

סטודנטים יקרים

לפניכם ספר תרגילים בקורס סטטיסטיקה הסקית. הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה, המועבר ברשת האינטרנט On-line.

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

הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי, לדוגמה [לחצו כאן](#).

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

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



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

גול זה בול. בשבילך!

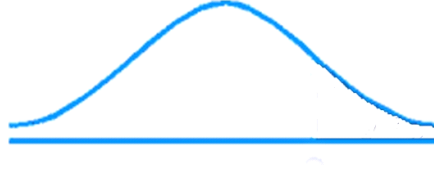
תוכן

3	פרק 1 - התפלגויות רציפות מיוחדות - התפלגות נורמלית.....
12	פרק 2 - הסקה סטטיסטית - הקדמה.....
15	פרק 3 - התפלגות הדגימה.....
15	ממוצע המדגם ומשפט הגבול המרכזי.....
22	פרק 4 - מושגים בסיסיים באמידה.....
29	פרק 5 - בדיקת השערות על פרמטרים.....
29	הקדמה.....
32	טעויות בבדיקת השערות.....
34	פרק 6 - בדיקת השערות על תוחלת (ממוצע).....
34	כאשר שונות האוכלוסיה ידועה.....
39	מובהקות התוצאה (P-VALUE) בבדיקת השערות על תוחלת עם שונות ידועה.....
44	בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסיה אינה ידועה.....
49	ניתוח פלטי SPSS.....
54	פרק 7 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים.....
54	כששונות האוכלוסיה לא ידועות ומניחים שהן שוות.....
58	ניתוח פלטים.....
65	פרק 8 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים).....
65	בדיקת השערות למדגמים מזווגים.....
71	מדגמים מזווגים - ניתוח פלטים.....
77	פרק 9 - מדדי קשר - מדד הקשר של קרמר.....
80	פרק 10 - מבחני חי בריבוע.....
82	פרק 11 - מדדי קשר - מדד הקשר של ספירמן.....
86	פרק 12 - מדדי קשר - מדד הקשר הלינארי (פירסון).....

פרק 1 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

רקע:

התפלגות נורמלית הינה התפלגות של משתנה רציף. ישנם משתנים רציפים מסוימים שנהוג להתייחס אליהם כנורמליים כמו: 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.

$$X \sim N(\mu, \sigma^2)$$

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

כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלבנטיים 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 .

$$Z \sim N(0,1^2)$$

תהליך התקנון מבוצע על ידי הנוסחה הבאה:

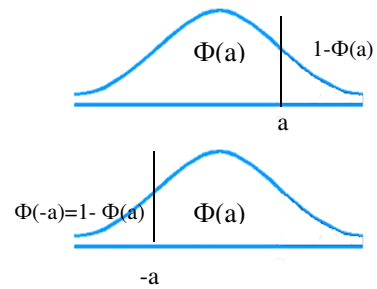
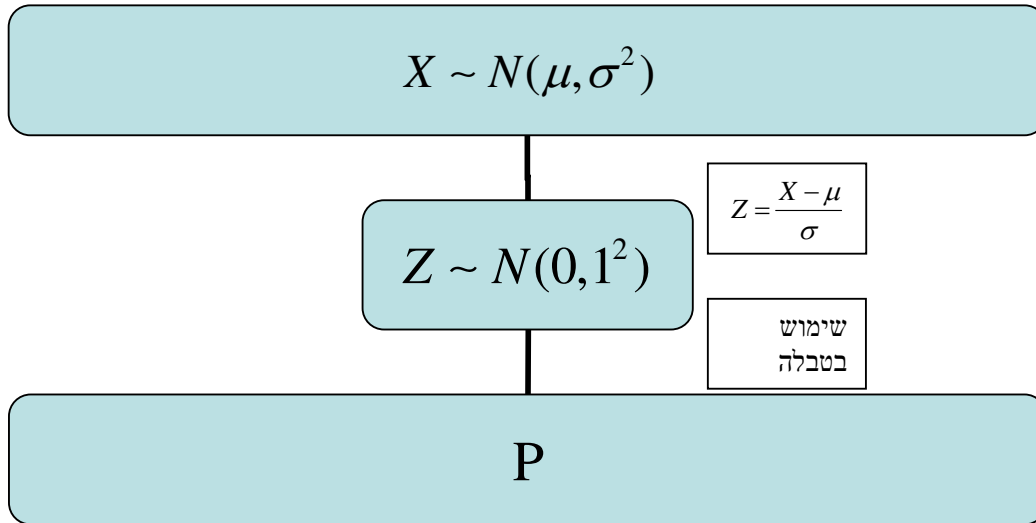
$$Z = \frac{X - \mu}{\sigma}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן.

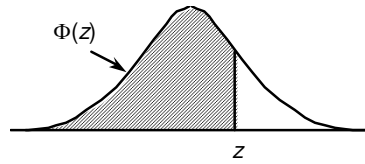
ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.

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.

- א. מה אחוז חפיסות השוקולד ששוקלות מתחת ל- 110 גרם?
- ב. מה אחוז חפיסות השוקולד השוקלות מעל 110 גרם?
- ג. מה אחוז חפיסות השוקולד השוקלות מתחת ל 92 גרם?
- ד. מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

תרגילים:

1. הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ.
 - א. מה אחוז האנשים שגובהם מתחת ל- 182.4 ס"מ?
 - ב. מה אחוז האנשים שגובהם מעל 190 ס"מ?
 - ג. מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
 - ד. מה אחוז האנשים שגובהם מתחת ל- 170 ס"מ?
 - ה. 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.
 - א. מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?
 - ב. מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?
 - ג. מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?
 - ד. 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.
 - א. מה אחוז האנשים שמשקלם נמוך מ- 55 ק"ג?
 - ב. מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
 - ג. מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל- 70 ק"ג?
 - ד. לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ- 4 ק"ג?
 - ה. מה 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.
 - א. מצאו את העשירון העליון.
 - ב. מצאו את האחוזון ה-95.
 - ג. מצאו את העשירון התחתון.

5. ציוני מבחן אינטליגנציה מתפלג נורמלית עם ממוצע 100 ושונוות 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, נתון ש-33% מהבקבוקים הם עם 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 סטיית בתקן של אורך חיי מכשיר?
 - מה הסיכוי שמכשיר אקראי יחיה פחות מ- 460 שעות?
 - מהו המאון העליון של אורח 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.



א. לאיזו התפלגות הממוצע הגבוה ביותר?
 ב. במה מבין המדדים הבאים התפלגות 1 ו 2 זהות?

א. בעשירון העליון.

ב. בממוצע.

ג. בשונות.

ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?

א. 1

ב. 2

ג. 3

ד. 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?

