הסקה וניתוח נתונים



תוכן העניינים

1. התפלגות נורמלית 1
2. הסקה סטטיסטית - הקדמה 7
3. התפלגות הדגימה ומשפט הגבול המרכזי 10
4. מושגי יסוד באמידה 16
5. רווח סמך לתוחלת (ממוצע) 21
6. רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים 31
7. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 33
8. מבוא לבדיקת השערות על פרמטרים 35
9. בדיקת השערות על תוחלת (ממוצע) 41
10. בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים 64
11. בדיקת השערות לתוחלת ההפרש במדגמים מזווגים 75
12. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות 85
13. שאלות מסכמות בבדיקת השערות 88
14. מבחני חי בריבוע 101
15. ניתוח שונות חד כיוונית 114
16. ניתוח שונות דו כיווני 123

הסקה וניתוח נתונים

פרק 1 - התפלגות נורמלית

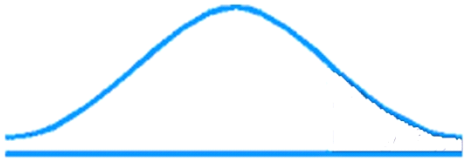
תוכן העניינים

1. כללי 1

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: $X \sim N\left(\mu,\sigma^2\right)$.

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

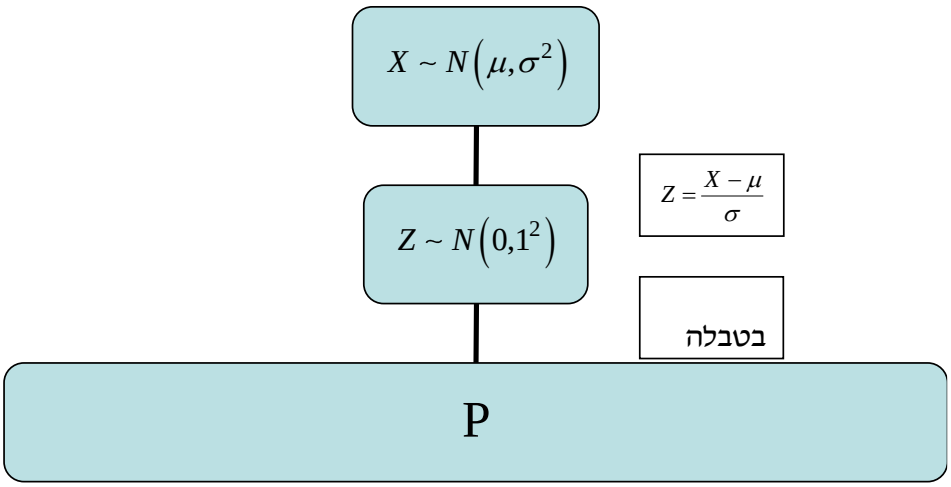
כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z: $Z \sim N\left(0,1^2\right)$.

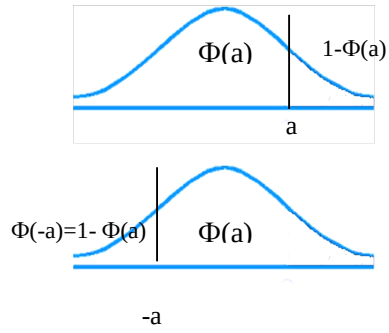
תהליך התקנון מבוצע על ידי הנוסחה הבאה: $Z=\frac{X-\mu}{\sigma}$.

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:

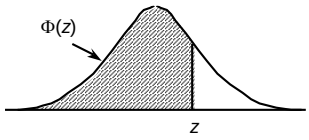
© כל הזכויות שמורות לברק קנדל





© כל הזכויות שמורות לברק קנדל

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
Φ(z)	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995

© כל הזכויות שמורות לברק קנדל

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 השוקלות מעל 110 גרם?

3) מה אחוז חפיסות השוקולד השוקלות מתחת ל 92 גרם?

4) מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

© כל הזכויות שמורות לברק קנדל

שאלות:

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 וסטיית תקן 400 גרם.

א. מצאו את העשירון העליון.

ב. מצאו את האחוזון ה-95.

ג. מצאו את העשירון התחתון.

© כל הזכויות שמורות לברק קנדל

5) ציוני מבחן אינטליגנציה מתפלגים נורמלית עם ממוצע 100 ושונות 225.

א. מה העשירון העליון של הציונים במבחן האינטליגנציה?

ב. מה העשירון התחתון של ההתפלגות?

ג. מהו הציון ש-20% מהנבחנים מקבלים מעליו?

ד. מהו האחוזון ה-20?

ה. מהו הציון ש-5% מהנבחנים מקבלים מתחתיו?

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

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

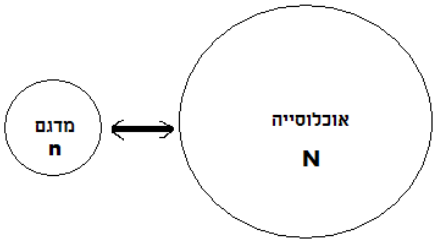
פרק 2 - הסקה סטטיסטית - הקדמה

תוכן העניינים

1. כללי 7

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.

© כל הזכויות שמורות לברק קנדל

הסימונים לפרמטר וסטטיסטי הם שונים:

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

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?

© כל הזכויות שמורות לברק קנדל

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.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב "העוגן".
נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. 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.

א. מהי האוכלוסייה?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. 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 בישוב 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.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 3 - התפלגות הדגימה ומשפט הגבול המרכזי

תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי..... 10

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: $V(\bar{x})=\sigma^2_{\bar{x}}=\frac{\sigma^2}{n}$.
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 סטית התקן של ממוצע המדגם?

© כל הזכויות שמורות לברק קנדל

דגימה מהתפלגות נורמאלית:

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

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

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?

