכוניות למשפחה ביישוב "עדן":

מספר משפחות	20	40	100	30	10
מספר מכוניות	0	1	2	3	4

נבחרה משפחה אקראית מן הישוב.

א. מה ההסתברות שאין לה מכוניות?

ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?

ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

- (6) נטיל מטבע רגיל 3 פעמים. בצד אחד של המטבע מוטבע עץ ובצד השני פלי.
 א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם הוא אחיד?
 ב. רשמו את כל האפשרויות לאירועים הבאים:
 i. התקבל פעם אחת עץ.
 ii. התקבל לפחות פלי אחד.
 ג. מהו המאורע המשלים ל-D?
 ד. חשבו את הסיכויים לאירועים שהוגדרו בסעיפים ב-ג.

תשובות סופיות:

- (1) א. $\Omega = \{EE, EF, EG, FE, FF, FG, GE, GF, GG\}$
 ב. $A = \{EE, EF, EG, FE, GE\}$, $B = \{EF, EG, FE, FG, GE, GF\}$
 ג. $\bar{A} = \{FF, FG, GF, GG\}$
- (2) א. $\Omega = \left\{ \begin{pmatrix} (1,1) & (2,1) & (3,1) & (5,1) & (4,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{pmatrix} \right\}$
 ב. $A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$, $C = \{(2,6), (3,4), (4,3), (6,2)\}$
 ג. הסיכוי ל-A: $\frac{1}{6}$. הסיכוי ל-B: $\frac{1}{9}$.
- (3) א. 0.4 ב. 0.4 ג. 0.5
 (4) א. 0.22 ב. 0.78 ג. 0.32
 (5) א. 0.1 ב. 0.2 ג. 0.8
 (6) א. $\Omega = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP, EEE\}$
 ב. $A = \{PPE, PEP, EPP\}$, $D = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP\}$
 ג. $\bar{D} = \{EEE\}$
 ד. $\frac{1}{8}$

סטטיסטיקה

פרק 32 - פעולות בין מאורעות (חיתוך ואיחוד) - מאורעות זרים ומכילים

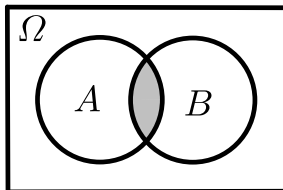
תוכן העניינים

1. כללי.....186

פעולות בין מאורעות (חיתוך ואיחוד) – מאורעות זרים ומכילים:

רקע:

פעולת חיתוך:

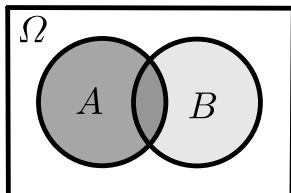


נותנת את המשותף בין המאורעות הנחתכים.
חיתוך בין המאורע A למאורע B יסומן כך: $A \cap B$.
מדובר בתוצאות שנמצאות ב- A וגם ב- B .

דוגמה:

בהטלת קובייה, למשל, האפשרויות לקבל לפחות 5 הן: $A = \{5, 6\}$.
האפשרויות לקבל תוצאה זוגית הן: $B = \{2, 4, 6\}$.
החיתוך שביניהם הוא: $A \cap B = \{6\}$.

פעולת איחוד:



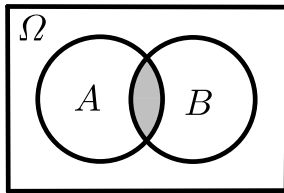
נותנת את כל האפשרויות שנמצאות לפחות באחת מהמאורעות, ומסומנת: $A \cup B$.
הפעולה נותנת את אשר נמצא ב- A או ב- B .
כלומר, לפחות אחד מהמאורעות קורה.

דוגמה:

בהטלת קובייה האפשרויות לקבל לפחות 5 הן: $A = \{5, 6\}$.
האפשרויות לקבל תוצאה זוגית: $B = \{2, 4, 6\}$.
האפשרויות לקבל לפחות 5 וגם תוצאה זוגית: $A \cup B = \{2, 4, 5, 6\}$.

דוגמה (הפתרון נמצא בהקלטה):

סטודנט ניגש בסמסטר לשני מבחנים. מבחן בסטטיסטיקה ומבחן בכלכלה. ההסתברות שלו לעבור את המבחן בסטטיסטיקה הוא 0.9, ההסתברות שלו לעבור את המבחן בכלכלה הוא 0.8 וההסתברות לעבור את המבחן בסטטיסטיקה ובכלכלה היא 0.75. מה ההסתברות שלו לעבור את המבחן בסטטיסטיקה בלבד? מה ההסתברות שלו להיכשל בשני המבחנים? מה ההסתברות לעבור לפחות מבחן אחד?

נוסחת החיבור לשני מאורעות:

ההסתברות של איחוד מאורעות תחושב ע"י הקשר הבא:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

חוקי דה מורגן לשני מאורעות:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

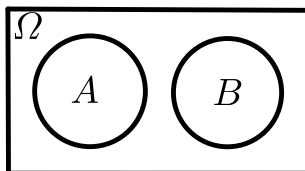
$$P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$$

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$$

שיטת ריבוע הקסם:

השיטה רלבנטית רק אם יש שני מאורעות במקביל בדומה לתרגיל הקודם:

	$\bar{A}$	A	
B	$P(\bar{A} \cap B)$	$P(A \cap B)$	$P(B)$
$\bar{B}$	$P(\bar{A} \cap \bar{B})$	$P(A \cap \bar{B})$	$P(\bar{B})$
	$P(\bar{A})$	$P(A)$	1

מאורעות זרים:

מאורעות זרים הם כאשר אין להם אף איבר משותף:
 $A \cap B = \{ \}$. כלומר, הם לא יכולים להתרחש בו זמנית.

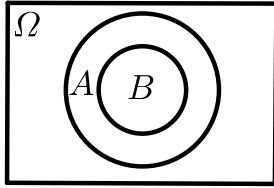
ההסתברות של חיתוך המאורעות היא אפס: $P(A \cap B) = 0$.

ההסתברות של איחוד המאורעות תחושב: $P(A \cup B) = P(A) + P(B)$.

דוגמה:

בהטלת קובייה, האפשרויות לקבל לפחות 5 הן: $A = \{5, 6\}$ והאפשרות לקבל 3

היא: $B = \{3\}$, ולכן החיתוך ביניהם הוא אפס, כלומר: $A \cap B = \{ \}$.

מאורעות מוכללים:

נתונים שני מאורעות A ו- B , השונים מאפס.
נאמר שהמאורע B מוכלל במאורע A אם כל איברי
המאורע B כלולים במאורע A ונרשום: $B \subset A$.

מאורע A מכיל את מאורע B כל התוצאות שנמצאות ב- B
מוכללות בתוך מאורע A .

קשר זה מסומן באופן הבא: $B \subset A$.

$$A \cap B = B \quad P(A \cap B) = P(B)$$

$$A \cup B = A \quad P(A \cup B) = P(A)$$

$$\begin{aligned} A &= \{2, 4, 6\} \\ B &= \{2, 4\} \end{aligned} \quad \text{למשל:}$$

שאלות:

- (1) מהאותיות E, F ו- G יוצרים מילה בת 2 אותיות – לא בהכרח בת משמעות. נגדיר את המאורעות הבאים:
- A - במילה נמצאת האות E .
 - B - במילה אותיות שונות.
- א. רשמו את כל האפשרויות לחיתוך A עם B .
- ב. רשמו את כל האפשרויות לאיחוד של A עם B .
- (2) תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה. נגדיר את המאורעות הבאים:
- A - לעבור את המבחן בסטטיסטיקה.
 - B - לעבור את המבחן בכלכלה.
- היעזרו בפעולות חיתוך, איחוד ומשלים בלבד כדי להגדיר את המאורעות הבאים וסמנו בדיאגרמת וון את השטח המתאים:
- א. התלמיד עבר רק את המבחן בכלכלה.
 - ב. התלמיד עבר רק את המבחן בסטטיסטיקה.
 - ג. התלמיד עבר את שני המבחנים.
 - ד. התלמיד עבר לפחות מבחן אחד.
 - ה. התלמיד נכשל בשני המבחנים.
 - ו. התלמיד נכשל בכלכלה.
- (3) נתבקשתם לבחור ספרה באקראי. נגדיר את A להיות הספרה שנבחרה היא זוגית. נגדיר את B להיות הספרה שנבחרה קטנה מ-5.
- א. רשמו את כל התוצאות למאורעות הבאים:
- $$A \cup B, A \cap B, \bar{B}, B, A$$
- ב. חשבו את ההסתברויות לכל המאורעות מהסעיף הקודם.
- (4) נסמן ב- Ω את מרחב המדגם וב- ϕ קבוצה ריקה.
- נתון כי A הינו מאורע בתוך מרחב המדגם.
- להלן מוגדרים מאורעות שפתרונם הוא Ω או ϕ או A .
- קבעו עבור כל מאורע מה הפתרון שלו:
- $$A \cup \bar{A}, \bar{\phi}, A \cap \bar{A}, A \cup \Omega, A \cap \Omega, A \cup \phi, A \cap \phi, \bar{A}$$

(5) הוגדרו המאורעות הבאים :

A - אדם שגובהו מעל 1.7 מטר

B - אדם שגובהו מתחת ל-1.8 מטר.

קבעו את גובהם של האנשים הבאים :

א. $A \cap B$

ב. $A \cup B$

ג. $\overline{A} \cap B$

ד. $\overline{A} \cup \overline{B}$

ה. $\overline{\overline{A}}$

(6) נגדיר את המאורעות הבאים :

A - אדם דובר עברית.

B - אדם דובר ערבית.

C - אדם דובר אנגלית.

השתמשו בפעולות איחוד, חיתוך והשלמה לתיאור המאורעות הבאים :

א. אדם דובר את כל שלוש השפות.

ב. אדם דובר רק עברית.

ג. אדם דובר לפחות שפה אחת מתוך השפות הללו.

ד. אדם אינו דובר אנגלית.

ה. קבוצת התלמידים שדוברים שתי שפות בדיוק (מהשפות הנ"ל).

(7) שתי מפלגות רצות לכנסת הבאה. מפלגת "גדר" תעבור את אחוז החסימה בהסתברות של 0.08 ומפלגת "עתיד" תעבור את אחוז החסימה בהסתברות של 0.20. בהסתברות של 76% שתי המפלגות לא תעבורנה את אחוז החסימה.

א. מה ההסתברות שלפחות אחת מהמפלגות תעבור את אחוז החסימה?

ב. מה ההסתברות ששתי המפלגות תעבורנה את אחוז החסימה?

ג. מה ההסתברות שרק מפלגות "עתיד" תעבור את אחוז החסימה?

(8) במקום עבודה מסוים 40% מהעובדים הם גברים. כמו כן, 20% מהעובדים הם אקדמאים. 10% מהעובדים הינן נשים אקדמאיות.

א. איזה אחוז מהעובדים הם גברים אקדמאיים?

ב. איזה אחוז מהעובדים הם גברים או אקדמאיים?

ג. איזה אחוז מהעובדים הם נשים לא אקדמאיות?

(9) הסיכוי של מניה A לעלות הנו 0.5 ביום מסוים והסיכוי של מניה B לעלות ביום מסוים הנו 0.4. בסיכוי של 0.7 לפחות אחת מהמניות תעלה ביום מסוים. חשבו את ההסתברויות הבאות לגבי שתי המניות הללו ביום מסוים:

א. ששתי המניות תעלנה.

ב. שאף אחת מהמניות לא תעלנה.

ג. שמניה A בלבד תעלה.