10. ההוצאה החודשית לבית אב בעיר "טרירה" מתפלגת נורמלית עם ממוצע של 2000 דולר וסטית תקן של 300 דולר. בחרו באקראי 5 בתי אב. ההסתברות שלפחות אחד מהם מוציא בחודש מעל ל-T דולר היא 0.98976.
- א. מה ערכו של T?
- ב. מה הסיכוי שההוצאה החודשית של בית אב בעיר תהיה לפחות סטיית תקן אחת מעל T?
- ג. מסתבר שנפלה טעות בנתונים, ויש להוסיף 100 דולר להוצאות החודשית של כל בתי האב בעיר. לאור זאת, מה ההסתברות שההוצאה החודשית של בית אב נמוכה מ-1800 דולר?
11. אורך שיר אקראי המשודר ברדיו מתפלג נורמלית עם תוחלת של 3.5 דקות וסטית תקן של שלושים שניות.
- א. מה ההסתברות שאורך של שיר אקראי המנוגן ברדיו יהיה בין 3 ל 2.5 דקות?
- ב. מהו הטווח הבין רבעוני של אורך שיר המשודר ברדיו?
- ג. ביום מסוים מנוגנים 200 שירים ברדיו. כמה שירים מתוכם תצפה שיהיו באורך הנמוך מ 3.5 דקות?
- ד. בשעה מסוימת שודרו 8 שירים. מה ההסתברות שרבע מהם בדיוק היו ארוכים מ-4 דקות והיתר לא?

פתרונות :

<u>שאלה 3</u>	<u>שאלה 1</u>
א. 26.43%	א. 89.25%
ב. 89.44%	ב. 2.28%
ג. 39.44%	ג. 0
ד. 0.383	ד. 50%
ה. 100%	

<u>שאלה 7</u>	<u>שאלה 5</u>
א. 500	א. 119.2
ב. 100	ב. 80.8
ג. 0.3446	ג. 112.6
ד. 733	ד. 87.4
ה. 267	

<u>שאלה 9</u>	<u>שאלה 8</u>
א. 0.1587	א. 3
ב. 0.0228	ב. בממוצע.
ג. 0.8563	ג. 1
ד. 0.3975	

<u>שאלה 11</u>	<u>שאלה 10</u>
א. 0.1359	א. 1925
ב. 0.675	ב. 0.2266
ג. 100	ג. 0.1587
ד. 0.25	

פרק 2 - הסקה סטטיסטית - הקדמה

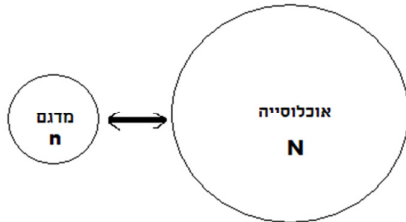
רקע:

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 תהייה דגימה מקרית הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכויי להיכלל במדגם.

סטטיסטי – גודל המחושב על המדגם.

פרמטר – גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים

למשל:

פרמטר (אוכלוסייה)	סטטיסטי (מדגם)	
μ	\bar{X}	ממוצע
P	\hat{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?

תרגילים :

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 בישוב "העוגן".
נגיד את x להיות מספר המקלטים של משפחה אקראית.
מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלויזיה במדגם.

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

- מי היא האוכלוסייה ומהו המשתנה הנחקר?
 - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?
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?

פרק 3 - התפלגות הדגימה

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 .

א. 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. מי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

ב. זגימה מהתפלגות נורמאלית

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2 ממוצע המדגם גם יתפלג נורמאלית:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

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

זוגמה: (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?

ג. משפט הגבול המרכזי

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק גדול ($n \geq 30$)

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

זוגמה: (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?

תרגילים :

1. מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
 - א. מי האוכלוסייה?
 - ב. מה המשתנה?
 - ג. מהם הפרמטרים?
 - ד. מהו גודל המדגם?
 - ה. מהו תוחלת ממוצע המדגם?
 - ו. מהי טעות התקן?

4. משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם
 - א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?

נתון כי ביום מסוים נולדו 4 תינוקות.

 - ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג ?
 - ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
 - ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
 - ה. הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

5. הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.
 - א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
 - ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
 - ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתוחלת הגבהים בפחות מ-5 ס"מ?
 - ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

6. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
 - 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?

7. נפח היין בבקבוק מתפלג נורמאלית 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?
8. משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4 .
- א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
- ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
- ג. הסבר את ההבדל בתשובות של שני הסעיפים.
10. 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 ₪?
12. אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ .
- א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?
- ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?
- ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. העזר במשפט הגבול המרכזי.

14. נתון ש $X \sim N(\mu, \sigma^2)$ דגמו 5 תצפיות מאותה התפלגות והתבוננו במוצע המדגם \bar{X} :

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

א. 0

ב. 0.5

ג. 1

ד. לא ניתן לדעת.

15. נתון ש X מתפלג כלשהו עם תוחלת μ ושונות σ^2 .

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

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

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

ג. $\bar{X} \sim N(\mu, \sigma^2)$

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

16. 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 :

(בחר 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.

פתרונות:**שאלה 4**

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

שאלה 6

א. 0.0465

ב. 27.71

ג. 0.2628

שאלה 7

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

שאלה 8

א. 0.5468

ב. 0.6826

שאלה 10

0.0475

שאלה 12

א. 0.9772

ב. 0.0228

ג. 271

שאלה 14

התשובה ב

שאלה 15

התשובה ד

שאלה 16

התשובה ג

פרק 4 - מושגים בסיסיים 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.

מה הפרמטר בדוגמה זו?

מהי טעות האמידה של ערוץ 10?

- $E(\hat{\theta}) = \theta$ יהיה אומד חסר הטיה ל θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל θ :
- טעות התקן של אומד היא סטיית התקן שלו, כלומר : $\sigma(\hat{\theta}) = S.E$

להלן פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה: μ

$$\bar{x} = \frac{\sum x}{n}$$

האומד הנקודתי שלו יהיה: ממוצע המדגם

$E(\bar{x}) = \mu$ לכן \bar{x} הינו אומר חסר הטויה ל μ .

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE$$

כמו כן טעות תקן: SE

פרופורציה באוכלוסייה: p

$$\hat{p} = \frac{y}{n}$$

האומד הנקודתי שלו יהיה: פרופורציה במדגם:

$E(\hat{p}) = p$ לכן \hat{p} הינו אומר חסר הטויה ל p .

$$\sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$$

כמו כן טעות התקן:

שוונת האוכלוסייה: σ^2

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

האומד הנקודתי שלו יהיה:

$E(S^2) = \sigma^2$ ולכן S^2 הינו אומד חסר הטויה 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.