© כל הזכויות שמורות לברק קנדל

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.

א. מיהי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו תוחלת ממוצע המדגם?

ו. מהי טעות התקן?

2) 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?

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.

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 מתחולת הגבהים בפחות מ-5 ס"מ?

ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

© כל הזכויות שמורות לברק קנדל

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 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?

5) 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. בקבוקיי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

6) 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?

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 ₪?

© כל הזכויות שמורות לברק קנדל

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-$X \sim N\left(\mu,\sigma^2\right)$. דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע המדגם $\bar{X}$. לכן: $P\left(\bar{X}>\mu\right)$ 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. $\bar{X} \sim N\left(\mu,\sigma^2\right)$.

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

11) נתון ש-$X \sim N\left(\mu,\sigma^2\right)$. 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.

© כל הזכויות שמורות לברק קנדל

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'.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 4 - מושגי יסוד באמידה

תוכן העניינים

1. כללי 16

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-$\hat{\theta}$. $\hat{\theta}$ הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את θ.
- שגיאת אמידה: $\left|\hat{\theta}-\theta\right|$ - 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: $E\left(\hat{\theta}\right)=\theta$.
- טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma\left(\hat{\theta}\right)=S.E$

© כל הזכויות שמורות לברק קנדל

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 אומר 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 אומר 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: 3, 1, 2, 5, 4, 1, 2, 3, 1, 2.
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.

© כל הזכויות שמורות לברק קנדל

שאלות:

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 האומד 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 האומד?

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. מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.

ב. מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.

ג. 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: $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X^2_i = 1502.2$.

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.

© כל הזכויות שמורות לברק קנדל

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) $X_1, X_2, ..., X_{16}$ 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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 0.25 ב. 0.019 ג. 0.24 ד. 0.01
2) א. אוכלוסייה: מקררים של יצרן, תוחלת: 2400, סטיית תקן: 500.
ב. 100 ג. 2342 ד. 58
3) א. 177.9 ב. 64.1 ג. 0.4
4) א. 8.1 ב. 3.16
5) ד'.
6) ג'.
7) ד'.
8) ד'.
9) ג'.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 5 - רווח סמך לתוחלת (ממוצע)

תוכן העניינים

1. רווח סמך כששונות האוכלוסיה ידועה 21
2. קביעת גודל מדגם 26
3. רווח סמך כששונות האוכלוסיה לא ידועה 28

רווח סמך כששונות האוכלוסייה ידועה:

רקע:

ממוצע המדגם הוא אומד לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמתי הוא אפסי.
מה שנהוג לעשות כדי לאמוד את ממוצע האוכלוסייה, זה לבנות רווח סמך.
נבנה מרווח בטחון שהסיכוי שהפרמטר μ ייכלל בתוכו הוא: $1-\alpha$.
$1-\alpha$: נקרא רמת בטחון או רמת סמך. כך ש: $P(A \le \mu \le B) = 1-\alpha$.
A - גבול התחתון של רווח הסמך.
B - הגבול העליון של רווח הסמך.
$L = B - A$ - אורך רווח הסמך.

דוגמה (פתרון בהקלטה):

חוקר דגם 25 חיילים שנבחנו במבחן הפסיכומטרי. הוא בנה רווח סמך לממוצע הציונים במבחן הפסיכומטרי בקרב אוכלוסיית החיילים וקיבל בין 510 ל-590.
רווח הסמך נבנה ברמת סמך של 95%.

1. מהי אוכלוסיית המחקר?
2. מה המשתנה באוכלוסייה?
3. מה הפרמטר שהחוקר רצה לאמוד?
4. מהו רווח הסמך?
5. מה אורך רווח הסמך?
6. מהי רמת הביטחון של רווח הסמך?

© כל הזכויות שמורות לברק קנדל

בפרק זה נרצה לבנות רווח סמך לתוחלת (μ) במקרה ש-σ^2 (שונות האוכלוסייה) ידועה.

פרמטר אותו נרצה לאמוד: μ.

אומד נקודתי: $\bar{x}$.

תנאים לבניית רווח הסמך: $X \sim N$ או $n \geq 30$.

σ^2 (שונות האוכלוסייה) ידועה.

נוסחה לרווח הסמך: $\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

דוגמה (פתרון בהקלטה):

על פי נתוני היצרן אורך חיי סוללה מתפלג נורמאלית עם סטיית תקן של 1 שעה. מעוניינים לאמוד את תוחלת חיי סוללה. נדגמו באקראי 4 סוללות, אורך החיים הממוצע שהתקבל הוא 13.5 שעות. בנו רווח סמך ברמת סמך של 95% לתוחלת אורך חיי סוללה.

שגיאת האמידה המקסימלית: $\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

ε - נותן את שגיאת האמידה המקסימלית, דבר שנקרא גם טעות סטטיסטית, טעות דגימה.

דוגמה (פתרון בהקלטה):

בהמשך לשאלה עם הסוללות. מה ניתן להגיד בביטחון של 95% על שגיאת האמידה?

קשרים מתמטיים ברווח הסמך:

- אורך רווח הסמך הוא פעמיים שגיאת האמידה המקסימלית: $L = 2\varepsilon$.
- ממוצע המדגם נופל תמיד באמצע רווח הסמך: $\bar{X} = \frac{A+B}{2}$.
- ככל שמספר התצפיות (n) גבוה יותר, כך יש יותר אינפורמציה ולכן האומד יותר מדויק, ולכן נקבל רווח סמך יותר קצר.
- ככל שרמת הביטחון ($1-\alpha$) גבוהה יותר, כך: $z_{1-\frac{\alpha}{2}}$ יותר גבוה, ורווח הסמך יותר ארוך.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800 ₪.