(10) מטילים זוג קוביות, אדומה ושחורה. נגדיר את המאורעות הבאים:

A - בקובייה האדומה התקבלה התוצאה 4 ובשחורה 2.

B - סכום התוצאות משתי הקוביות הוא 6.

C - מכפלת התוצאות בשתי הקוביות היא 10.

א. האם A ו-B מאורעות זרים?

ב. האם המאורע B מכיל את המאורע A?

ג. האם A ו-C מאורעות זרים?

ד. האם A ו-C מאורעות משלימים?

(11) עבור המאורעות A ו-B ידועות ההסתברויות הבאות: $P(A) = 0.6$,

$$P(\bar{A} \cap \bar{B}) = 0.1, P(B) = 0.3$$

א. האם A ו-B מאורעות זרים?

ב. חשבו את $P(\bar{A} \cap B)$.

(12) מטבע הוטל פעמיים. נגדיר את המאורעות הבאים:

A - קיבלנו עץ בהטלה הראשונה.

B - קיבלנו לפחות עץ אחד בשתי ההטלות.

איזו טענה נכונה?

א. A ו-B מאורעות זרים.

ב. A ו-B מאורעות משלימים.

ג. B מכיל את A.

ד. A מכיל את B.

(13) בהגרלה חולקו 100 כרטיסים. על 3 מהם רשום חופשה ועל 2 מהם רשום מחשב. שאר הכרטיסים ריקים. אדם קיבל כרטיס אקראי.

א. מה הסיכוי לזכות בחופשה או במחשב? האם המאורעות הללו זרים?

ב. מה ההסתברות לא לזכות בפרס?

14 נתון כי: $P(A) = 0.3$, $P(B) = 0.25$, $P(A \cup B) = 0.49$

- חשבו את הסיכוי ל- $P(A \cap B)$.
- האם A ו- B מאורעות זרים?
- מה ההסתברות שרק A יקרה או שרק B יקרה?

15 A ו- B מאורעות זרים. נתון ש: $2 \cdot P(B \cap \bar{A}) = P(A \cap \bar{B}) = P(\bar{A} \cap \bar{B})$

מה הסיכוי למאורע A ומה ההסתברות למאורע B ?

16 קבעו אילו מהטענות הבאות נכונות:

- $A \cap B = B \cap A$.
- $\overline{A \cup B} = A \cap B$.
- $A \cap B \cap C = A \cap B \cap (C \cup B)$.
- $\overline{A \cap B \cap C} = \bar{A} \cup \bar{B} \cup \bar{C}$.

17 נתון ש- A ו- B מאורעות במרחב מדגם. נתון ש- $P(A) = 0.3$, $P(B) = 0.2$.

- האם יתכן ש- $P(A \cup B) = 0.4$?
- האם יתכן ש- $P(A \cup B) = 0.6$?
- אם A ו- B זרים מה הסיכוי $P(A \cup B)$?
- אם A מכיל את B מה הסיכוי $P(A \cup B)$?

18 מתוך אזרחי המדינה הבוגרים ל-30% חשבון בבנק הפועלים. ל-28% חשבון בבנק לאומי ול-15% חשבון בבנק מזרחי. כמו כן נתון כי 6% מחזיקים חשבון בבנק לאומי ובבנק הפועלים. ל-5% חשבון בבנק פועלים ומזרחי. ול-4% חשבון בבנק לאומי ומזרחי. כמו כן ל-1% מהאוכלוסייה הבוגרת חשבון בנק בשלושת הבנקים יחד.

- מה אחוז האזרחים להם חשבון בבנק לאומי בלבד?
- מה ההסתברות שאזרח כלשהו יחזיק חשבון בבנק פועלים ולאומי אבל לא בבנק מזרחי?
- מה ההסתברות שלאזרח יהיה חשבון בפועלים או במזרחי אבל לא בבנק לאומי?
- מה אחוז האזרחים שיש להם חשבון בנק אחד בלבד?
- מה אחוז האזרחים שיש להם בדיוק חשבון בשני בנקים בלבד?
- מה ההסתברות שלאזרח בוגר אין חשבון בנק באף אחד מהבנקים הללו?
- לאיזה אחוז מהאזרחים יש חשבון בנק בלפחות אחד מהבנקים הללו?

- 19** חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21. הנתונים שהתקבלו היו: 40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראכרט", 20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים כרטיס ויזה וגם ישראכרט, 8% מחזיקים כרטיס ישראכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם אמריקן אקספרס. כמו כן, 13% לא מחזיקים באף אחד משלושת הכרטיסים הנ"ל.
- א. מה אחוז מחזיקי שלושת כרטיסי האשראי גם יחד?
- ב. מה אחוז מחזיקי ישראכרט וויזה אך לא את אמריקן אקספרס?
- ג. מה אחוז מחזיקי כרטיס אחד בלבד?

20 הוכיחו: $P(\bar{A} \cap \bar{B}) = 1 - P(A) - P(B) + P(A \cap B)$.

- 21** A ו- B מאורעות במרחב המדגם. האם נכון לומר שהסיכוי שיתרחש בדיוק מאורע אחד הוא: $P(A) + P(B) - 2P(A \cap B)$?

תשובות סופיות:

- (1) א. $A \cap B = \{EG, EF, FE, GE\}$ ב. $A \cup B = \{EG, EF, EE, FE, GE, EG, GF\}$
- (2) א. $B \cap \bar{A}$ ב. $A \cap \bar{B}$ ג. $A \cap B$ ד. $A \cup B$ ה. $\bar{A} \cap \bar{B}$ ו. $\bar{B}$
- (3) א. $\bar{B} = 5, 6, 7, 8, 9, B = 0, 1, 2, 3, 4, A = 0, 2, 4, 6, 8$
 $A \cup B = 0, 2, 4, 6, 8, 1, 3, A \cap B = 0, 2, 4$
- ב. $P(A \cup B) = 0.7, P(A \cap B) = 0.3, P(\bar{B}) = 0.5, P(B) = 0.5, P(A) = 0.5$
- (4) $\bar{\bar{A}} = A, A \cap \phi = \phi, A \cup \phi = A, A \cap \Omega = A, A \cup \Omega = \Omega, A \cup \bar{A} = \Omega, \bar{\phi} = \Omega, A \cap \bar{A} = \phi$
- (5) א. $A \cap B$: גובה בין 1.7 ל-1.8. ב. $A \cup B$: כל גובה אפשרי.
 ג. $\bar{A} = \bar{A} \cap B$: גובה לכל היותר 1.7. ד. $\bar{A} \cup \bar{B}$: לכל היותר 1.7 או לפחות 1.8.
 ה. $A = \bar{\bar{A}}$: גובה מעל 1.7.
- (6) א. $A \cap B \cap C$ ב. $A \cap \bar{B} \cap \bar{C}$ ג. $A \cup B \cup C$ ד. $\bar{C}$ ה. $(A \cap B \cap \bar{C}) \cup (B \cap C \cap \bar{A}) \cup (A \cap C \cap \bar{B})$
- (7) א. $P(A \cup B) = 0.24$ ב. $P(A \cap B) = 0.04$ ג. $P(B \cap \bar{A}) = 0.16$
- (8) א. $P(A \cap B) = 10\%$ ב. $P(A \cup B) = 50\%$ ג. $P(\bar{A} \cap \bar{B}) = 50\%$
- (9) א. $P(A \cap B) = 0.2$ ב. $P(\bar{A} \cap \bar{B}) = 0.3$ ג. $P(A \cup \bar{B}) = 0.3$
- (10) א. לא. ב. כן. ג. כן. ד. לא.
- (11) א. כן. ב. $P(\bar{A} \cap B) = 0.3$
- (12) הטענה הנכונה היא ג'.
- (13) א. 0.05. ב. 0.95.
- (14) א. $P(A \cap B) = 0.06$ ב. לא. ג. $P((A \cap \bar{B}) \cup (B \cap \bar{A})) = 0.43$
- (15) $P(B) = \frac{1}{5}, P(A) = \frac{2}{5}$
- (16) א. נכון. ב. לא נכון. ג. לא נכון. ד. נכון.
- (17) א. כן. ב. לא. ג. $P(A \cup B) = 0.5$ ד. $P(A \cup B) = 0.3$
- (18) א. 19%. ב. 0.05. ג. 0.31. ד. 46%. ה. 12%. ו. 0.41.
- (19) א. 5%. ב. 10%. ג. 67%.
- (20) שאלת הוכחה.
- (21) נכון.

סטטיסטיקה

פרק 33 - הסתברות מותנית-במרחב מדגם אחד

תוכן העניינים

1. כללי 195

הסתברות מותנית – במרחב מדגם אחיד:

רקע:

לעיתים אנו נדרשים לחשב הסתברות למאורע כלשהו כאשר ברשותנו אינפורמציה לגבי מאורע אחר. הסתברות מותנית הינה סיכוי להתרחשות מאורע כלשהו כאשר ידוע שמאורע אחר התרחש/ לא התרחש.

ההסתברות של A בהינתן ש- B כבר קרה: $P(A|B)$.

כשמרחב המדגם אחיד: $P(A|B) = \frac{|A \cap B|}{|B|}$

דוגמה (פתרון בהקלטה):

נטיל קובייה.

נגדיר:

A - התוצאה זוגית.

B - התוצאה גדולה מ-3.

נרצה לחשב את: $P(A|B)$.

שאלות:

- (1) נבחרה ספרה זוגית באקראי. מה הסיכוי שהספרה גדולה מ-6?
- (2) יוסי הטיל קובייה.
מה הסיכוי שקיבל את התוצאה 4, אם ידוע שהתוצאה שהתקבלה זוגית?
- (3) הוטלו צמד קוביות. נגדיר:
A - סכום התוצאות בשתי ההטלות הינו 7.
B - מכפלת התוצאות 12.
חשבו את $P(A|B)$.
- (4) מטבע הוטל פעמיים.
ידוע שהתקבל לכל היותר ראש אחד, מה הסיכוי שהתקבלו שני ראשים?
- (5) זוג קוביות הוטלו והתקבל שהתוצאות זהות.
מה הסיכוי שלפחות אחת התוצאות 5?
- (6) זוג קוביות הוטלו והתקבל לפחות פעם אחת 4.
מה הסיכוי שאחת התוצאות 5?
- (7) נבחרה משפחה בת שני ילדים, שמהם אחד הוא בן.
מה ההסתברות שבמשפחה שני בנים בקרב הילדים?

תשובות סופיות:

(1) 0.2

(2) $\frac{1}{3}$

(3) 0.5

(4) 0

(5) $\frac{1}{6}$

(6) $\frac{2}{11}$

(7) $\frac{1}{3}$

סטטיסטיקה

פרק 34 - הסתברות מותנית - מרחב לא אחיד

תוכן העניינים

1. כללי 198

הסתברות מותנית – מרחב לא אחיד:

רקע:

הסיכוי שמאורע A יתרחש, בהינתן שמאורע B כבר קרה: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

במונה: הסיכוי לחיתוך של שני המאורעות, זה הנשאל וזה הנתון שהתרחש.
במכנה: הסיכוי למאורע נתון שהתרחש.

דוגמה (פתרון בהקלטה):

נבחרו משפחות שיש להם שתי מכוניות. ל-30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן, בקרב 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית. אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?