תרגילים:

1. מתוך 500 טירונים נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
- ב. מהי טעות התקן של האומדן כשהמדגם בגודל 500?
- ג. מהו האומדן לפרמטר?
- ד. 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.
- א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
- ב. מהי טעות התקן של האומדן?
- ג. מהו האומדן לפרמטר?
- ד. 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.
- א. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
- ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
- ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
4. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו:
- $$\sum_{i=1}^{20} X_i^2 = 1502.2 \quad \sum_{i=1}^{20} X_i = 162$$
- א. אמדו את השכר הממוצע של השכירים במשק.
- ב. אמדו את סטיית התקן של שכר השכירים במשק.

5. במטרה לאמוד את ממוצע האוכלוסייה. דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
 א. סטיית התקן של האוכלוסייה.
 ב. סטיית התקן של ממוצע האוכלוסייה.
 ג. סטיית התקן של המדגם.
 ד. 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 תהיה:

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 באומדן 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 האומדן 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.

פתרונות:**שאלה 3**

א. 177.9

ב. 64.1

ג. 0.4

שאלה 4

א. 8.1

ב. 3.16

שאלה 5

התשובה היא ד.

שאלה 6

התשובה היא ג.

שאלה 7

התשובה היא ד.

שאלה 8

התשובה היא ד.

שאלה 9

התשובה היא ג.

פרק 5 - בדיקת השערות על פרמטרים

הקדמה

רקע:

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

שלב א: נזהה 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 שניקרא כלל הכרעה:
 הכלל יוצר אזור שניקרא אזור דחייה (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה)
ואזור קבלה (קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי.

אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שניקרא 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 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:

$$n = 20$$

$$\bar{X} = 3120$$

$$S = 280$$

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

תרגילים:

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.

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

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

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.

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

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.

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

טעויות בבדיקת השערות

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

הכלל יוצר אזור שניקרא אזור דחייה (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ואזור 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. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

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

הגדרת הטעויות:

טעות מסוג ראשון- להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.

טעות מסוג שני- להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.

דוגמה: (פתרון בהקלטה)

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

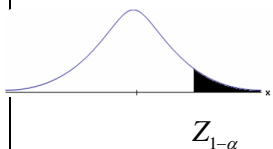
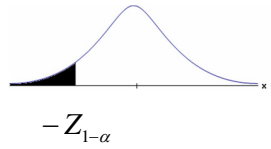
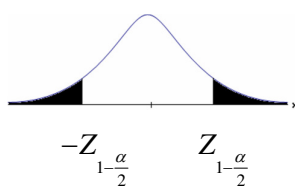
תרגילים:

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.
- א. רשמו את השערות המחקר.
 ב. מה מסקנת המחקר?
 ג. איזו סוג טעות יתכן וביצעו במחקר?
2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
- א. האם ניתן לדעת אם בוצע טעות במחקר?
 ב. 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.
- א. מהי אוכלוסיית המחקר?
 ב. מה המשתנה הנחקר?
 ג. מה הפרמטר הנחקר?
 ד. מה השערות המחקר?
 ה. מה מסקנת המחקר?
 ו. מהי סוג הטעות האפשרית במחקר?

פרק 6 - בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסייה ידועה

רקע:

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

סטטיסטי המבחן:

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

חלופה אחרת לכלל הכרעה:

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

דוגמה : (פתרון בהקלטה)

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 שזובלו בשיטה החדשה. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדוק את ההשערה ברמת מובהקות של 1%.

9

תרגילים:

1. ממוצע הציונים בבחינת הברגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהייה 15 מה מסקנתכם ברמת מובהקות של 5%?
2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 ס"מ"ק וסטיית תקן 20 ס"מ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 ס"מ"ק במדגם בגודל 25.
 - א. מה מסקנתכם ברמת מובהקות של 2.5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?
3. מהנדס האיכות מעוניין לבדוק אם מכונה מכוילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
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? בחר בתשובה הנכונה.

א. הגדלת רמת המובהקות לא תשתנה 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: (בחר בתשובה הנכונה)

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.

פתרונות :**שאלה 1:**

נקבל H_0

שאלה 2:

נדחה H_0

שאלה 3:

נדחה H_0

שאלה 4:

נדחה H_0

שאלה 5:

נקבל H_0

שאלה 6:

א

שאלה 7:

ג

שאלה 8:

א. אותה מסקנה

ב. לא ניתן לדעת.

מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות 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 :

$$\text{אם } p_v \leq \alpha \text{ דוחים את } H_0$$

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

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

אם ההשערה היא דו צדדית :

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

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 :
3. σ ידועה			תנאים :
4. $X \sim N$ או מדגם מספיק גדול			
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	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

כאשר בהנחת השערת האפס : $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$

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

דוגמה: (פתרון בהקלטה)

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 נורמאלית 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%?

תרגילים:

1. לפניך 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 מאוד, 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%? בחר 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 : (בחר בתשובה הנכונה)
 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.
8. בבדיקת השערות מסוימת התקבל $p\text{ value}=0.0254$ לכן (בחר 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 .

פתרונות :**שאלה 1:**

0.0228

שאלה 2:

עבור כל רמת מובהקות סבירה.

שאלה 3:

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

שאלה 4:

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיוול.

שאלה 5:

נכון

שאלה 6:

תשובה: א

שאלה 7:

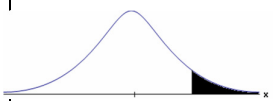

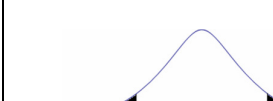
תשובה: א

שאלה 8:

תשובה: ג

בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה

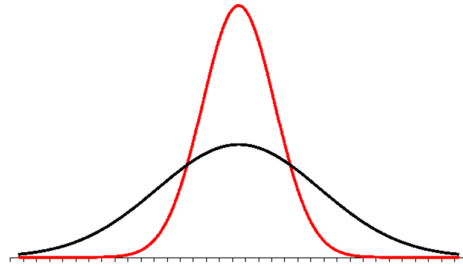
רקע:

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

סטטיסטי המבחן :

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

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

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%.

תרגילים:

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 מבארה"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.

א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.

ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

ד. 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 :

$$H_0 : \mu = \mu_0$$

$$H_0 : \mu < \mu_0$$

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.