א. מי האוכלוסייה במחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר שאותו רוצים לאמוד?

ד. מה רווח הסמך לפרמטר?

ה. מהי רמת הסמך לפרמטר?

ו. מה אורך רווח הסמך?

ז. מה הסיכוי שטעות הדגימה תעלה על 300 ₪?

2) מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנו את רווח הסמך.

3) מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלית עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.

א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.

ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.

ג. הסבירו כיצד ומדוע השתנה רווח הסמך.

4) דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.

א. בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.

ב. מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?

ג. מה היה צריך להיות גודל המדגם אם היינו רוצים להקטין את רווח הסמך ב-50%?

ד. אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?

© כל הזכויות שמורות לברק קנדל

5) בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (105,99). שחזרו את:

א. ממוצע המדגם.

ב. שגיאת האמידה המקסימאלית.

ג. רמת הסמך.

6) זמן החלמה מאנגינה מתפלג עם סטיית תקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.

א. בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.

ב. מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.

ג. מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.

7) חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא: $82 < \mu < 92$.
נתון שסטיית התקן בהתפלגות שווה ל-10 ושהמדגם מתבסס על 16 תצפיות. התפלגות המשתנה היא נורמאלית.

א. מהו ממוצע המדגם?

ב. מהי רמת הסמך של רווח הסמך שנבנה?

ג. מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5?

8) חוקר בנה רווח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותם נתונים רווח סמך ברמת סמך קטנה מ-95%, איזה מהמשפטים הבאים לא יהיה נכון.

א. אורך רווח הסמך החדש יהיה קטן יותר.

ב. גודל המדגם יהיה כעת קטן יותר.

ג. המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.

ד. רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.

© כל הזכויות שמורות לברק קנדל

9) חוקר בנה רווח סמך ל- μ וקיבל: $48 < \mu < 54$. מה נכון בהכרח:

א. $\mu = 51$.

ב. $\overline{X} = 6$.

ג. $\overline{X} = 51$.

ד. אורך רווח הסמך הינו 3.

10) איזה מהגורמים הבאים אינו משפיע על גודלו של רווח בר סמך, כאשר שונות האוכלוסייה ידועה (בחרו בתשובה הנכונה):

א. רמת הביטחון.

ב. סטיית התקן באוכלוסייה.

ג. מספר המשתתפים.

ד. סטיית התקן במדגם.

תשובות סופיות:

1) א. העובדים במשק. ב. שכר ב-₪. ג. μ. ד. $9200 < \mu < 9800$.
ה. 0.95. ו. 600. ז. 0.05.

2) $4920.6 < \mu < 4979.4$.

3) א. $223.42 < \mu < 236.58$. ב. $222.16 < \mu < 237.84$.
ג. ראה סרטון.

4) א. $9284 < \mu > 10,116$. ב. הסטיה המירבית בין $\bar{x}$ ל- μ היא 416 ₪ בבטחון של 95%.
ג. 800. ד. לא.

5) א. 102. ב. 3. ג. 0.9544.

6) א. $83.5 < \mu < 4.42$. ב. יקטן פי 2. ג. גדל.

7) א. 87. ב. 5. ג. 0.9544.

8) ב'.

9) ג'.

10) ד'.

© כל הזכויות שמורות לברק קנדל

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

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 האמתי לא יעלה על 0.5 ק"ג? (תשובה: 87).

© כל הזכויות שמורות לברק קנדל

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.

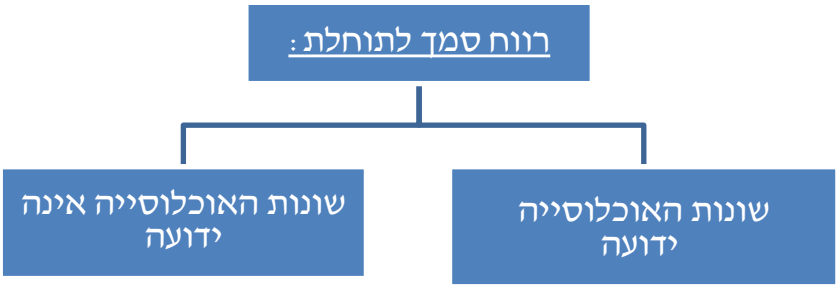
3) $n=62$.

© כל הזכויות שמורות לברק קנדל

רווח סמך כששונות האוכלוסייה לא ידועה:

רקע:

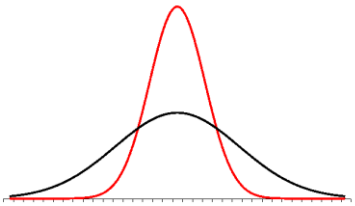
בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:



בפרק זה נעסוק במקרה ששונות האוכלוסייה $\left(\sigma^2\right)$ אינה ידועה לנו.

מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול.



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

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

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: 5.3, 4.6, 5.2, 4.7.
בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

© כל הזכויות שמורות לברק קנדל

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 האמיתית 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: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X^2_i = 643.19$.
בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

© כל הזכויות שמורות לברק קנדל

6) נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו: $S=2$, $\bar{x}=13.8$.
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 $\sigma=20$ 2.
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.

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'.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 6 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

תוכן העניינים

1. כששונויות האוכלוסיה לא ידועות ובהנחת שוויון שונויות..... 31

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. $\sigma^2_1=\sigma^2_2$.
2. $X_1, X_2 \sim N$.
3. מדגמים בלתי תלויים.

השונות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונות הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי הנוסחה הבאה: $S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

דרגות החופש: $d.f=n_1+n_2-2$.

רווח סמך: $(\bar{x}_1-\bar{x}_2)\pm t^{n_1+n_2-2}_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{S^2_p}{n_1}+\frac{S^2_p}{n_2}}$.