שאלות:

- (1) תלמיד ניגש בסמסטר לשני מבחנים: מבחן בכלכלה ומבחן בסטטיסטיקה. נגדיר את המאורעות הבאים:
- A - לעבור את המבחן בסטטיסטיקה.
 - B - לעבור את המבחן בכלכלה.
- כמו כן נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8, הסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9 והסיכוי לעבור את שני המבחנים הנו 0.75. חשבו את הסיכויים למאורעות הבאים:
- א. התלמיד עבר בסטטיסטיקה, מה ההסתברות שהוא עבר בכלכלה?
 - ב. התלמיד עבר בכלכלה, מה ההסתברות שהוא עבר בסטטיסטיקה?
 - ג. התלמיד עבר בכלכלה, מה ההסתברות שהוא נכשל בסטטיסטיקה?
 - ד. התלמיד נכשל בסטטיסטיקה, מה ההסתברות שהוא נכשל בכלכלה?
 - ה. התלמיד עבר לפחות מבחן אחד, מה ההסתברות שהוא יעבור את שניהם?
- (2) במדינה שתי חברות טלפון סלולארי: "סופט" ו"בל". 30% מהתושבים הבוגרים רשומים אצל חברת "בל", 60% מהתושבים הבוגרים רשומים אצל חברת "סופט" ול-15% מהתושבים הבוגרים אין טלפון סלולארי כלל.
- א. איזה אחוז מהתושבים הבוגרים רשומים אצל שתי החברות?
 - ב. נבחר אדם שרשום אצל חברת "סופט", מה ההסתברות שהוא רשום גם אצל חברת "בל"?
 - ג. אם אדם לא רשום אצל חברת "בל", מה ההסתברות שהוא כן רשום בחברת "סופט"?
 - ד. אם אדם רשום אצל חברה אחת בלבד, מה ההסתברות שהוא רשום בחברת "סופט"?
- (3) במכללה שני חניונים: חניון קטן וחניון גדול. בשעה 08:00 יש סיכוי של 60% שבחניון הגדול יש מקום, סיכוי של 30% שבחניון הקטן יש מקום וסיכוי של 20% שבשני החניונים יש מקום.
- א. מה ההסתברות שיש מקום בשעה 08:00 רק בחניון הגדול של המכללה?
 - ב. ידוע שבחניון הקטן יש מקום בשעה 08:00, מה הסיכוי שבחניון הגדול יש מקום?
 - ג. אם בשעה 08:00 בחניון הגדול אין מקום, מה ההסתברות שבחניון הקטן יהיה מקום?
 - ד. נתון שלפחות באחד מהחניונים יש מקום בשעה 08:00, מה ההסתברות שבחניון הגדול יש מקום?

(4) נלקחו 200 שכירים ו-100 עצמאים. מתוך השכירים 20 הם אקדמאיים, ומתוך העצמאיים 30 הם אקדמאיים.

- א. בנו טבלת שכיחות משותפת לנתונים.
- ב. נבחר אדם אקראי מה ההסתברות שהוא שכיר?
- ג. מה ההסתברות שהוא שכיר ולא אקדמאי?
- ד. מה ההסתברות שהוא שכיר או אקדמאי?
- ה. אם האדם שנבחר הוא עצמאי מהי ההסתברות שהוא אקדמאי?
- ו. אם האדם שנבחר הוא לא אקדמאי, מה ההסתברות שהוא שכיר?

(5) חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21 :
 40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראלכרט",
 20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים ויזה וגם ישראלכרט,
 8% מחזיקים ישראלכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם
 אמריקן אקספרס. כמו כן, 5% מחזיקים בשלושת הכרטיסים הנ"ל.
 א. אם לאדם יש ויזה, מה הסיכוי שאין לו ישראלכרט?
 ב. אם לאדם שני כרטיסי אשראי, מה הסיכוי שאין לו ישראלכרט?
 ג. אם לאדם לפחות כרטיס אחד, מה הסיכוי שאין לו ישראלכרט?

תשובות סופיות:

(1) א. 0.833 ב. 0.9375 ג. 0.0625 ד. 0.5 ה. 0.789

(2) א. 5% ב. 0.0833 ג. 0.786 ד. 0.6875

(3) א. 0.4 ב. $\frac{2}{3}$ ג. 0.25 ד. $\frac{6}{7}$

(4) א. להלן טבלה: ב. $\frac{2}{3}$ ג. 0.6 ד. $\frac{23}{30}$

סה"כ	לא אקדמאי	אקדמאי	
200	180	20	שכיר
100	70	30	עצמאי
300	250	50	סה"כ

ה. 0.3 ו. 0.72

(5) א. 0.625 ב. 0.133 ג. 0.402

סטטיסטיקה

פרק 35 - דיאגרמת עצים - נוסחת בייס ונוסחת ההסתברות השלמה

תוכן העניינים

1. כללי 202

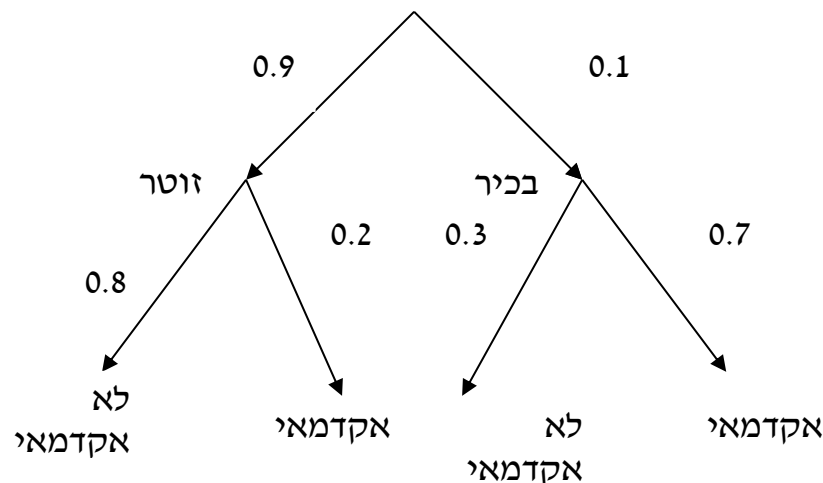
דיאגרמת עצים – נוסחת בייס ונוסחת ההסתברות השלמה:

רקע:

נשתמש בשיטה זו כאשר יש תרגיל שבו התרחשות המאורעות היא בשלבים, כך שכל תוצאה של כל שלב תלויה בשלב הקודם, פרט לשלב הראשון:

דוגמה:

בחברה מסוימת 10% מוגדרים בכירים והיתר מוגדרים זוטרים. מבין הבכירים 20% הם אקדמאים ומבין הזוטרים 80% הם אקדמאים. נשרטט עץ שיתאר את הנתונים, השלב הראשון של העץ אינו מותנה בכלום ואילו השלב השני מותנה בשלב הראשון.



כדי לקבל את הסיכוי לענף מסוים נכפיל את כל ההסתברויות על אותו ענף. נבחר אדם באקראי מאותה חברה.

(1) מה הסיכוי שהוא בכיר אקדמאי ? $0.1 \cdot 0.7 = 0.07$.

(2) מה הסיכוי שהוא זוטר לא אקדמאי ? $0.9 \cdot 0.8 = 0.72$.

כדי לקבל את הסיכוי לכמה ענפים נחבר את הסיכויים של כל ענף (רק אחרי שבתוך הענף הכפלנו את ההסתברויות).

(3) מה הסיכוי שהוא אקדמאי ? $0.1 \cdot 0.7 + 0.9 \cdot 0.2 = 0.25$.

(4) נבחר אקדמאי מה ההסתברות שהוא עובד זוטר? מדובר כאן על שאלה בהסתברות מותנה ולכן נשתמש בעיקרון של הסתברות

$$P(zutar | academay) = \frac{0.9 \cdot 0.2}{0.25} = \frac{0.18}{0.25} = 0.72 \text{ מותנה :}$$

נוסחת ההסתברות השלמה:

בהינתן B , מאורע כלשהו, וחלוקה של מרחב המדגם Ω ל- $A_1, \dots, A_n$ כך ש- $\bigcup_i A_i = \Omega$,

$$. P(B) = \sum_{i=1}^n P(A_i) \cdot P\left(\frac{B}{A_i}\right) \quad \text{אזי:}$$

נוסחת בייס:

$$. P\left(\frac{A_j}{B}\right) = \frac{P(A_j) P\left(\frac{B}{A_j}\right)}{\sum_{i=1}^n P(A_i) \cdot P\left(\frac{B}{A_i}\right)}$$

שאלות:

- (1) בשקית סוכריות 4 סוכריות תות ו-3 לימון. מוציאים באקראי סוכריה. אם היא בטעם תות אוכלים אותה ומוציאים סוכריה נוספת, ואם היא בטעם לימון מחזירים אותה לשקית ומוציאים סוכריה נוספת.
 - א. מה ההסתברות שהסוכריה הראשונה שהוצאה בטעם תות והשנייה בטעם לימון?
 - ב. מה ההסתברות שהסוכריה השנייה בטעם לימון?
- (2) באוכלוסייה מסוימת 30% הם ילדים, 50% בוגרים והיתר קשישים. לפי נתוני משרד הבריאות הסיכוי שילד יחלה בשפעת במשך החורף הוא 80%, הסיכוי שמבוגר יחלה בשפעת במשך החורף הוא 40% והסיכוי שקשיש יחלה בשפעת במשך החורף הוא 70%.
 - א. איזה אחוז מהאוכלוסייה הינו קשישים שלא יחלו בשפעת במשך החורף?
 - ב. מה אחוז האנשים שיחלו בשפעת במשך החורף?
 - ג. נבחר אדם שחלה במשך החורף בשפעת, מה ההסתברות שהוא קשיש?
 - ד. נבחר ילד, מה ההסתברות שהוא לא יחלה בשפעת במשך החורף?
- (3) בכד א' 5 כדורים כחולים ו-5 כדורים אדומים. בכד ב' 6 כדורים כחולים ו-4 כדורים אדומים. בוחרים באקראי כד, מוציאים ממנו כדור ומבלי להחזירו מוציאים כדור נוסף.
 - א. מה ההסתברות ששני הכדורים שיוצאו יהיו בצבעים שונים?
 - ב. אם הכדורים שהוצאו הם בצבעים שונים, מה ההסתברות שהכדור השני שהוצא יהיה בצבע אדום?
- (4) חברת סלולר מסווגת את לקוחותיה לפי 3 קבוצות גיל: נוער, בוגרים ופנסיונרים. נתון כי: 10% מהלקוחות בני נוער, 70% מהלקוחות בוגרים והיתר פנסיונרים. מתוך בני הנוער 90% מחזיקים בסמארט-פון, מתוך האוכלוסייה הבוגרת ל-70% יש סמארט-פון ומתוך אוכלוסיית הפנסיונרים 30% מחזיקים בסמארט-פון.
 - א. איזה אחוז מלקוחות החברה הם בני נוער עם סמארט-פון?
 - ב. נבחר לקוח אקראי ונתון שיש לו סמארט-פון. מה ההסתברות שהוא פנסיונר?
 - ג. אם ללקוח אין סמארט-פון, מה ההסתברות שהוא לא בן נוער?