פתרונות:**שאלה 1:** H_0 נדחה**שאלה 2:** H_0 נדחה**שאלה 3:** H_0 נקבל**שאלה 4:** H_0 נקבל**שאלה 5:**

התשובה היא : ב

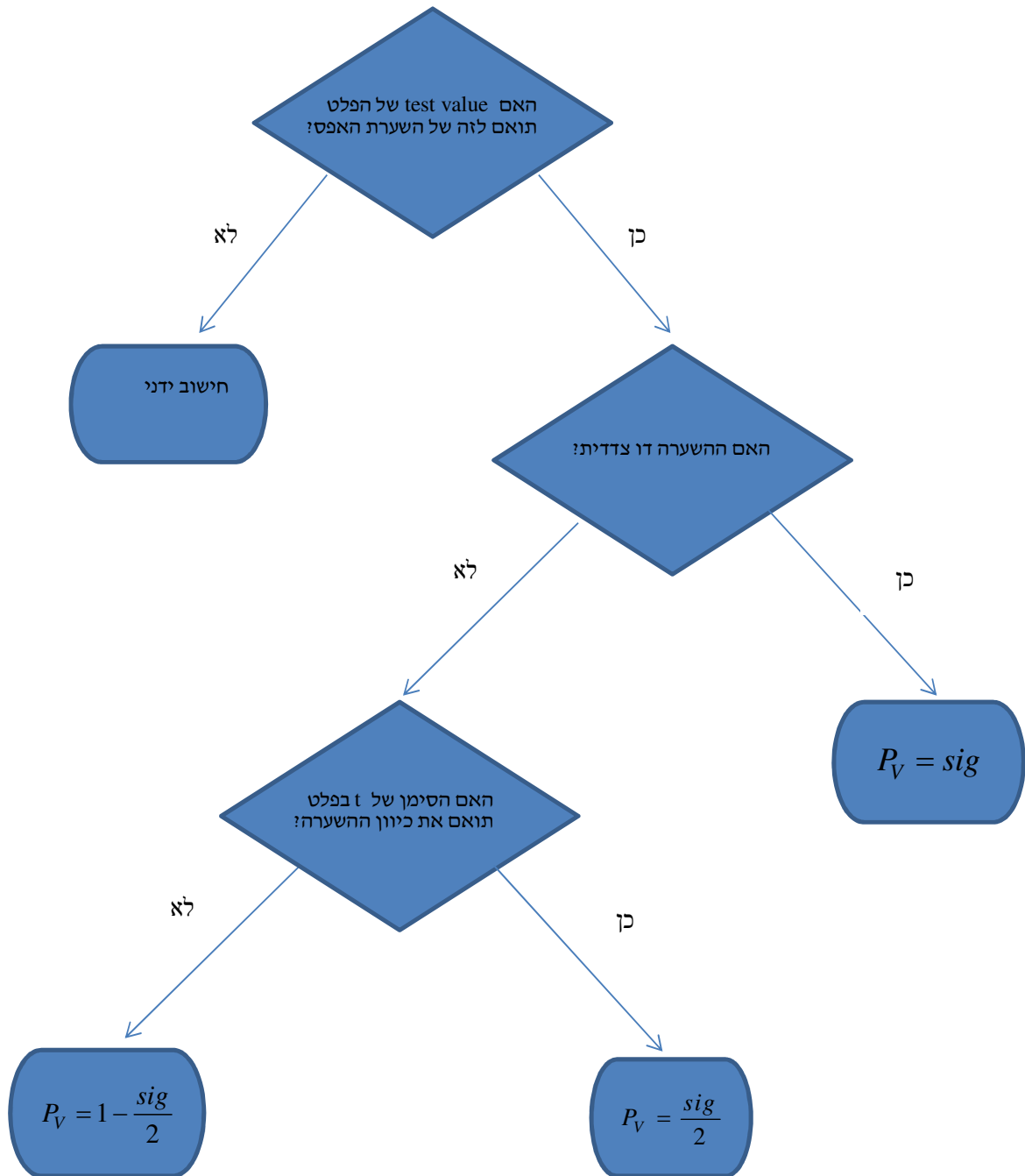
שאלה 6:

התשובה היא : ג

ניתוח פלטי SPSS

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS



דוגמה: (פתרון בהקלטה)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
x	25	87.6400	64.90434	12.98087

One-Sample Test

	Test Value = 60					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
x	2.129	24	.044	27.64000	.8488	54.4312

ממוצע הציונים במבחן המיצב בחשבון הוא 60. הוחלט לדגום כיתה אקראית של 25 תלמידים וללמד אותם בשיטת לימוד חדשה.

א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?

ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?

ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

תרגילים :

1. באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלי. בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פלט תוכנת SPSS על מדגם של 16 סטודנטים אקראיים מתואר ראשון:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	16	23.4375	2.50250	.62562

One-Sample Test

	Test Value = 23					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	.699	15	.495	.43750	-.8960	1.7710

- א. מהו המבחן הסטטיסטי שנעשה כאן?
 ב. מה ערכו של הפרמטר לפי השערת האפס?
 ג. רשום רווח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.
 ד. בדוק ברמת מובהקות של 5% האם הגיל הממוצע כיום שונה מבעבר?
2. קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עבור כל ילד בדקו כמה זמן לקח לו לסיים את המשימה בדקות. להלן תוצאות הניתוח הסטטיסטי:

One-Sample Test

	Test Value = 4.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
time	-1.853	24	.076	-.09200	-.1944	.0104

- א. כמה ילדים השתתפו במחקר?
 ב. מצא רווח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עבור ילדים בני 6.
 ג. מה יש להניח כדי שרווח הסמך מסעיף א' יהיה תקף?
 ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמוך מ-4.5 דקות.

3. לפניך פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS. הניתוח הוא עבור מדגם אקראי של קבוצת נבחנים בבגרות באנגלית.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
grade	???	???	19.62787	2.95901

One-Sample Test

	Test Value = 75					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
grade	???	43	.017	-7.34091	-13.3083	???

- א. השלם את הגדלים החסרים המסומנים בסמני שאלה בפלט.
 ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
 ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
 ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

4. יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ- 27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות:
 21, 21, 20, 24, 22 מ"ג. הנח כי כמות הניקוטין בסיגריות מפולג נורמאלי.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
nicotine	5	21.6000	1.51658	.67823

One-Sample Test

	Test Value = 27					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
nicotine	-7.962	4	.001	-5.40000	-7.2831	-3.5169

- א. האם ברמת מובהקות של 5% ניתן להסיק שיש אמת בפרסום?
 ב. אם הינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
 ג. בדוק האם ניתן להגייד שתוחלת רמת הניקוטין שונה מ- 26 ברמת מובהקות של 5%.

פתרונות :**שאלה 1:**

ב. 23

ג. (22.104, 24.771)

ד. נקבל H_0 **שאלה 2:**

א. 25

ב. (4.3056, 4.5104)

ד. נדחה H_0 **שאלה 3**א. $n = 44$

$$\bar{X} = 67.66$$

$$t = -2.48$$

$$upper = -1.3735$$

ב. 0.017

ג. 0.0085

ד. 0.9915

שאלה 4:

א. נכריע שיש אמת בפרסום.


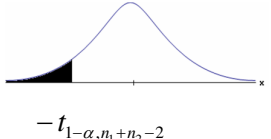
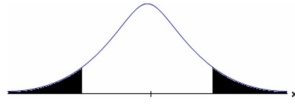
ב. מסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

פרק 7 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסייה לא ידועות ומניחים שהן שוות

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:
1. מדגמים בלתי תלויים 2. σ_1, σ_2 לא ידועות אך שוות 3. המשתנים בכל אוכלוסייה מתפלגים נורמלית			תנאים:
$t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha, n_1+n_2-2}$ 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	או $t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2}, n_1+n_2-2} \quad t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ דוחים את H_0	אזור הדחייה של H_0:

סטטיסטי המבחן :

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \quad \text{השונות המשוקללת} \quad t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

חלופה אחרת לכלל הכרעה:

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

דוגמה : (פתרון בהקלטה)

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשמשים 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%.