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:

	תל אביב	באר שבע
מספר האקדמאים	20	10
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.

© כל הזכויות שמורות לברק קנדל

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: $N\left(\mu_x,\sigma^2\right)$, ומשתנה Y שמתפלג: $N\left(\mu_y,\sigma^2\right)$.

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:
1. השונות שווה.
2. שהציונים מתפלגים נורמלית.
3. המדגמים אינם תלויים זה בזה.

$-5.52<\mu_1-\mu_2<17.52$.

2) $-9.6<\mu_y-\mu_x<1.6$.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 7 - רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים

תוכן העניינים

1. רווח סמך לתוחלת (ממוצע) ההפרשים במדגמים מזווגים 33

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. $x, y \sim N$.
2. המדגם מזווג.

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

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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) $-0.013 < \mu < 0.185$.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 8 - מבוא לבדיקת השערות על פרמטרים

תוכן העניינים

1. הקדמה 35
2. סוגי טעויות 39

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 המסומנת 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 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. אזור הדחיה 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.

© כל הזכויות שמורות לברק קנדל

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:

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. נבחנים בבגרות באנגלית.
ב. ציון.
ג. ממוצע הציונים בשיטת לימוד חדשה.
ד. $\begin{matrix} H_0:\mu=72 \\ H_1:\mu>72 \end{matrix}$.

2) א. משקאות בבקבוק של חברה מסוימת.
ב. נפח משקה בסמ"ק.
ג. ממוצע נפח המשקה בבקבוק.
ד. $\begin{matrix} H_0:\mu=500 \\ H_1:\mu<500 \end{matrix}$.

3) א. מועמדים לפקולטה למשפטים.
ב. משתנה דיכוטומי (התקבל, לא התקבל).
ג. אחוז הקבלה.
ד. $\begin{matrix} H_0:p=0.25 \\ H_1:p<0.25 \end{matrix}$.

4) א. אזרחים בוגרים במשק.
ב. משתנה דיכוטומי (מובטל, עובד).
ג. אחוז האבטלה כיום.
ד. $\begin{matrix} H_0:p=0.08 \\ H_1:p\neq 0.08 \end{matrix}$.

© כל הזכויות שמורות לברק קנדל

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
מציאות	H_0	אין טעות	טעות מסוג 1
	H_1	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?

© כל הזכויות שמורות לברק קנדל

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.

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?

תשובות סופיות:

1) א. $\begin{matrix} H_0:\mu=500 \\ H_1:\mu<500 \end{matrix}$. 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. $\begin{matrix} H_0:\mu=2.3 \\ H_1:\mu<2.3 \end{matrix}$.

ה. לא לדחות את H_0. ו. טעות מסוג שני.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

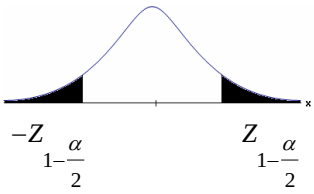
פרק 9 - בדיקת השערות על תוחלת (ממוצע)

תוכן העניינים

1. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה ידועה..... 41
2. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה ידועה)..... 45
3. בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסיה לא ידועה..... 50
4. מובהקות תוצאה - אלפא מינימלית (ששונות האוכלוסיה לא ידועה)..... 54
5. הקשר בין רווח סמך לבדיקת השערות על תוחלת (ממוצע)..... 57
6. ניתוח פלטים..... 59

בדיקת השערות על תוחלת (ממוצע) כששונות האוכלוסייה ידועה:

רקע:

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

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

חלופה אחרת לכלל הכרעה:

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

© כל הזכויות שמורות לברק קנדל

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

פיתרון:

<u>אוכלוסייה</u>: עגבניות.

<u>המשתנה</u>: X = יבול העגבניות בטון לעונה.

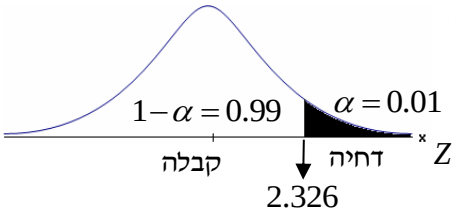
<u>הפרמטר</u>: μ = תוחלת היבול בשיטת הזיבול החדשה.

<u>השערות</u>: $H_0 : \mu = 10$
$H_1 : \mu > 10$

<u>תנאים</u>:

1. $X \sim N$.
2. $\sigma = 2.5$.

<u>כלל הכרעה</u>:



נדחה את H_0 אם $Z_{\bar{x}} > 2.326$

<u>תוצאות</u>: $n = 4$, $\bar{x} = 12.5$

<u>סטטיסטי המבחן</u>: $Z_{\bar{x}} = \dfrac{\bar{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$.

<u>נציב</u>: $Z_{\bar{X}} = \dfrac{1.25 - 10}{\dfrac{2.5}{\sqrt{4}}} = 2 < 2.326$

<u>מסקנה</u>:

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.

© כל הזכויות שמורות לברק קנדל

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 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?
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.

© כל הזכויות שמורות לברק קנדל

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, ברי"מ של 5% לא נקבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.

2) א. נדחה H_0, ברי"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, ברי"מ של 5% נקבע שהמכונה לא מאופסת.

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

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

6) א'.

7) ג'.

8) א. לדחות. ב. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

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 $p_v \le \alpha$, 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 אלטרנטיבה:**	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu<\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu>\mu_0$
תנאים:	1. σ ידועה 2. $X\sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x}>\mu_0 \Rightarrow 2\cdot P_{H_0}\left(\bar{X}\ge\bar{x}\right)$ אם $\bar{x}<\mu_0 \Rightarrow 2\cdot P_{H_0}\left(\bar{X}\le\bar{x}\right)$	$P_{H_0}(\bar{X}\le\bar{x})$	$P_{H_0}(\bar{X}\ge\bar{x})$

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

© כל הזכויות שמורות לברק קנדל

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 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%?