- (5) כדי להתקבל למקום עבודה יש לעבור שלושה מבחנים. המבחנים הם בשלבים, כלומר לאחר כישלון במבחן מסוים אין אפשרות לגשת למבחן הבא אחריו. 70% מהמועמדים עוברים את המבחן הראשון. מתוכם, 50% עוברים את המבחן השני. מבין אלה שעוברים את המבחן השני 40% עוברים את המבחן השלישי.
- א. מה ההסתברות להתקבל לעבודה?
 ב. מועמד לא התקבל לעבודה. מה ההסתברות שהוא נכשל במבחן הראשון?
 ג. מועמד לא התקבל לעבודה. מה ההסתברות שהוא עבר את המבחן השני?
- (6) משרד הבריאות פרסם את הנתונים הבאים:
- מתוך אוכלוסיית הילדים והנוער 80% חולים בשפעת בזמן החורף.
 מתוך אוכלוסיית המבוגרים (עד גיל 65) 60% חולים בשפעת בזמן החורף.
 30% מהתושבים הם ילדים ונוער. 50% הם מבוגרים. היתר קשישים.
 כמו כן נתון ש 68% מהאוכלוסייה תחלה בשפעת בחורף.
- א. מה אחוז החולים בשפעת בקרב האוכלוסייה הקשישה?
 ב. נבחר אדם שלא חלה בשפעת, מה ההסתברות שהוא לא קשיש?
- (7) רדאר שנמצא על החוף צריך לקלוט אנייה הנמצאת ב-1 מ-4 האזורים: A, B, C, D. אם האנייה נמצאת באזור A הרדאר מזהה אותה בסיכוי 0.8, סיכוי זה פוחת ב-0.1 ככל שהאנייה מתקדמת באזור. כמו כן נתון שבהסתברות חצי האנייה נמצאת באזור D, בהסתברות 0.3 באזור C, באזור B היא נמצאת בסיכוי 0.2, אחרת היא נמצאת באזור A.
- א. מה הסיכוי שהאנייה תתגלה ע"י הרדאר?
 ב. אם האנייה התגלתה ע"י הרדאר, מה ההסתברות שהיא נמצאת באזור C?
 ג. אם האנייה התגלתה ע"י הרדאר, מה הסיכוי שהיא לא נמצאת באזור B?
- (8) סימפטום X מופיע בהסתברות של 0.4 במחלה A, בהסתברות של 0.6 במחלה B ובהסתברות של 0.5 במחלה C. סימפטום X מופיע אך ורק במחלות הללו, אדם לא יכול לחלות ביותר ממחלה אחת מבין המחלות הללו. לקליניקה מגיעים אנשים כדלקמן: 8% חולים במחלה A, 10% במחלה B, 2% במחלה C והיתר בריאים. כמו כן נתון שבמחלה A, סימפטום X מתגלה בסיכוי של 80%, ובמחלות B, C הסימפטום מתגלה בסיכוי של 90% בכל מחלה.
- א. מה ההסתברות שאדם הגיע לקליניקה וגילו אצלו את סימפטום X?
 ב. אם התגלה אצל אדם סימפטום X, מה ההסתברות שהוא חולה במחלה A?
 ג. אם לאדם יש את סימפטום X, מה ההסתברות שהוא חולה במחלה A?
 ד. אם לא גילו אצל אדם את סימפטום X, מה ההסתברות שהוא בריא?

- (9) סטודנט ניגש למבחן אמריקאי. הסיכוי שהוא יודע תשובה לשאלה מסוימת הוא P , ואם הוא לא יודע את התשובה הוא מנחש. בכל מקרה הוא עונה על השאלה. נתון שלשאלה יש k תשובות אפשריות. אם הסטודנט ענה נכון על השאלה, מה הסיכוי שהוא ידע אותה?
- (10) אדם משחק נגד שני מתמודדים, רונית ודולב. האדם צריך לשחק שלושה משחקים ויש לו לבחור איזה סדר משחקים עדיף לו:
- דולב, רונית, דולב.
 - רונית, דולב, רונית.
- בכל משחק מישו חייב לנצח(אין תיקו). האדם ינצח בטורניר רק אם ינצח בשני משחקים ברציפות. נתון שדולב שחקן טוב יותר מאשר רונית. איזו אפשרות עדיפה יותר על האדם כדי לנצח בטורניר?

תשובות סופיות:

- (1) א. $\frac{2}{7}$ ב. $\frac{23}{49}$
- (2) א. 6% ב. 58% ג. 0.241 ד. 0.2
- (3) א. 0.544 ב. 0.5
- (4) א. 9% ב. 0.09375 ג. 0.9722
- (5) א. 0.14 ב. 0.3488 ג. 0.2442
- (6) א. 70% ב. 0.8125
- (7) א. 0.57 ב. 0.3158 ג. 0.7543
- (8) א. 0.0886 ב. 0.2889 ג. 0.3137 ד. 0.8778
- (9) $\frac{kp}{1+p(k-1)}$
- (10) א'.

סטטיסטיקה

פרק 36 - תלות ואי תלות בין מאורעות

תוכן העניינים

1. אי תלות בין מאורעות (מורחב) 207

תלות ואי תלות בין מאורעות:

רקע:

אם מתקיים ש: $P(B|A) = P(B)$, נגיד שמאורע B בלתי תלוי ב- A .
 הדבר גורר גם ההפך: $P(A|B) = P(A)$, כלומר, גם A אינו תלוי ב- B .
 כשהמאורעות בלתי תלויים מתקיים ש: $P(A \cap B) = P(A) \cdot P(B)$.
 הוכחה לכך: $P(A/B) = P(A) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A) \cdot P(B)$.

נשתמש בנוסחאות של מאורעות בלתי תלויים רק אם נאמר במפורש שהמאורעות בלתי תלויים בתרגיל או שמההקשר אפשר להבין ללא צל של ספק שהמאורעות בלתי תלויים.

למשל,

חוקר מבצע שני ניסויים בלתי תלויים הסיכוי להצליח בניסוי הראשון הנו 0.7 והסיכוי להצליח בניסוי השני הוא 0.4.

א. מה הסיכוי להצליח בשני הניסויים יחדו?

כיוון שהמאורעות הללו בלתי תלויים:

$$P(A \cap B) = P(A) \cdot P(B) = 0.7 \cdot 0.4 = 0.28$$

ב. מה הסיכוי להיכשל בשני הניסויים?

באופן דומה:

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = (1 - 0.7)(1 - 0.4) = 0.18$$

הרחבה: אי תלות בין n מאורעות:

n מאורעות $A_1, \dots, A_n$ הם בלתי תלויים אם ורק אם: $P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$.

שאלות:

- (1) נתון: $P(A) = 0.2$, $P(B) = 0.5$, $P(A \cup B) = 0.6$. האם המאורעות הללו בלתי תלויים?
- (2) תלמיד ניגש לשני מבחנים שהצלחתם לא תלויה זו בזו. הסיכוי שלו להצליח במבחן הראשון הוא 0.7 והשני 0.4.
א. מה הסיכוי להצליח בשני המבחנים יחד?
ב. מה הסיכוי שניכשל בשני המבחנים?
- (3) במדינה מסוימת יש 8% אבטלה, נבחרו באקראי שני אנשים מהמדינה.
א. מה ההסתברות ששניהם מובטלים?
ב. מה ההסתברות שלפחות אחד מהם מובטל?
- (4) מוצר צריך לעבור בהצלחה ארבע בדיקות בלתי תלויות לפני שיווק, אחרת הוא נפסל ולא יוצא לשוק. הסיכוי לעבור בהצלחה כל אחת מהבדיקות הוא 0.8. בכל מקרה מבוצעות כל 4 הבדיקות.
א. מה הסיכוי שהמוצר יפסל?
ב. מה ההסתברות שהמוצר יעבור בהצלחה לפחות בדיקה אחת?
- (5) במדינה מסוימת יש 8% אבטלה, נבחרו באקראי חמישה אנשים מהמדינה.
א. מה ההסתברות שכולם מובטלים במדגם?
ב. מה ההסתברות שלפחות אחד מהם מובטל?
- (6) עבור שני מאורעות A ו- B המוגדרים על אותו מרחב מדגם נתון ש:
 $P(A|B) = 0.6$, $P(A \cap \bar{B}) = 0.3$, $P(A \cup B) = 0.9$.
האם A ו- B מאורעות בלתי תלויים?
- (7) הוכיח שאם: $P(A/B) = P(B/A)$, אז: $P(A) = P(B)$.

(8) קבעו אילו מהטענות הבאות נכונות. נמק!

- א. אם : $P(A \cup B) = P(A) \cdot P(B)$, אזי המאורעות בלתי תלויים.
 ב. מאורע A כלול במאורע B : $0 < P(B) < 1$, $P(A) > 0$, לכן : $P(A/B) < P(A)$.
 ג. A ו- B מאורעות זרים שסיכוייהם חיוביים לכן הם מאורעות תלויים.
 ד. A ו- B מאורעות תלויים שסיכוייהם חיוביים לכן A ו- B מאורעות זרים.
 ה. $\bar{P}(\bar{A} \cap \bar{B}) = 1 - P(A) - P(B)$ לכן A ו- B מאורעות זרים.

(9) זוג מעוניין להביא ילד לעולם, בבדיקה גנטית שהם עשו לאב התגלה שהאב אינו נשא של מחלה Q . מערכים את הסיכוי של האם להיות נשאית למחלץ Q להיות 0.2. אם האם נשאית היא תלד בכל פעם ילד חולה בסיכוי 0.5 באופן בלתי תלוי בין הלידות. האם ילדה שני ילדים. האם המאורעות "הילד הראשון בריא" ו-"הילד השני בריא" הם מאורעות בלתי תלויים?

(10) מטילים פעמיים מטבע עם הסתברות p לעץ בכל הטלה, $0 < p < 1$.

A – יצא עץ בהטלה ראשונה.

B – יצאו תוצאות שונות.

עבור איזה ערכים של p המאורעות A ו- B בלתי תלויים?

(11) הוכח אם A ו- B בלתי תלויים, אזי $\bar{A}$, $\bar{B}$ בלתי תלויים.

(12) נתונה מערכת חשמלית שבשרטוט, כל מתג יכול להיות פתוח או סגור

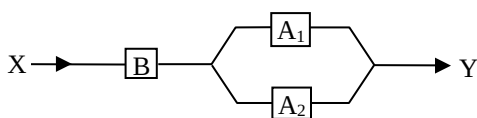
בהסתברויות שונות, אך באופן בלתי תלוי זה בזה.

להלן ההסתברות של כל מתג להיות סגור :

$$P(A_1) = 0.7, P(A_2) = 0.8, P(B) = 0.9$$

א. מה ההסתברות שיעבור זרם במערכת החשמלית?

ב. אם לא עובר זרם במערכת, מה הסיכוי שמתג B סגור?



(13) מטילים שתי קוביות הוגנות. נגדיר שלושה מאורעות :

A – תוצאה של קובייה ראשונה זוגית.

B – תוצאה של קובייה שניה אי זוגית.

C – סכום התוצאות של שתי הקוביות זוגי.

האם המאורעות בלתי-תלויים?

(14) ענה על הסעיפים הבאים :

א. המאורעות A ו- B הם מאורעות זרים של ניסוי כלשהו.
 חוזרים על אותו ניסוי שוב באופן בלתי תלוי זה בזה.
 הוכיחו שהסיכוי שמאורע A יתרחש בניסוי לפני שמאורע B יתרחש

$$\frac{P(A)}{P(A) + P(B)}$$

בניסוי הוא :

ב. מטילים קובייה הוגנת פעם אחרי פעם, מה הסיכוי שתתקבל תוצאה זוגית עוד לפני שתתקבל התוצאה 3?
 ג. מטילים קובייה הוגנת פעם אחרי פעם, מה הסיכוי שתתקבל תוצאה זוגית עוד לפני שתתקבל תוצאה גדולה מ-4?

תשובות סופיות:

- (1) כן.
- (2) א. 0.28 ב. 0.18
- (3) א. 0.0064 ב. 0.1536
- (4) א. 0.5904 ב. 0.9984
- (5) א. 0.08^5 ב. 0.3409
- (6) לא, הם תלויים.
- (7) שאלת הוכחה.
- (8) א. לא נכון. ב. לא נכון. ג. נכון. ד. לא נכון. ה. נכון.
- (9) תלויים.
- (10) 0.5
- (11) שאלת הוכחה.
- (12) א. 0.846 ב. 0.3506
- (13) תלויים.
- (14) א. שאלת הוכחה. ב. $\frac{3}{4}$ ג. $\frac{1}{2}$

סטטיסטיקה

פרק 37 - שאלות מסכמות בהסתברות

תוכן העניינים

1. כללי 211

שאלות מסכמות בהסתברות:

שאלות:

- (1) נלקחו משפחות שיש להם שתי מכוניות. ל-30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית.