תרגילים:

1. להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (מטרים רבועים):

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 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.

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

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 W100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

פתרונות :**שאלה 1:**

לא נדחה H_0

שאלה 2:**שאלה 3:**

א. נדחה H_0

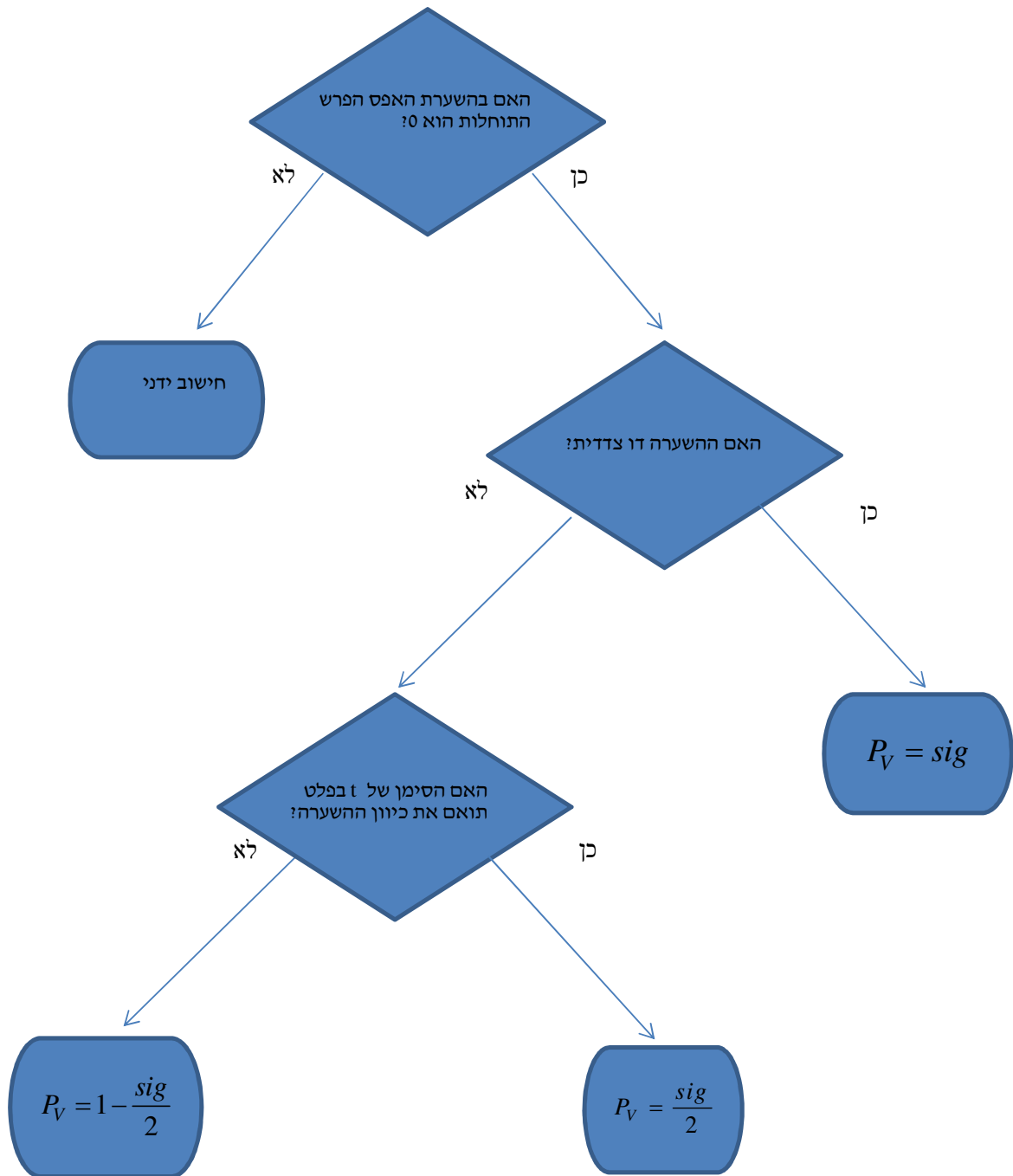
ב. רמות מובהקות של לפחות 5%

ג. לא נדחה H_0

ניתוח פלטים

רקע:

מובהקות התוצאה על סמך הפלט:



דוגמה: (פתרון בהקלטה)

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

כתב ופתר - ברק קנדל ©

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים משני אזורים שונים במדינה על מס' האחים והאחיות שלהם. להלן הפלט שהתקבל:

Group Statistics

	Region of the United States	N	Mean	Std. Deviation	Std. Error Mean
Number of Brothers and Sisters	North East	676	3.76	2.939	.113
	South East	410	4.05	2.993	.148

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Number of Brothers and Sisters	Equal variances assumed	.173	.677	-1.583	1084	.114	-.293	.185	-.657	.070
	Equal variances not assumed			-1.576	850.945	.115	-.293	.186	-.658	.072

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?

ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.

ד. מהי מובהקות התוצאה לבדיקת הטענה שהפרש הממוצע בין-"South East"

לבין-"North East" חיובי ?

תרגילים :

1. להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים.
רמת האופטימיות נמדדה בסקאלת ציונים של 1 עד 5.

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
optimizm	MALE	633	2.6053	.49781	.01979
	FEMALE	568	2.5503	.48483	.02034

Independent Samples Test

		Levene's Test for quality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	190.977	.053	.05500	.02838	-.00068	.11067

- א. האם ניתן להניח ששוונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
 ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
 ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.
 ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

2. פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר.
להלן הפלט שהתקבל:

T-Test

Group Statistic:

	N	Mean	Std. Deviation	Std. Error Mean
grade A	9	96.8889	9.40006	3.13335
B	11	108.4545	11.46616	3.45718

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
grade	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828
Equal variances assumed									
			-2.479	17.997	.023	-11.56566	4.66583	-21.36832	-1.76299
Equal variances not assumed									

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
- כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
- האם ניתן לומר כי מידת הפיזור של ציוני האנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha=0.05$.
- האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha=0.01$.

3. כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז. לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

T-Test

Group Statistics

מקום המגורים	N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין עיר א	50	24.8072	1.38978	.19654
עיר ב	50	23.0131	1.62070	.22920

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
גיל הנישואין	Equal variances assumed	.330	.567	5.942	98	.000	1.79415	.30193	1.19497	2.39332
	Equal variances not assumed			5.942	95.772	.000	1.79415	.30193	1.19480	2.39350

- א. מהו המבחן הסטטיסטי שנעשה כאן?
 ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.
 ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

4. להלן פלט של תוכנת SPSS:

T-Test

	N	Mean	Std. Deviation	Std. Error Mean
X	26	36.3077	13.23259	2.59513
Y	24	46.4583	20.96369	4.27920

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2- ailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.446	.040	-2.164	???	.044	-10.15064	???	-20.03781	-.26347
Equal variances not assumed			-2.038	38.267	.048	???	5.00462	-20.27964	-.02164

א. השלם את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y?

ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y?

ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y?

פתרונות:**שאלה 1:**

- א. נכריע שוויון שוניות.
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.
 ג. 0.11076
 ד. $2.5665 \leq \mu \leq 2.6441$

שאלה 2:

- א. מבחן t למדגמים בלתי תלויים.
 ב. מבחן t למדגמים מזווגים.
 ג. נקבע שקיים שוויון שוניות.
 ד. לא נקבל את טענת הפסיכולוגים.

שאלה 3:

- א. מבחן t להשוואת תוחלת במדגמים בלתי תלויים.
 ב. $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$
 ג. כן.



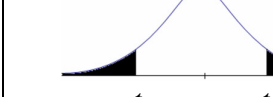
שאלה 4:

- א. -10.15, 4.69, 48
 ב. 0.04
 ג. 0.978
 ד. 0.022

**פרק 8 - בדיקת השערות על תוחלת הפרשים 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:
<p align="center">4. σ_D אינה ידועה</p> <p align="center">5. $D \sim N$ או מדגם מספיק גדול</p>			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
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ או $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:

סטטיסטי המבחן:

$$t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}$$

דוגמה: (פתרון בהקלטה)

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה נמוכים מהמחירים מרשת השיווק "שופרסל".

לצורך בדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים:

המוצר	מגה בעיר	שופרסל
שמפו	17	18
גיל 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".

תרגילים:

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

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

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

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

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

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

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

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

- פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נקודות לעומת סמסטר א'. הנח 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 נורמאלית 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	5.4	6.9	7.0	5.2
משקל במכשיר 2	5.3	6.9	7.1	5.0

נניח שהמשקלים מתפלגים נורמלית.

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.

6. 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 :

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

פתרונות:**שאלה 1:**

לא נדחה H_0

שאלה 2:

לא נדחה H_0

שאלה 3:

א. $0.25 \leq p \leq 0.5$

ב. 0.5

ג. לא נדחה H_0

שאלה 4:

התשובה היא ד.

שאלה 5:

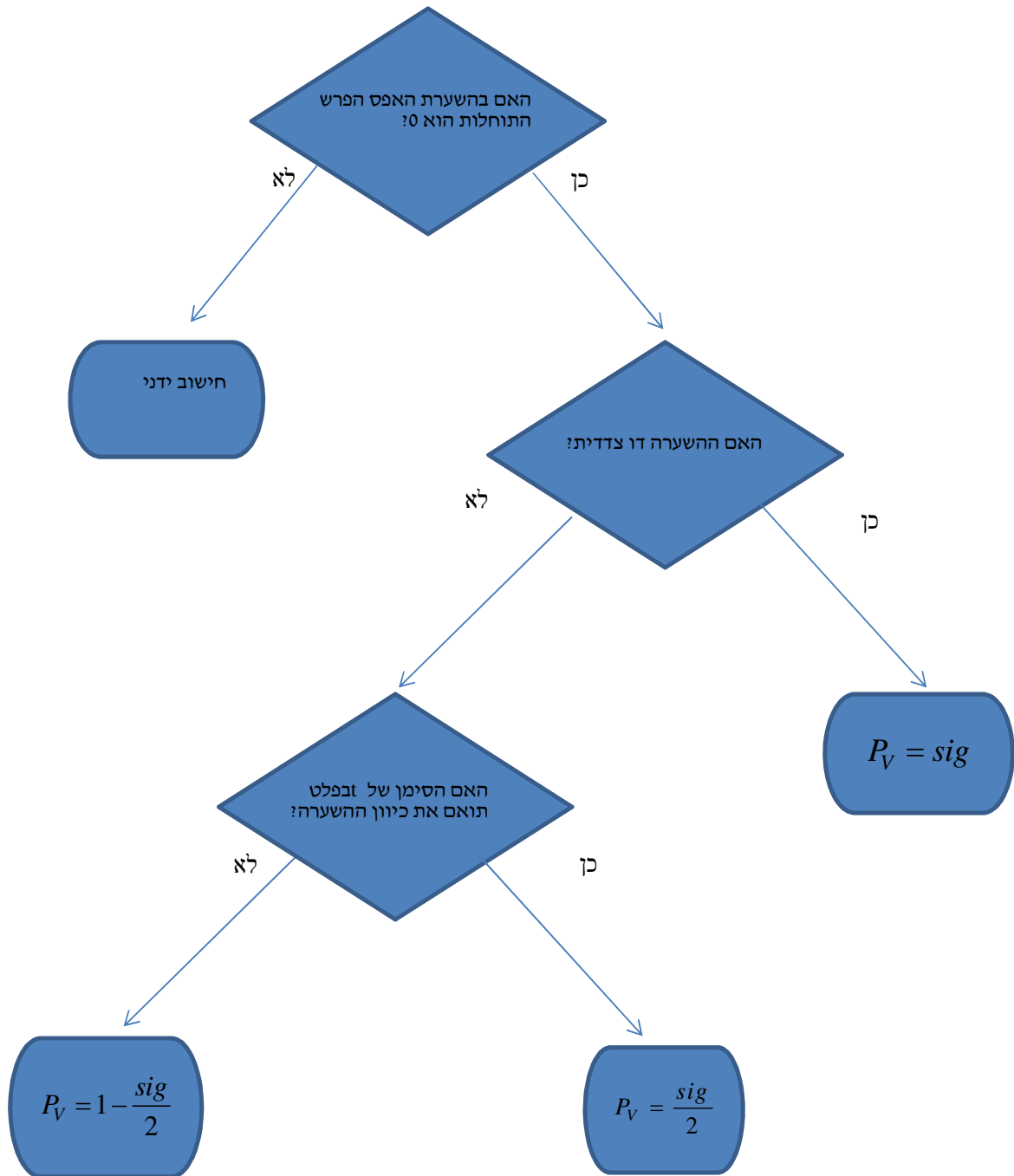
התשובה היא ד.

שאלה 6:

התשובה היא ג.