פתרון:

א. <u>אוכלוסייה</u>: המתגייסים לצבא כיום.

<u>משתנה</u>: X = משקל בק"ג.

<u>פרמטר</u>: μ.

<u>השערות</u>: $\begin{matrix} H_0:\mu=65 \\ H_1:\mu>65 \end{matrix}$.

<u>תנאים</u>:

1. $X\sim N$.
2. $\sigma=12$.

<u>תוצאות מדגם</u>:

$n=16$

$\bar{X}=71$

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

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

$\alpha_{\min}=0.0228$

© כל הזכויות שמורות לברק קנדל

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: $\bar{x}=74$, $n=100$.
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 נורמאלית 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.

© כל הזכויות שמורות לברק קנדל

6) בבדיקת השערות התקבל שה- $p-value=0.02$.
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: $p\ value=0.0254$ 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 .

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) 0.0228.

2) עבור כל רמת מובהקות סבירה.

3) א. $\begin{matrix} H_0:\mu=100 \\ H_1:\mu<100 \end{matrix}$. 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'.

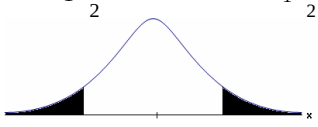
7) א'.

8) ג'.

© כל הזכויות שמורות לברק קנדל

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

רקע:

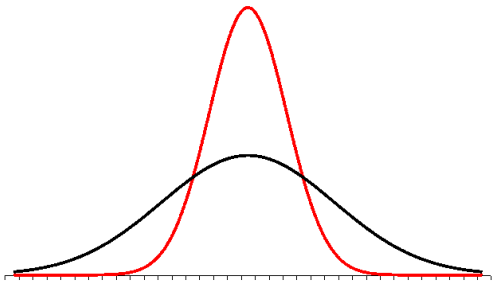
השערת האפס: השערה אלטרנטיבה:	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu<\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu>\mu_0$
תנאים:	1. σ אינה ידועה 2. $X\sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0:	$t_{\bar{x}}>t^{(n-1)}_{1-\frac{\alpha}{2}}$ או $t_{\bar{x}}<-t^{(n-1)}_{1-\frac{\alpha}{2}}$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

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

$S^2=\dfrac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1}=\dfrac{\sum_{i=1}^{n}X^2{}_i-n\bar{X}^2}{n-1}$.

© כל הזכויות שמורות לברק קנדל

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%.

© כל הזכויות שמורות לברק קנדל

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: 125, 80, 100, 95, 90 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:

$$\sum_{i=1}^{10} X_i = 750$$
$$\sum_{i=1}^{10} \left(X_i - \bar{X} \right)^2 = 900$$

נתון שההתפלגות היא נורמלית.
בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80 .

© כל הזכויות שמורות לברק קנדל

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: $\begin{matrix} H_0:\mu=\mu_0 \\ H_1:\mu<\mu_0 \end{matrix}$.
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'.

© כל הזכויות שמורות לברק קנדל

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

רקע:

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם $p_v \le \alpha$ 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.

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

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

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

$$d.f = n-1$$

© כל הזכויות שמורות לברק קנדל

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: 27, 32, 30, 40, 34.
הניחו שזמן הנסיעה מתפלג נורמלית.

א. רשמו את השערות המחקר.

ב. מצאו חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5%?

פתרון:

<u>אוכלוסייה</u>: כלל הנסיעות לעבודה בדרך החלופית.

<u>משתנה</u>: X = זמן נסיעה בדקות.

<u>תנאים</u>: $X \sim N$.

<u>פרמטר</u>: μ.

א. <u>השערות</u>:
$$\begin{matrix} H_0: \mu = 40 \\ H_1: \mu < 40 \end{matrix}.$$

ב. <u>תוצאות המדגם</u>:

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

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

$$S=\sqrt{23.4}$$

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

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

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

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

$P_V<\alpha=0.05$, לכן דוחים את H_0.

<u>מסקנה</u>: ברי"מ של 5% נכריע שהדרך החלופית מהירה יותר.

© כל הזכויות שמורות לברק קנדל

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:
997, 996, 1005, 1008, 1024.

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:
$\sum(x_i-\bar{x})^2=2832$, $\bar{x}=176.2$.
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) א. $H_0:\mu=1000$, $H_1:\mu\neq1000$. ב. $20\%\le P_v\le 50\%$.

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

2) 10%.

3) 1.01, נקבל את H_0.

© כל הזכויות שמורות לברק קנדל

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 ← נקבל את H_0 .

אם μ_0 לא נופל ברווח ← נדחה את H_0 .

דוגמה:

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו:

$H_0:\mu=80$, $H_1:\mu\neq80$, $\alpha=5\%$.

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 .

© כל הזכויות שמורות לברק קנדל

שאלות:

1) חוקר רצה לבדוק את ההשערות הבאות: $H_1:\mu\neq 90$, $H_0:\mu=90$.
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 פנצילין 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.