 - א. מה ההסתברות שמשפחה אקראית בת שתי מכוניות תהיה ללא מכוניות מתוצרת אירופה?
 - ב. מה ההסתברות שלפחות מכונית אחת תהיה אירופאית?
 - ג. ידוע שלמשפחה יש מכונית אירופאית.
 - ד. אם ההסתברות שרק המכונית החדשה שלה היא מתוצרת אירופאית?
 - ה. אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?
- (2) במדינת "שומקום" 50% מהחלב במרכולים מיוצר במחלבה א', 40% במחלבה ב' והיתר במחלבה ג'. 3% מתוצרת מחלבה א' מגיעה חמוצה למרכולים ואילו במחלבה ב' 10%.

כמו כן ידוע שבמדינת "שומקום" בסך הכול 7.5% מהחלב חמוץ.

 - א. איזה אחוז מהחלב שמגיע למרכול ממחלבה ג' חמוץ?
 - ב. אם נרכש חלב חמוץ במרכול. מה הסיכוי שהוא יוצר במחלבה ג'?
 - ג. ברכישת חלב נמצא שהוא אינו חמוץ. מה הסיכוי שהוא יוצר במחלבה א'?
 - ד. האם המאורעות: "חלב חמוץ" ו-"יוצר במחלבה א'" בלתי תלויים?
- (3) רוני ורונה יצאו לבלות במרכז בילויים עם מספר אפשרויות בילוי: בהסתברות של 0.3 הם ייצאו לבאולינג, בהסתברות של 0.5 הם ייצאו לבית קפה ובהסתברות של 0.7 הם ייצאו לפחות לאחד מהם (באולינג/קפה).

 - א. מה ההסתברות שהם ייצאו רק לבאולינג?
 - ב. האם המאורעות "לצאת לבאולינג" ו-"לצאת לבית קפה" זרים?
 - ג. האם המאורעות "לצאת לבאולינג" ו-"לצאת לבית קפה" תלויים?
 - ד. מה ההסתברות שיום אחד הם ייצאו רק לבאולינג וביום למחרת לא ייצאו לאף אחד מהמקומות?

(4) 70% מהנבחנים בסטטיסטיקה עוברים את מועד א'. כל מי שלא עובר את מועד א' ניגש לעשות מועד ב', מתוכם 80% עוברים אותו. מבין אלה שנכשלים בשני המועדים 50% נרשמים לקורס מחדש, והיתר פורשים מהתואר.

- מה הסיכוי שסטודנט אקראי עבר את הקורס?
- אם סטודנט אקראי עבר הקורס, מה הסיכוי שעבר במועד ב'?
- מה אחוז הסטודנטים שפורשים מהתואר?
- נבחרו 2 סטודנטים אקראיים רונית וינאי, מה ההסתברות שרונית עברה במועד א' ושינאי עבר במועד ב'?

(5) באוכלוסייה מסוימת 40% הם גברים והיתר הן נשים. מבין הגברים 10% מובטלים. בסך הכול 13% מהאוכלוסייה מובטלת.

- מה אחוז האבטלה בקרב הנשים?
- נבחר אדם מובטל, מה ההסתברות שזו אישה?
- נגדיר את המאורעות הבאים: A - נבחר אדם מובטל, B - נבחר גבר. האם המאורעות הללו זרים? והאם הם בלתי תלויים?

(6) בתיבה 10 מטבעות, מתוכם 7 מטבעות רגילים (ראש, זנב) ו-3 מטבעות שבשני צדדיהם טבוע ראש. אדם בוחר באקראי מטבע ומטיל אותו פעמיים. נסמן ב-A את ההטלה הראשונה ראש וב-B את ההטלה השנייה ראש.

- חשבו את הסיכויים למאורעות A ו-B.
- האם המאורע A ו-B בלתי תלויים?
- ידוע שבהטלה הראשונה התקבל ראש, מה ההסתברות שהמטבע שהוטל הוא מטבע הוגן?

(7) ערן מעוניין למכור את רכבו והוא מפרסם מודעה באינטרנט ומודעה בעיתון. מבין אלה שמעוניינים לרכוש רכב משומש 30% יראו את המודעה באינטרנט, 50% יראו את המודעה בעיתון ו-72% יראו את המודעה בלפחות אחת מהמדיות.

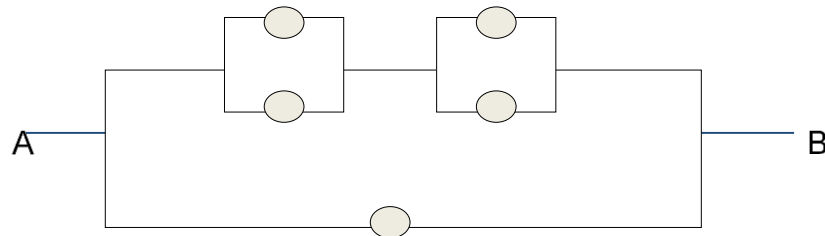
- מה אחוז האנשים, מאלה שמעוניינים לרכוש רכב משומש, שיראו את 2 המודעות?
- אם אדם ראה את המודעה באינטרנט, מה ההסתברות שהוא לא ראה את המודעה בעיתון?

ג. האם המאורעות: "לראות את המודעה באינטרנט" ו-"לראות את המודעה בעיתון" בלתי תלויים?

ד. אדם שראה את המודעה באינטרנט בלבד יתקשר לערן בהסתברות של 0.7, אם הוא ראה את המודעה בעיתון בלבד הוא יתקשר לערן בהסתברות של 0.6 ואם הוא ראה את שתי המודעות הוא יתקשר לערן בהסתברות של 0.9.

- מה ההסתברות שאדם המעוניין לרכוש רכב משומש יתקשר לערן?
- אדם המעוניין לרכוש רכב משומש התקשר לערן. מה ההסתברות שהוא ראה את שתי המודעות?

(8) נתונה המערכת החשמלית הבאה :



כל יחידה עובדת באופן בלתי תלוי ובהסתברות p .
 כדי שהמערכת תפעל צריך לעבור זרם מהנקודה A לנקודה B.
 הוכיחו שהסיכוי שהמערכת תפעל הוא: $P + (1 - P)(2P - P^2)^2$.

(9) ליאת מעוניינת לתרגל לבחינה בהסתברות. היא מצאה באינטרנט מאגר הכולל 25 שאלות מבחינות. השאלות ממוספרות ו-6 מתוכן עוסקות במשתנה מקרי רציף. ליאת החליטה לבחור באקראי 7 שאלות מהמאגר במטרה לפתור אותן. כל שאלה שלא עוסקת במשתנה הרציף תיפתר על ידי מיכל בסיכוי של 90%, אך אם השאלה עוסקת במשתנה הרציף היא תיפתר בסיכוי של 60%.
 א. מה הסיכוי שהשאלות שנבחרו הן כולן ממוספרות בסדר עוקב?
 ב. מה הסיכוי ששאלה 20 היא השאלה עם המספור המקסימלי מבין השאלות שנבחרו?
 ג. ידוע שליאת בחרה 2 שאלות שעוסקות במשתנה הרציף והיתר לא. מה הסיכוי שתצליח לפתור 6 מתוך השאלות שבחרה?

(10) נתונים שלושה מאורעות: A, B ו-C. ידוע ש: $P(A|B) = 1$, $P(A|C) = 1$.
 תנו דוגמא ספציפית למאורעות: A, B ו-C שבה המאורעות B ו-C תלויים.

(11) הוכיחו או הפריכו (על ידי דוגמה נגדית) את הטענה הבאה:
 אם A ו-B בלתי תלויים, אז A ו- $\bar{B}$ בלתי תלויים.

(12) משחקים משחק מזל פעמיים, כך שבכל משחק בודד יש אפשרות לנצח או להפסיד. הסיכוי לנצח בכל משחק הוא P כאשר: $0 < P < 1$.
 נגדיר את המאורעות הבאים:
 A - תוצאות המשחקים שונות זו מזו.
 B - המשחק הראשון היה ניצחון.
 מה ערכו של P, עבורו A ו-B יהיו מאורעות בלתי תלויים?

(13) טל מניח בשורה N קוביות בצבעים שונים. בין שתי קוביות אקראיות כלשהן ערן מניח מכחול. הוכיחו שהסיכוי שהקובייה הכחולה והאדומה יהיו בצדדים

$$\text{שונים של המכחול הוא : } \frac{N+1}{3(N-1)}.$$

(14) הוכיחו באמצעות אינדוקציה את אי שוויון בול: $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$.

תשובות סופיות:

- (1) א. 0.25 ב. 0.75 ג. 0.6 ד. 0.5
- (2) א. 0.2 ב. 0.267 ג. 0.524 ד. תלויים.
- (3) א. 0.2 ב. אינם זרים. ג. תלויים. ד. 0.06
- (4) א. 0.94 ב. 0.255 ג. 0.03 ד. 0.168
- (5) א. 15% ב. 0.692 ג. לא זרים ותלויים.
- (6) א. 0.65 ב. תלויים. ג. 0.5384
- (7) א. 8% ב. 0.733 ג. תלויים.
- (8) ד. i. 0.478 ד. ii. 0.15 שאלת הוכחה.
- (9) א. $\frac{19}{480,700}$ ב. $\frac{27,132}{480,700}$ ג. 0.4015
- (10) ראו סרטון.
- (11) שאלת הוכחה.
- (12) $\frac{1}{2}$
- (13) שאלת הוכחה.
- (14) שאלת הוכחה.

סטטיסטיקה

פרק 38 - מקדם המתאם (מדד קשר) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלינארי (פירסון) 216
2. חישוב מקדם המתאם הלינארי (פירסון) 227
3. בדיקת השערות על מקדם המתאם הלינארי..... 232
4. בדיקת השערות על מקדם המתאם הלינארי באמצעות טבלה של ערכים קריטיים..... 236

מקדם המתאם (מדד קשר) הלינארי ומובהקותו

מדד הקשר הלינארי (פירסון) – מבוא

מעוניינים לבדוק עד כמה קיים קשר מסוג קשר לינארי (קו ישר) בין שני משתנים. שני המשתנים שאנו בודקים לגביהם קשר צריכים להיות משתנים כמותיים. מבחינת סולמות מדידה כל משתנה נחקר צריך להיות מסולם רווחים או מנה. בדרך כלל המשתנה המוצג כ- Y הוא המשתנה התלוי והמשתנה המוצג ב- X הוא המשתנה הבלתי תלוי. תיאור גרפי לנתונים נעשה על ידי דיאגרמת פיזור. בדיאגרמת פיזור אנחנו מסמנים כל תצפית בנקודה לפי שיעור ה- X ושיעור ה- Y שלה. דיאגרמת הפיזור נותנת אינדיקציה גרפית על הקשר בין שני המשתנים.