מדגמים מזווגים - ניתוח פלטים

רקע:



דוגמה : (פתרון בהקלטה)

כדי לבדוק את ההשפעה של קורס לגמילה מעישון נלקח מדגם מקרי של 5 נבדקים. עבור כל אחד מהם נמדדה צריכת הסיגריות היומית לפני הקורס וחודשיים אחריו. הניחו שצריכת הסיגריות מתפלגת נורמלית. להלן התוצאות:

נבדק	1	2	3	4	5
לפני	40	22	25	28	30
אחרי	30	24	13	10	12

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 BEFORE	29.0000	5	6.85565	3.06594
AFTER	17.8000	5	8.72926	3.90384

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 BEFORE - AFTER	11.20000	8.19756	3.66606	3.38452	19.01548	3.055	4	.038

בדוק ברמת מובהקות של 5% האם הקורס יעיל.

תרגילים :

1. בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים על השכלת הוריהם, להלן הפלט שהתקבל:

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	973	.943

- א. תן אומדן להפרש הממוצעים.
 ב. תן אומדן לטעות התקן של הפרש הממוצעים.
 ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

2. בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות:

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Italy	???	300	.86742	.05008
	South Korea	8.9183	???	.81992	.04734

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Italy - South Korea	-.42233	.36153	.02087	-.46341	-.38126	-20.234	???	???

- א. השלימו את החלקים החסרים בפלט. (מסומנים בסימני שאלה).
 ב. בדוק את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
 ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים.
 ד. בנה את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

3. בדקו את ציוניהם של 44 נבדקים אקראיים במבחן הפסיכומטרי. פעם אחת לפני הכנה (Before) ופעם אחת אחרי הכנה (After).

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Before - After	-7.45455	19.28303	2.90703	-13.31712	-1.59197	-2.564	43	.014

- א. רשום מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ב. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ג. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ד. מצא רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	x	54.0000	6	5.86515	2.39444
	y	46.5000	6	10.72847	4.37988

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	x - y	7.50000	??	4.72405	-4.64356	19.64356	??	5	.173

- א. מלא את החלקים החסרים בטבלה.
- ב. מהי רמת המובהקות המינימלית לקבלת הטענה שיש הבדל בין X ל- Y בממוצע ?
- ג. האם התשובה לסעיף הקודם הייתה משתנה , ואם כן גדלה או קטנה , אם הינו מוסיפים עוד תצפית שההפרש בין X ל- Y הוא 0.
- ד. מהי מובהקות התוצאה לבדיקת הטענה ש X גדול מ- Y בממוצע?
- ה. מהי מובהקות התוצאה לבדיקת הטענה ש X קטן מ- Y בממוצע?
- ו. בנו רווח סמך לתוחלת של X ברמת סמך של 90% .

פתרונות:**שאלה 1:**

א. -0.007

ב. 0.1

ג. אין הבדל מובהק

שאלה 2:א. $d.f = 299$ ב. $n = 300$ ג. $\bar{X} = 8.496$

ד. Sig=0

א. נדחה H_0

ב. (0.46341, 0.38126)

ג. (0.4708, 0.3738)

שאלה 3:א. נדחה H_0 ב. לא נדחה את H_0

ג. (1.592, 13.317)

שאלה 4:

א. 1.5876, 11.5715

ב. 0.173

ג. יגדל

ד. 0.0865

ה. 0.9135

ו. $49.18 < \mu < 58.82$

פרק 9 - מדדי קשר - מדד הקשר של קרמר

רקע:

מתי משתמשים במדד הזה?
 כאשר אחד המשתנים הוא מסולם שמי והשני מכל סולם אפשרי.
 מדד הקשר מקבל ערכים בין 0 ל-1.
 ככל שהמדד יותר קרוב לאחד קיים קשר בעוצמה יותר חזקה בין המשתנים.

דוגמה: (פתרון בהקלטה)

במחקר רוצים לבדוק את הקשר בין מין לדעה בנושא מסוים, שאלו 100 גברים ו-100 נשים האם הם בעד/נגד/נמנע/נשא. להלן טבלת השכיחויות המשותפת שהתקבלה.

f(x)	נמנע	נגד	בעד	Y X
100	10	40	50	גבר
100	10	60	30	אישה
n=200	20	100	80	f(y)

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

X- מין (גבר/אישה) – סולם שמי

Y-דעה (בעד/נמנע/נגד) – סולם שמי/סדר

שלבים בחישוב: r_c

שלב א': נבנה את טבלת E (Expected)

נעתיק את המסגרת של טבלת O ואז כל $E_i = (f_{(x)} * f_{(y)}) / n$

f(x)	נמנע	נגד	בעד	Y X
100				גבר
100				אישה
n=200	20	100	80	f(y)

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad \text{שלב ב': נחשב}$$

$$r_c = \sqrt{\frac{1}{n(L-1)} \chi^2} \quad \text{שלב ג': נחשב}$$

כאשר L מבטא את המספר הקטן מבין מספר השורות או העמודות.

תרגילים:

1. להלן תוצאות מחקר שבדק את הקשר בין מין להשכלה. לגבי כל נחקר נבדק המין שלו והשכלתו. להלן התוצאות:

גבוהה	תיכונית	נמוכה	השכלה
			מין
20	40	120	גבר
80	20	20	אישה

האם קיים קשר בין מין להשכלה? נמק!

2. נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.
- א. בנה טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
- ב. האם קיים קשר בין פעילות גופנית למצב בריאותי? חשב לפי מדד הקשר של קרמר.

פתרונות :**שאלה 1**

0.595

שאלה 2

ב. 0.19

פרק 10 - מבחני חי בריבוע

מבחן אי תלות

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

	יום	ערב	לילה
פגומים	50	60	70
תקינים	600	700	800

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

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

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

צבע	שחור	לבן	אדום	כחול
מספר הפריטים	15	20	15	50

כמו כן בדקו את התפלגות הצבעים בחנות שכנה B :

צבע	שחור	לבן	אדום	כחול
מספר הכדורים	60	20	10	20

- א. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החניות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

תשובות סופיות - מבחני חי בריבוע

מבחן לאי תלות

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

פרק 11 - מדדי קשר - מדד הקשר של ספירמן

רקע:

מתי נשתמש במדד ספירמן ?

כאשר אחד המשתנים מסולם סדר והשני מסולם סדר ומעלה.

הקשר שהמדד בודק הוא קשר דירוגי.

מדד הקשר בודק:

1 כיוון של הקשר.

2 בודק את עצמת הקשר.

המדד מקבל ערכים בסקלה מ -1 ועד 1.



אם מדד הקשר של ספירמן יוצא 1 המשמעות היא שיש קשר דירוגי חיובי מלא : ככל המשתנה אחד עולה השני עולה ללא יוצא מן הכלל.

קשר דירוגי חיובי חלקי (שמקדם המתאם בין 0 ל-1) אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות אך לא באופן מוחלט.

אם מדד הקשר של ספירמן יוצא -1 המשמעות היא שיש קשר דירוגי שלילי מלא : ככל שהמשתנה אחד עולה השני יורד ללא יוצא מן הכלל.

קשר דירוגי שלילי חלקי (שמקדם המתאם הוא בין 0 ל- -1) אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אך לא באופן מוחלט.