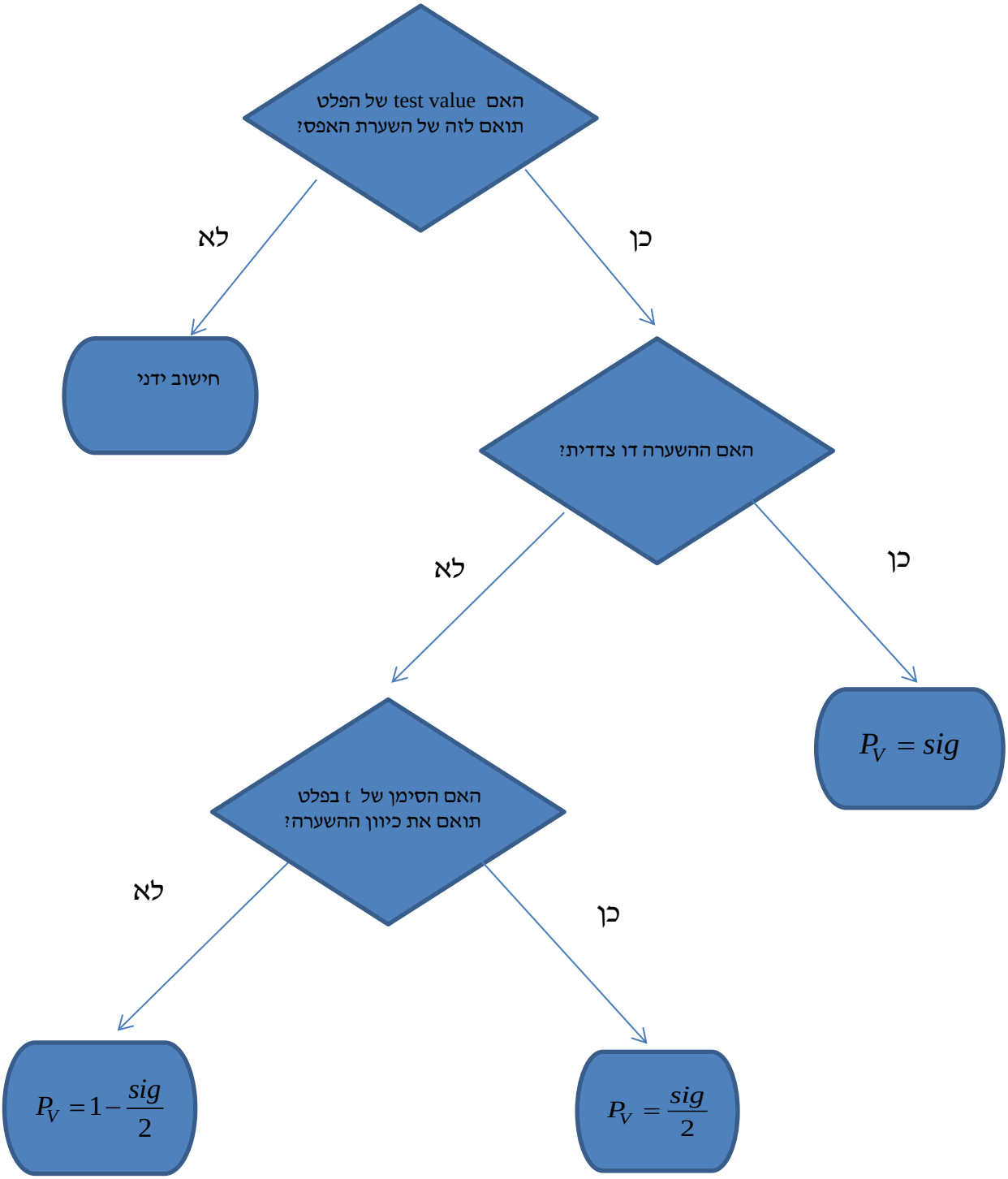
3) א. $191.8\leq\mu\leq 200.2$. ב. נכריע שיש אמת בפרסום.

© כל הזכויות שמורות לברק קנדל

ניתוח פלטים:

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS:



© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלית. 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					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean 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					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean 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.

© כל הזכויות שמורות לברק קנדל

3) להלן 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. השלימו 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.
הניחו 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 היינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
ג. בדקו האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ-26 ברמת מובהקות של 5%.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. הסקה של תוחלת אחת. ב. 23. ג. (22.104, 24.771).

ד. נקבל H_0.

2) א. 25. ב. (4.3056, 4.5104). ג. המשתנה מתפלג נורמלית.

ד. נדחה H_0.

3) א. $n=44$, $\bar{X}=67.66$, $t=-2.48$, $upper=-1.3735$.

ב. 0.017. ג. 0.0085. ד. 0.9915.

4) א. נכריע שיש אמת בפרסום. ב. המסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

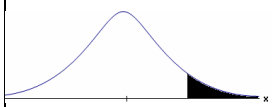
פרק 10 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

תוכן העניינים

1. כששוניויות האוכלוסיה לא ידועות ומניחים שהן שוות..... 64
2. ניתוח פלטים..... 68

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

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

סטטיסטי המבחן:

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

השונות המשוקללת:

$$S_p{}^2=\frac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{n_1+n_2-2}$$

© כל הזכויות שמורות לברק קנדל

חלופה אחרת לכלל הכרעה:

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

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%.

© כל הזכויות שמורות לברק קנדל

שאלות

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

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

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 W 60 דולקות בממוצע יותר מאשר נורות מסוג 100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W 60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

© כל הזכויות שמורות לברק קנדל

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%.