דוגמה (פתרון בהקלטה) :

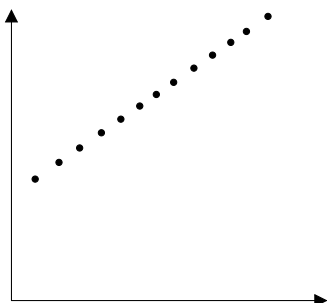
בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

4	4	3	3	2	3	2	2	מספר חדרים בדירה
5	4	4	3	2	2	1	0	מספר הנפשות בדירה

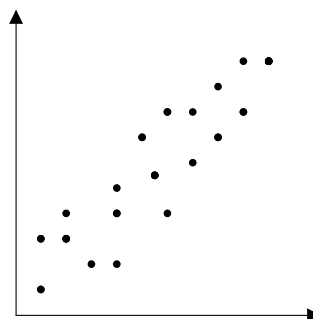
- (1) כמה תצפיות ישנן בדוגמה?
- (2) כמה משתנים ישנם בדוגמה, מי הם?
- (3) שרטטו לנתונים דיאגרמת פיזור.
- (4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

דיאגרמות פיזור לקשר בין משתנים וניתוחם

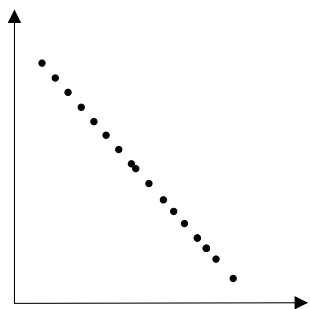
קשר לינארי חיובי מלא



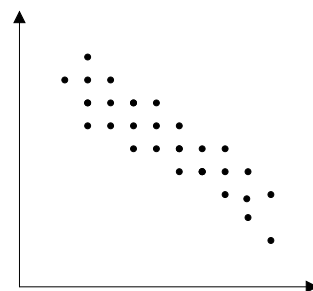
קשר לינארי חיובי חלקי



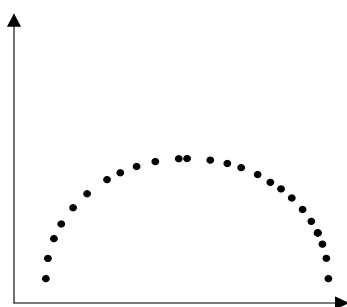
קשר לינארי שלילי מלא



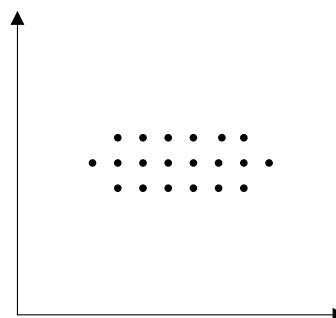
קשר לינארי שלילי חלקי



אין קשר לינארי



אין קשר



משמעות מקדם המתאם:

כדי לבדוק עד כמה קיים קשר לינארי בין שני המשתנים ישנו מדד קשר שנקרא גם מקדם המתאם הלינארי הידוע גם בשם מקדם המתאם של פירסון. מקדם מתאם זה מקבל ערכים בין 1 ל-1.

-1

0

1

מקדם מתאם 1-או 1 אומר שקיים קשר לינארי מלא בין המשתנים שניתן לבטאו על ידי נוסחה של קו ישר: $y = ax + b$.

מתאם חיובי מלא (מקדם מתאם 1):

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

מתאם חיובי חלקי:

ככל 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 שמקדם המתאם קרוב לאפס עוצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לסיכום, מקדם המתאם בודק את עוצמת הקשר הלינארי, ואת כיוון הקשר.

מקדם המתאם הלינארי אינו מושפע מיחידות המדידה. כל שינוי ביחידות המדידה של המשתנים, לא ישנה את מקדם המתאם.

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:

r - מדד הקשר הלינארי במדגם שמהווה אומדן לפרמטר ρ .

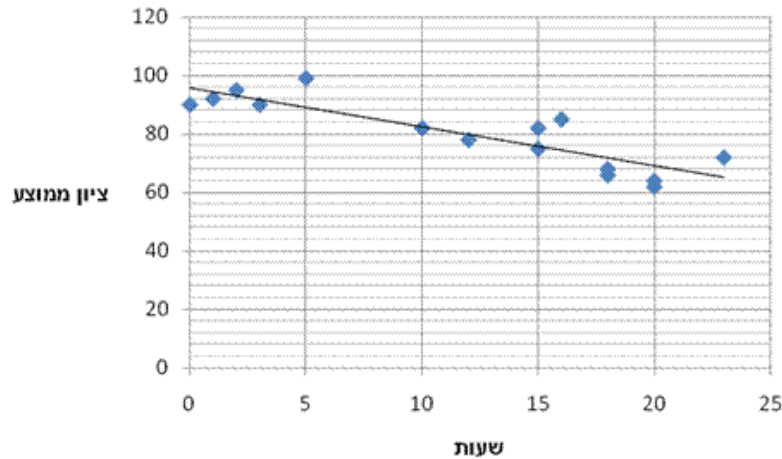
קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את X ב- Y התוצאה תהיה זהה.

דוגמה (פתרון בהקלטה):

- מה ניתן להגיד על מקדם המתאם של שני המשתנים על סמך דיאגרמת הפיזור ששרטטנו?
- אם היינו משנים את השרטוט כך שבציר האנכי היה המשתנה "מספר החדרים" ובציר האופקי היה "מספר הנפשות", האם הדבר היה משפיע על מדד הקשר של פירסון?

שאלות

- (1) 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 דיאגרמת פיזור:



- א. מיהו המשתנה הבלתי תלוי?
 ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?
- (2) להלן טבלה המסכמת את מקדמי המתאם הלינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

מתמטיקה	לשון	ספורט	
?	-0.7	?	ספורט
0.6	?	?	לשון
?	?	-0.1	מתמטיקה

- א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.
 ב. בין אילו שני ציוני מקצועות שונים קיים מתאם בעל העוצמה החזקה ביותר?

- (3) במחקר נתבקשו לבדוק את הקשר בין מספר שעות התרגול של קורס לבין הציון הסופי שלו. להלן תוצאות מדגם שהתקבל:

שעות תרגול	ציון סופי
20	90
25	90
30	95
15	60
30	90
20	85
10	50

- א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי בדוגמה זו?
 ב. שרטטו דיאגרמת פיזור לנתונים.
 ג. מה ניתן לומר על הקשר בין המשתנים במדגם?
 ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

4) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8.

א. השלימו את המשפט:

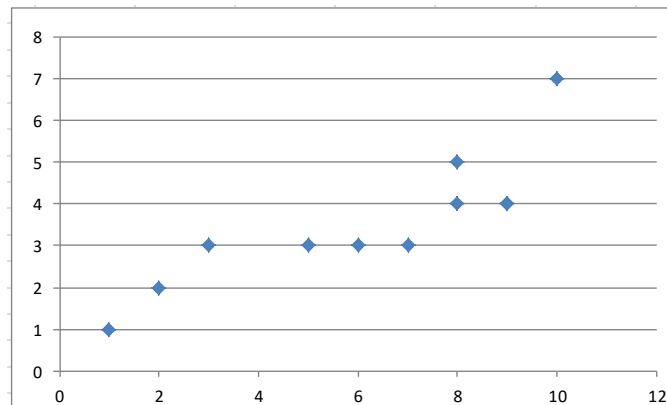
בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה _____.

ב. הוחלט להעביר את הטמפרטורה למעלות פרנהייט על מנת שיוכלו

להשוות אותה לנתונים מארה"ב. נוסחת המעבר היא $F^0 = 32 + \frac{9}{5}C^0$.

כיצד הדבר ישפיע על מקדם המתאם בין הטמפרטורה במעלות פרנהייט לכמות המשקעים במ"מ?

5) להלן דיאגרמת פיזור המראה קשר בין שני משנים:



א. השלימו: ניתן לראות שהקשר הוא לינארי _____ (מלאו חלקי) כיוון הקשר הוא (חיובי/שלילי).

ב. השלימו: אם היינו מוסיפים תצפית שערך ה- X שלה הוא 4 וערך ה- Y שלה הוא 7, מקדם המתאם של פירסון היה _____ (גדלו קטן/לא משתנה).

שאלות רב ברירה (יש לבחור את התשובה הנכונה):

6) חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאם שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאם השלילי במדגם:

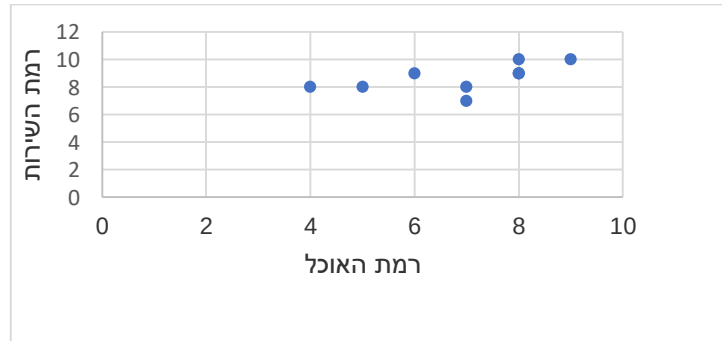
א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.

ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.

ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.

ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.

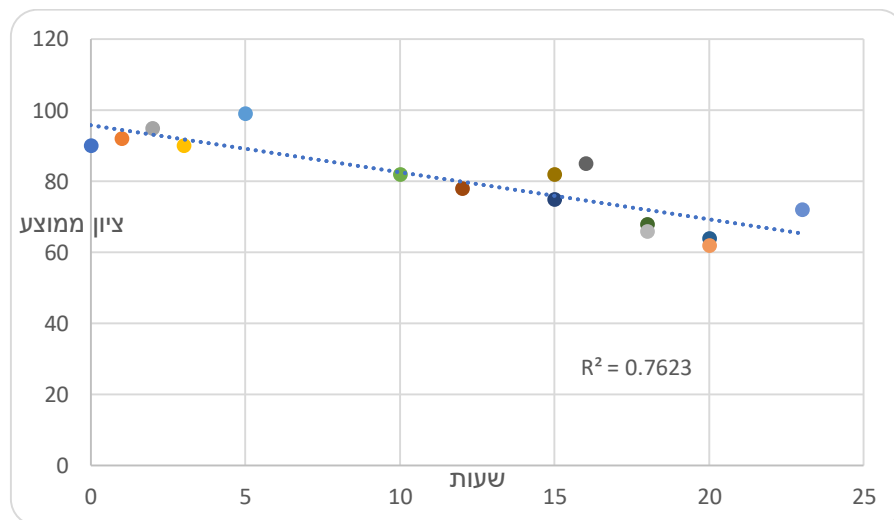
- 7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם (r)?

- א. $r = -0.3$
 ב. $r = 0$
 ג. $r = 1.125$
 ד. $r = 0.593$

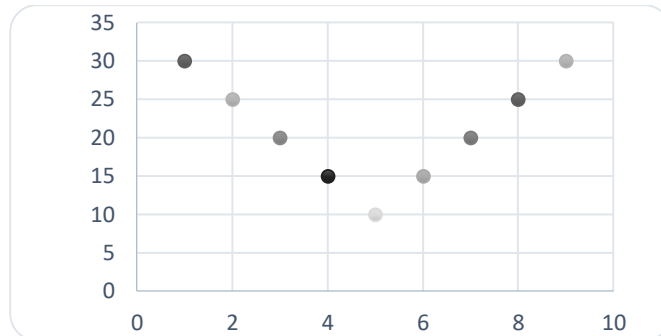
- 8) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.