על מנת לחשב את הקשר יש לבצע פעולת דירוג (RANK) נלמד את פעולת הדירוג דרך הדוגמה הבאה (פתרון בהקלטה)

שם התלמיד	הערכה	דירוג R
ערן	בינוני	
מיכל	מצוין	
עודד	חלש	
רוני	טוב	
יעל	טוב	

כאשר מדרגים אם יש כמה תצפיות שתופסות את אותו הערך אז הדירוג שלהם הוא הממוצע של המקומות שהן תופסות.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

הנוסחה של מדד הקשר :

דוגמה : (פתרון בהקלטה)

בתחרות רוקדים עם כוכבים השתתפו 7 זוגות, 2 שופטים נתנו את ציוניהם לריקוד של כל זוג.

מהי מידת ההתאמה בין ציוני השופטים?

X- ציון שופט א (סולם סדר)

Y- ציון שופט ב (סולם סדר)

להלן התוצאות שהתקבלו :

מספר הזוג	ציון שופט א	R _x	ציון שופט ב	R _y	$d=r_x-r_y$	d^2
1	4		5			
2	5		5			
3	6		7			
4	5		7			
5	8		9			
6	7		9			
7	3		7			

תרגילים:

1. בתחרות יופי חילקו שני שופטים ציונים למועמדות:

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

האם קיים קשר בין שתי הערכות השופטים? נמק והסבר!

2. משרד רצה לבחון האם קיים קשר בין מידת המוטיבציה של העובדים שלו לבין מספר

החיסורים של העובדים בחודש עבודה. להלן התוצאות שהתקבלו:

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

האם קיים קשר בין רמת המוטיבציה של העובד ומספר החיסורים שלו? חשב באמצעות מדד

הקשר המתאים והסבר.

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

פתרונות:**שאלה 1:**

0.973

שאלה 2:

-0.85

שאלה 3:

לא נכון

פרק 12 - מדדי קשר - מדד הקשר הלינארי (פירסון)

רקע:

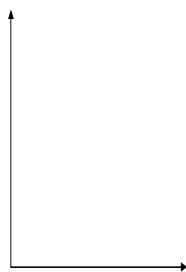
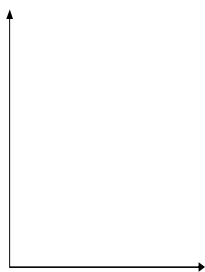
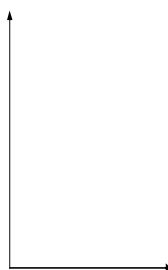
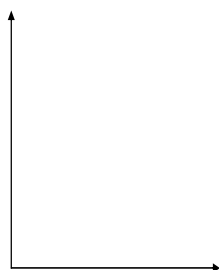
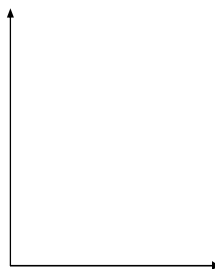
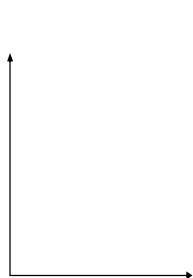
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X מסבירה את ההכנסה שלו Y . במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו. בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים. למשל, בבניין של 5 דירות בדקו את הנתונים הבאים: X - מס' חדרים בדירה. Y - מס' נפשות הגרות בדירה. להלן התוצאות שהתקבלו:

מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

נשרטט מנתונים הללו דיאגרמת פיזור:



נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבדק 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

$$y = bx + a$$

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

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

מתאם חיובי חלקי אומר שככל שמשנתנה 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

שמשנתנה 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.

$$COV(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \frac{\sum xy}{n} - \bar{x} \cdot \bar{y} : \text{שונות משותפת}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 : \text{שונות של המשתנה X}$$

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 : \text{שונות המשתנה Y}$$

$$r_{xy} = \frac{COV(x, y)}{s_x \cdot s_y} : \text{מקדם המתאם הלינארי}$$

תרגילים:

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

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

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

2. 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:

x	y
10	12
14	15
15	15
18	17
20	21

- 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:

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

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

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

- א. חשב את מדד הקשר הלינארי בין X ל-Y. מיהו המשתנה התלוי?
 ב. 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:

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

$$\sum_{i=1}^{20} Y_i^2 = 2080 \quad \sum_{i=1}^{20} X_i^2 = 2960$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשב 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 קבע נכון/לא נכון ונמק!

- א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.
 ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ $S_x = S_y = 1$ לכן מדד הקשר של פירסון יהיה 1.
 ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות אמריקאיות:

7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :

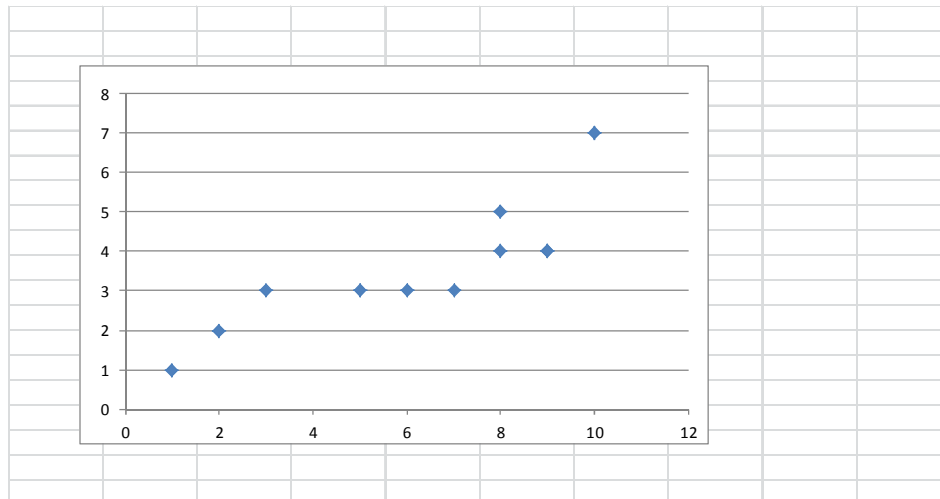
- א. הדבר מעיד שהציונים בכתה היו שליליים.
- ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
- ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
- ד. אף אחת מהתשובות לא נכונה.

8. נלקחו 20 מוצרים וניבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו

היום ערך הדולר היה - 4.2 ש"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

פתרונות:**שאלה 1:**

- א. בהקלטה
ב. -0.9325

שאלה 2:

- א. $\bar{y} = 16$
ב. $r_{xy} = 0.96$
- $\bar{x} = 15.4$

שאלה 3:

- א : 0.8

שאלה 4:

- 0.8

שאלה 5:

- 1

שאלה 6:

- א. נכון
ב. לא נכון
ג. נכון

שאלה 7:

- התשובה : ג

שאלה 8:

- התשובה : א

שאלה 9:

- התשובה : ב