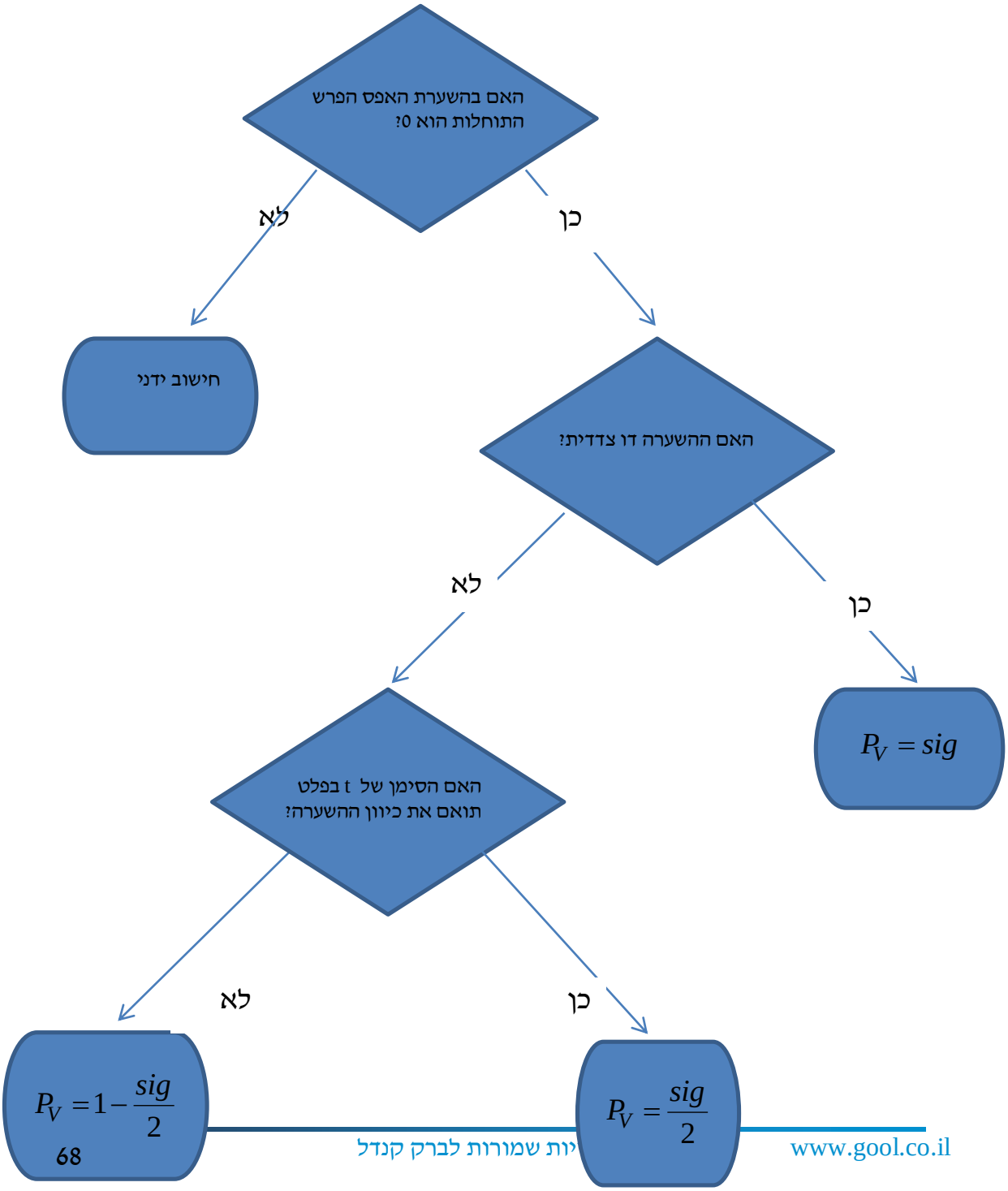
ג. לא נדחה את H_0.

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

ניתוח פלטים – רקע

מובהקות התוצאה על סמך הפלט:



דוגמה (פתרון בהקלטה):

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						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error 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 שההפרש הממוצע בין-" South East" לבין-"North East" חיובי?

© כל הזכויות שמורות לברק קנדל

שאלות

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

Group Statistics

	GENDER	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 Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	1190.977	.053	.05500	.02838	-.00068	.11067

א. האם ניתן להניח ששונות 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%.

© כל הזכויות שמורות לברק קנדל

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

Group Statistics

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

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha = 0.05$.

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha = 0.01$.

© כל הזכויות שמורות לברק קנדל

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?

© כל הזכויות שמורות לברק קנדל

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						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-ailed)	Mean Difference	Std. Error 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. השלימו 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?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. נקבל את H_0 ונכריע שיש שוויון שוניות.

ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.

ג. 0.11076 ד. $2.5665 \le \mu \le 2.6441$.

2) א. מבחן T להפרש ממוצעים במדגמים בלתי תלויים.

ב. מבחן T למדגמים מזווגים.

ג. נקבל את H_0 , נקבע לקיום שוויון שוניות.

ד. נקבל את H_0 , לא נקבל את טענת הפסיכולוגים.

3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.

ב. $1.19497 \le \mu_1 - \mu_2 \le 2.39332$ ג. כן.

4) א. 10.15, 4.69,-48 ב. 0.04

ג. 0.978 ד. 0.022.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

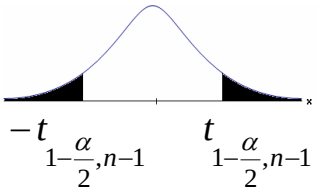
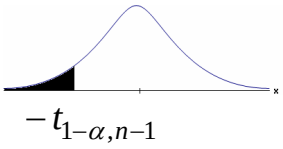
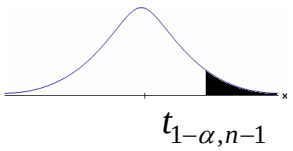
פרק 11 - בדיקת השערות לתוחלת ההפרש במדגמים מזווגים

תוכן העניינים

1. בדיקת השערות למדגמים מזווגים 75
2. ניתוח פלטים 79

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

בדיקת השערות למדגמים מזווגים – רקע

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

סטטיסטי המבחן: $t_{\bar{D}} = \dfrac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$, $S_D^2 = \dfrac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1} = \dfrac{\sum_{i=1}^{n} D_i^2 - n\bar{D}^2}{n-1}$

דוגמה (פתרון בהקלטה):

© כל הזכויות שמורות לברק קנדל

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה 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".

© כל הזכויות שמורות לברק קנדל

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 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 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 בממוצע 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 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.

© כל הזכויות שמורות לברק קנדל

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	4.5	9.6	0.7	2.5
משקל במכשיר 2	3.5	6.9	1.7	0.5

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 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:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

תשובות סופיות

1) לא נדחה H_0.