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

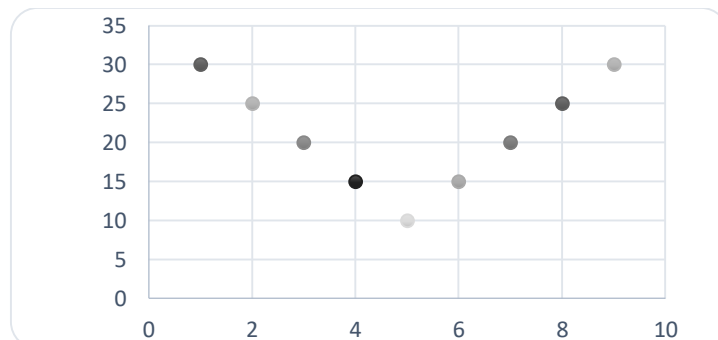
- א. ככל שמבליים יותר הציון נוטה לרדת.
 ב. אין קשר בין שעות הבילוי לציון.
 ג. ככל שמבליים פחות הציון נוטה לרדת.
 ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.

9) התרשים הבא מתאר קשר בין שני משתנים, איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



- א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.
- ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.
- ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.
- ד. $r = \pm 1$ היות ויש קו עולה וגם קו יורד.

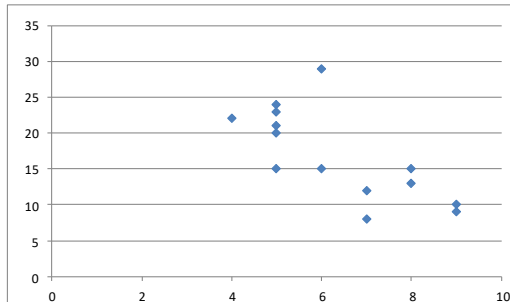
10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

- א. בתרשים מוצג הקשר בין שני משתנים.
- ב. בתרשים מוצג הקשר בין 9 משתנים.
- ג. בתרשים מוצג הקשר בין 10 משתנים.
- ד. אין לדעת כמה משתנים מוצגים בתרשים.

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים :



X - (משתנה בלתי תלוי בציר האופקי)

Y - (משתנה תלוי).

במדגם התקבל $r^2 = 0.52$.

11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של r ?

א. -0.52

ב. 0.72

ג. -0.72

ד. 0.52

12) אם מקדם המתאם בין שני משתנים הוא 1, אזי :

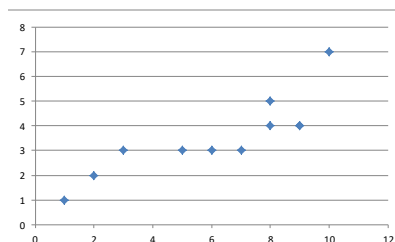
א. הערכים של המשתנים הם חיוביים.

ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.

ג. הקשר הלינארי הוא בעוצמה חזקה.

ד. אף אחת מהתשובות לא בהכרח נכונה.

13) להלן דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים ?

א. 1

ב. 0.85

ג. 0.15

ד. 0

14) בבדיקת קשר בין שני משתנים התקבל : $r = -1$.

א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.

ב. לא קיים קשר בין שני המשתנים.

ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.

ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

15) לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר _____ בין קרבה פיזית לקרבה נפשית.

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

16) מבחן אמי"ר הינו מבחן מיון באנגלית של המרכז הארצי לבחינות והערכה. הציון המינימלי בבחינה הינו 150 והמקסימלי הינו 250. בקורס הכנה למבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פלט שהתקבל:

	159
	170
	180
	185
	204
	224
	236
	212
	168
	189
	195
	163
	187
	206
	201
	223
	242
	203
	205
197.47	AVERAGE
536.25	VARPA

יש להוסיף עמודה נוספת לצד עמודת הציונים שתראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציון המקסימלי בבחינה.

מה יהיה מקדם המתאם בין שתי העמודות (כלומר, מקדם המתאם בין הציון לבין הנקודות החסרות)?

- א. -1
- ב. 1
- ג. -0.5
- ד. 0.5

17) מקדם המתאם בין שטחי דירה למחיר שלהם חושב ונמצא 1.2. מה נובע מכך?

- א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.
- ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.
- ג. לא קיים קשר בין שטח הדירה למחיר הדירה.
- ד. מצב כזה שמתואר הנתונים לא אפשרי.

18) אם ניקח 10 אנשים ונרשום לכל אדם את הגובה במטר וכמו כן את הגובה בס"מ. מה יהיה מקדם המתאם בין גובה האדם במטר לגובה האדם בס"מ?

- א. 1
- ב. 0
- ג. -1
- ד. לא ניתן לדעת.

- 19** נמצא מתאם חיובי בעוצמה גבוהה בין X – ציון בבגרות בלשון ל Y – ציון בבגרות במתמטיקה. אילו מהמשפטים הבאים נכון?
- א. ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.
- ב. קיימת נוסחה של קו ישר שקושרת בין ציון בבגרות במתמטיקה לציון בבגרות בלשון.
- ג. ללא יוצא מן הכלל, ניתן להגיד שכל תלמיד שמצליח יותר מתלמיד אחר בלשון גם יצליח יותר מאותו תלמיד במתמטיקה.
- ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

- 20** עבור סדרה של תצפיות מדדו את X ואת Y . נמצא שעבור כל התצפיות שהערך של Y ירד הערך של X בהכרח ירד ללא יוצא מן הכלל. מקדם המתאם של פירסון יהיה בהכרח :

- א. 1
- ב. -1
- ג. 0
- ד. אף אחת מהתשובות.

תשובות סופיות

- (1) א. שעות בילוי.
ב. הקשר חלקי, כיוון הקשר שלילי.
(2) א. להלן טבלה:
ב. ספורט ולשון.

מתמטיקה	לשון	ספורט	
0.1	-0.7	1	ספורט
0.6	1	-0.7	לשון
1	0.6	-0.1	מתמטיקה

- (3) א. ב"ת- מס' שעות התרגול, תלוי- ציון.
ג. קשר לינארי חיובי חלקי.
(4) א. לעלות.
(5) א. חלקי, חיובי.
ב. קטן.
ב. ראה גרף בפתרון וידאו.
ד. מקדם המתאם לא היה משתנה.
ב. לא ישפיע על מקדם המתאם.

(6) ד' (7) ד' (8) א' (9) ג' (10) א'

(11) ג' (12) ד' (13) ב' (14) א' (15) א'

(16) א' (17) ד' (18) א' (19) ד' (20) ד'

מדדי קשר – מדד הקשר הלינארי (פירסון) – רקע

המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו- Y הוא המשתנה המוסבר (התלוי).

דוגמה:

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.

שלב ראשון: נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים.

דוגמה:

מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

בבניין של 5 דירות בדקו את הנתונים הבאים:
 X - מס' חדרים בדירה. Y - מס' נפשות הגרות בדירה.
 להלן התוצאות שהתקבלו:

נשרטט מנתונים אלה דיאגרמת פיזור (הדיאגרמה המלאה בסרטון). נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן (הדיאגרמות המלאות בסרטון).

שלב שני: מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (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):

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

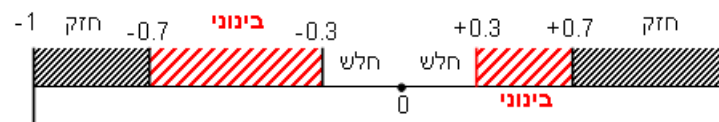
מתאם חיובי חלקי:

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

מתאם שלילי חלקי:

ככל 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.

$$COV(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \frac{\sum xy}{n} - \bar{x} \cdot \bar{y} : \text{שונות משותפת}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 : \text{שונות של המשתנה } X$$

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 : \text{שונות המשתנה } Y$$

$$r_{xy} = \frac{COV(x, y)}{S_x \cdot S_y} : \text{מקדם המתאם הלינארי}$$

שאלות

- 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	2	1	0	2	3	4
ציון	80	90	90	70	70	50

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

X	Y
10	12
14	15
15	15
18	17
20	21

- 2) 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 התוצאות שהתקבלו:
- א. מה הממוצע של כל רמת הורמון?
- ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

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

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

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

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

- א. חשב את מדד הקשר הלינארי בין X ל-Y. מיהו המשתנה התלוי?
- ב. מה המשמעות של התוצאה שקיבלת בסעיף א'?

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

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

$$\sum_{i=1}^{20} Y_i^2 = 2080 \quad \sum_{i=1}^{20} X_i^2 = 2960$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשבו את מדד הקשר הלינארי בין X ל- Y .

- (5) במוסד אקדמי ציון ההתאמה מחושב כך: מכפילים את הציון הממוצע בבגרות ב-3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה 2.
מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם?

- (6) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו.
- 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.
 - לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$, $S_x = S_y = 1$. לכן, מדד הקשר של פירסון יהיה 1.
 - אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

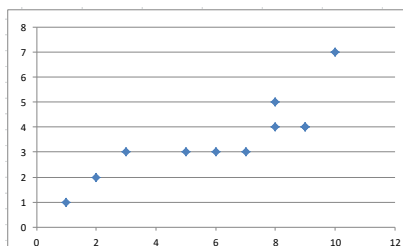
שאלות רב-ברירה:

- (7) נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן:
- הדבר מעיד שהציונים בכיתה היו שליליים.
 - ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
 - ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
 - אף אחת מהתשובות לא נכונה.

8) נלקחו 20 מוצרים ונבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה-4.2ש"ח). מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

- א. 1
- ב. 0
- ג. 4.2
- ד. לא ניתן לדעת.

9) להלן דיאגרמת פיזור: מה יהיה מקדם המתאם בין שני המשתנים?



- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

תשובות סופיות

- 1) א. משתנה תלוי: ציון, משתנה ב"ת: מס' חיסורים. ראה דיאגרמה בוידאו. ניתן להסיק שקיים קשר לינארי שלילי וחלקי בין מספר החיסורים לציון התלמיד.
ב. -0.9325.
ג. הקשר יישאר לינארי שלילי חלקי אך עוצמתו תחלש.
- 2) א. $\bar{y} = 16$, $\bar{x} = 15.4$ ב. $r_{xy} = 0.96$.
- 3) א. 0.8
- 4) 0.8
- 5) 1
- 6) א. נכון. ב. לא נכון. ג. נכון.
- 7) ג'.
- 8) א'.
- 9) ב'.

בדיקת השערות על מקדם המתאם הלינארי – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר: r - מדד הקשר הלינארי במדגם שמהווה אומדן לפרמטר ρ .

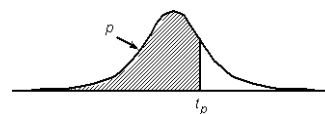
השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים $H_0: \rho = 0$.
ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$$

סטטיסטי המבחן:

סטטיסטי זה מתפלג t עם $n-2$ דרגות חופש.

$H_0: \rho = 0$	$H_0: \rho = 0$	$H_0: \rho = 0$	השערת האפס:
$H_1: \rho > 0$	$H_1: \rho < 0$	$H_1: \rho \neq 0$	השערת המחקר:
$t \geq t_{1-\alpha}$	$t \leq -t_{1-\alpha}$	$t \geq t_{1-\alpha}$ $\vee$ $t \leq -t_{1-\alpha}$	כלל ההכרעה: אזור דחייה של השערת האפס

טבלת ערכים קריטיים של t - נספח: טבלת התפלגות T
 P 

דרגות חופש	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.000	3.078	6.314	12.709	31.821	63.657	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.291

שאלות

(1) להלן נתונים על הוותק בעבודה (בשנים) ועל השכלה (בשנים) במדגם של 10 עובדים:

נחקר	1	2	3	4	5	6	7	8	9	10
X - וותק	13	18	2	8	16	9	5	28	17	24
Y - השכלה	12	14	17	8	11	12	13	8	12	15

מקדם המתאם חושב והתקבל: -0.31 .

- א. האם קיים מתאם בין וותק העובד להשכלתו? בדקו ברמת מובהקות של 5%?
 ב. אם הוותק של העובד היה נמדד בחודשים האם התשובה לסעיף א' הייתה משתנה?

(2) מחקר התעניין לבדוק את הקשר בין גיל נשים בהריון לרמת ההמוגלובין שלהן בדם בזמן הריון. נדגמו 7 נשים והתקבלו התוצאות הבאות:

נחקרת	1	2	3	4	5	6	7
המוגלובין	14.7	13.5	9.7	12	10.8	13	10.3
גיל	39	34	30	29	28	26	23

במדגם חושב מדד הקשר של פירסון להיות 0.7 .

- א. האם ניתן לומר שבמדגם אם אישה היא יותר מבוגרת אזי בהכרח יש לה יותר המוגלובין בדם?
 ב. האם ניתן לומר, ברמת מובהקות של 5%, שקיים מתאם בין גיל האישה שבהריון לבין רמת ההמוגלובין שלה בדם?

(3) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה -0.8 .

- א. בדקו ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במעלות צלזיוס.
 ב. כיצד הייתה משתנה התשובה לסעיף א אם הינו מוסיפים עוד תצפיות למדגם?
 ג. על סמך טבלת T המצורפת עבור אילו רמות מובהקות ניתן להחליט שקיים קשר לינארי שלילי מובהק?

(4) מתווך דירות חישב את מקדם המתאם בין שטח דירה במרכז תל אביב לבין המחיר של הדירה עבור 17 דירות. מקדם המתאם שקיבל היה 0.6 .

- א. בדקו ברמת מובהקות של 5% האם ניתן להגיד שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה במרכז תל אביב?
 ב. מהי מובהקות התוצאה לבדיקת ההשערה שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה בתל אביב.

תשובות סופיות

- | | |
|---------------------------|---------------------------|
| ב. לא תשתנה. | (1) א. לא נדחה את H_0 . |
| ב. לא נדחה את H_0 . | (2) א. לא |
| ב. לא ניתן לדעת. | (3) א. נדחה את H_0 . |
| | ג. לפחות 0.005. |
| ב. $0.005 < P_v < 0.01$. | (4) א. נדחה את H_0 . |

בדיקת השערות על מקדם המתאם הלינארי (באמצעות טבלה של ערכים קריטיים) – רקע

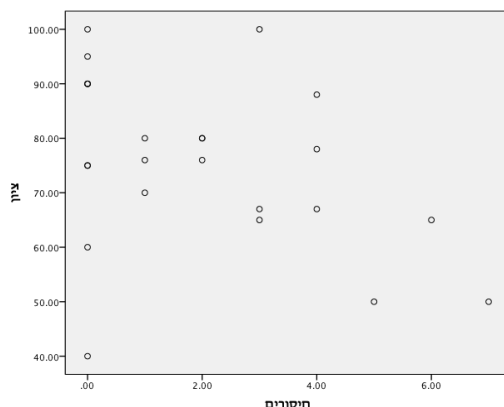
מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר: r - מדד הקשר הלינארי במדגם שמהווה אומדן לפרמטר ρ .

השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים: $H_0: \rho = 0$.
ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו-נורמלית.
את מקדם המדגם הקריטי, שנסמן ב- r_c , נוציא מתוך טבלה של ערכים קריטיים שמצורפת בהמשך.

$H_0: \rho = 0$	$H_0: \rho = 0$	$H_0: \rho = 0$	השערת האפס:
$H_1: \rho > 0$	$H_1: \rho < 0$	$H_1: \rho \neq 0$	השערת המחקר:
$r \geq r_c$	$r \leq -r_c$	$r \geq r_c$ ו $r \leq -r_c$	כלל ההכרעה: אזור דחייה של השערת האפס

דוגמה (פתרון בהקלטה):

הדיקן ביקש לדגום סטודנטים כדי לבדוק את הקשר בין ציון הסטודנט בקורס למספר הפעמים שהוא החסיר שיעור בקורס.
דיאגרמת הפיזור שהתקבלה במדגם שבוצע:



מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי במחקר?
מה ניתן לראות לגבי הקשר הלינארי בין המשתנים שהתקבל במדגם?

חושב האומדן למקדם המתאם הלינארי על סמך 24 הסטודנטים שנדגמו והתקבל: -0.389.

מה משמעות של מקדם המתאם שהתקבל במדגם?

האם ניתן להגיד ברמת מובהקות של 5% שקיים מתאם לינארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?

טבלת ערכים קריטיים של מקדם המתאם הלינארי



0.0005	0.005	0.025	0.05	α / n
0.999	0.990	0.950	0.900	4
0.991	0.959	0.878	0.805	5
0.974	0.917	0.811	0.729	6
0.951	0.875	0.754	0.669	7
0.925	0.834	0.707	0.621	8
0.898	0.798	0.666	0.582	9
0.872	0.765	0.632	0.549	10
0.847	0.735	0.602	0.521	11
0.823	0.708	0.576	0.497	12
0.801	0.684	0.553	0.476	13
0.780	0.661	0.532	0.458	14
0.760	0.641	0.514	0.441	15
0.742	0.623	0.497	0.426	16
0.725	0.606	0.482	0.412	17
0.708	0.590	0.468	0.400	18
0.693	0.575	0.456	0.389	19
0.679	0.561	0.444	0.378	20
0.665	0.549	0.433	0.369	21
0.652	0.537	0.423	0.360	22
0.640	0.526	0.413	0.352	23
0.629	0.515	0.404	0.344	24
0.618	0.505	0.396	0.337	25
0.607	0.496	0.388	0.330	26
0.597	0.487	0.381	0.323	27
0.588	0.479	0.374	0.317	28
0.579	0.471	0.367	0.311	29
0.570	0.463	0.361	0.306	30
0.532	0.430	0.334	0.283	35

שאלות

- (1) להלן נתונים על הוותק בעבודה (בשנים) ועל השכלה (בשנים) במדגם של 10 עובדים:

נחקר	1	2	3	4	5	6	7	8	9	10
X - וותק	13	18	2	8	16	9	5	28	17	24
Y - השכלה	12	14	17	8	11	12	13	8	12	15

מקדם המתאם חושב והתקבל: -0.31 .

- א. האם קיים מתאם בין וותק העובד להשכלתו? בדקו ברמת מובהקות של 5%.
- ב. אם הוותק של העובד היה נמדד בחודשים האם התשובה לסעיף א' הייתה משתנה?

- (2) מחקר התעניין לבדוק את הקשר בין גיל נשים בהריון לרמת ההמוגלובין שלהן בדם בזמן הריון. נדגמו 7 נשים והתקבלו התוצאות הבאות:

נחקרת	1	2	3	4	5	6	7
המוגלובין	14.7	13.5	9.7	12	10.8	13	10.3
גיל	39	34	30	29	28	26	23

במדגם חושב מדד הקשר של פירסון להיות 0.7 .

- א. האם ניתן לומר שבמדגם אם אישה היא יותר מבוגרת אזי היא בהכרח יש לה יותר המוגלובין בדם?
- ב. האם ניתן לומר, ברמת מובהקות של 5%, שהמתאם בין גיל האישה שבהריון לבין רמת ההמוגלובין שלה בדם הוא חיובי?

- (3) בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה -0.8 .

- א. בדוק ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במ"מ?
- ב. כיצד הייתה משתנה התשובה לסעיף א אם הינו מוסיפים עוד תצפיות למדגם?

תשובות סופיות

- (1) א. לא נדחה את H_0 .
ב. לא תשתנה.
- (2) א. לא.
ב. נדחה את H_0 .
- (3) א. נדחה את H_0 .
ב. לא ניתן לדעת.

סטטיסטיקה

פרק 39 - רגרסיה

תוכן העניינים

1. כללי 239

מדדי קשר – רגרסיה ליניארית:

רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבוי. 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 - 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: $Y = bX + a$, $b = r \frac{S_y}{S_x}$.

לצורך בניית קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

שאלות:

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

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

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

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

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

- (2) נסמן ב- X את ההשכלה של אדם בשנות לימוד. נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_x = 2, \quad S_y = 5, \quad \bar{X} = 14, \quad \bar{Y} = 8, \quad \text{COV}(X, Y) = 7.5$$

- א. חשבו את מדד הקשר של פירסון בין ההשכלה להכנסה.
 ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?
 ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000 ₪?

- (3) חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה 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.
 א. על פי משוואת הרגרסיה, שעת הכנה נוספת משפרת את ציון המבחן ב-?
 ב. על פי משוואת הרגרסיה, תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?
 ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

- (4) נתונים 2 משתנים X ו- Y . כמו כן נתון: $\bar{X} = 1.5, S_x = S_y = 4$,
 וכן שקו הרגרסיה של Y על בסיס X הינו: $Y = -0.2X + 0.5$.
 חשבו מהו מקדם המתאם בין X ל- Y .

תשובות סופיות:

- | | | |
|----------------------|-----------------------|-------------|
| ג. 12.4 אלפי ₪. | ב. $Y = 0.8X + 0.4$. | א. 0.8. (1 |
| ג. 14.6 שנים. | ב. 4.25 אלפי ₪. | א. 0.75. (2 |
| ג. $Y = 1.2X + 29$. | ב. 29. | א. 1.2. (3 |
| | | א. -0.2. (4 |

סטטיסטיקה

פרק 40 - מדדי קשר-רגרסיה - שונות מוסברת ושונות לא מוסברת

תוכן העניינים

1. כללי 242

מדדי קשר – רגרסיה – שונות מוסברת ושונות לא מוסברת:

רקע:

המטרה ברגרסיה היא להסביר את השונות של המשתנה התלוי. למשל, להסביר את השונות של המשכורת באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

r^2 - החלק מהשונות של המשתנה התלוי מוסבר. השונות המוסברת נקראת גם שונות ניבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

שאלות:

- (1) נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן, נתון שסטיית התקן של מחירי הדירות הינה 200.
- איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?
 - איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?
 - מהי השונות המוסברת ומהי השונות הלא מוסברת של מחירי הדירות?
- (2) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו:
- אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.
 - אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 0.
 - אם השונות המשותפת של X ושל Y היא 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות רב-ברירה:

- (3) בקשר בין שני משתנים התקבל: $r^2 = 0.64$, לכן:
- ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.
 - 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.
 - הקשר בין שני המשתנים הוא בעוצמה של 0.64.
 - כל התשובות נכונות.
- (4) אם מגדילים את r^2 , ניתן לומר כי:
- אחוז השונות המוסברת יקטן.
 - אחוז השונות המוסברת יגדל.
 - אחוז השונות המוסברת יישאר ללא שינוי.
 - סטיית התקן משתנה.
 - לא ניתן לדעת.

- (5) בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים : מבחן בסוף סמסטר א' X ומבחן בסוף סמסטר ב' Y . כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב' לפי הציון במבחן סוף סמסטר א' התקבלה שונות טעויות של 80, ושונות ניבויים של 20.
- לפי נתונים אלו, מקדם המתאם בין הציון במבחן סוף סמסטר א' לבין הציון במבחן סוף סמסטר ב' הוא :
- א. 0.44
 - ב. - 0.44
 - ג. עוצמת ההקשר הלינארי היא 0.44, אך אין אפשרות לדעת את סימנה.
 - ד. אין אפשרות לחשב את מקדם המתאם.
 - ה. 0.35

תשובות סופיות:

- (1) א. 49% ב. 51%
ג. שונות מוסברת : 19,600, שונות לא מוסברת : 20,400
- (2) א. לא נכון. ב. נכון. ג. נכון.
- (3) ב'.
- (4) ב'.
- (5) ג'.