2) לא נדחה H_0.

3) א. $0.25 \le pv \le 0.5$ ב. 0.5 ג. לא נדחה H_0.

4) ד'.

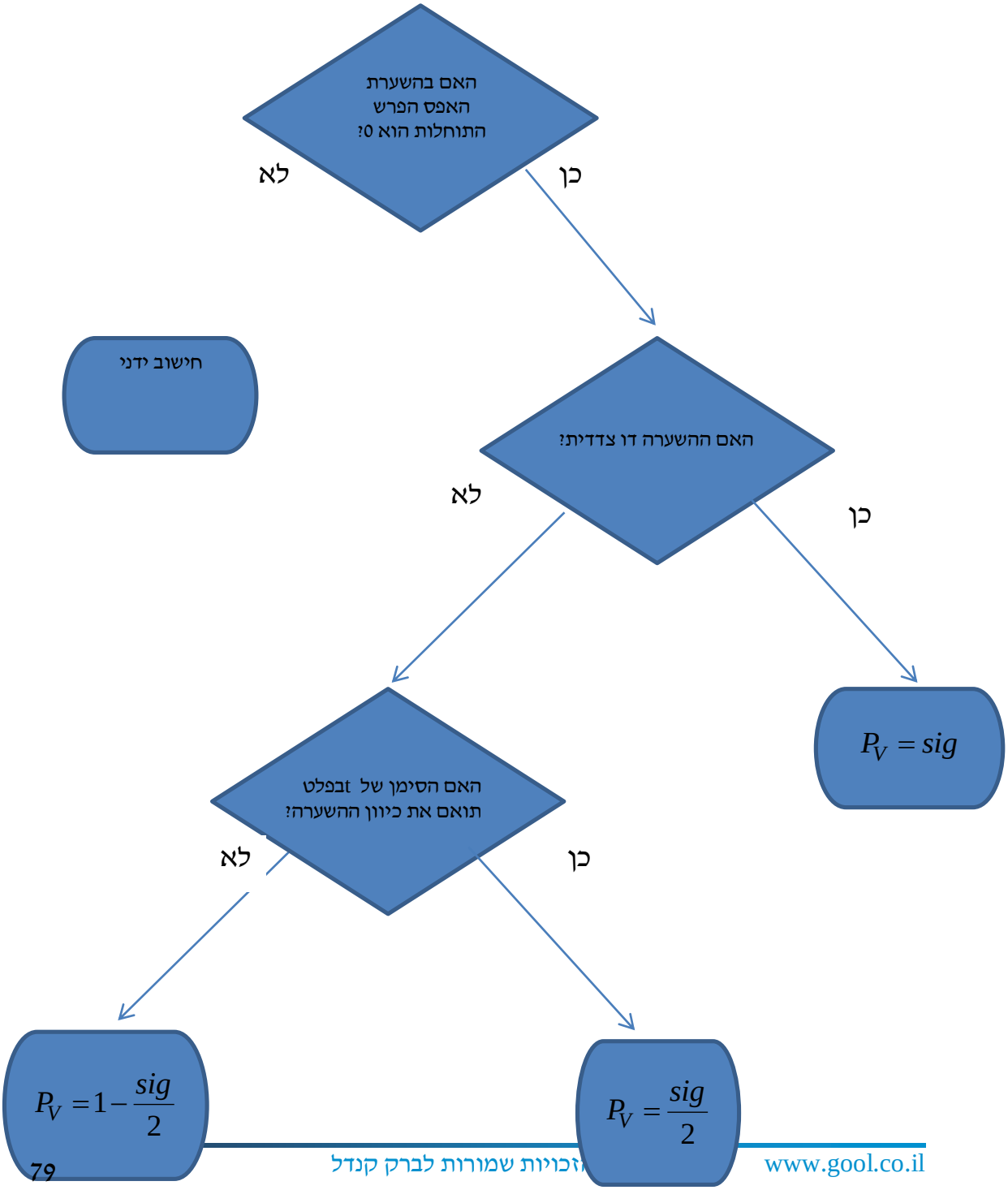
5) ד'.

6) ג'.

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

מדגמים מזווגים – ניתוח פלטים – רקע



כל הזכויות שמורות לברק קנדל

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

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

© כל הזכויות שמורות לברק קנדל

שאלות

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

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	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	973	.943

א. תנו אומדן להפרש הממוצעים.

ב. תנו 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:

Paired Samples Statistics

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

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	Italy - South Korea	-.42233	.36153	.02087	-.46341	-.38126	-20.234	???	???

א. השלימו את החלקים החסרים בפלט (מסומנים בסימני שאלה).

ב. בדקו 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. בנו את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

© כל הזכויות שמורות לברק קנדל

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. רשמו מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.

ב. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.

ג. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.

ד. מצאו רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

© כל הזכויות שמורות לברק קנדל

4) להלן פלט של תכנת SPSS:

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%.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. -0.007 ב. 0.1 ג. אין הבדל מובהק.

2) א. $d.f = 299$, $n = 300$, $\bar{X} = 8.496$, $Sig = 0$. ב. נדחה את H_0. ג. $(0.46341, 0.38126)$ ד. $(0.4708, 0.3738)$.

3) א. ראה וידאו. ב. נדחה את H_0. ג. לא נדחה את H_0. ד. (1.592,13.317).

4) א. 1.5876, 11.5715 ב. 0.173 ג. יגדל. ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 12 - הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רווח סמך לבדיקת השערות להפרש תוחלות..... 85

הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע

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

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

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

אם C נופל ברווח ← נקבל את H_0 .

אם C לא נופל ברווח ← נדחה את H_0 .

דוגמה (פתרון בהקלטה) :

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג.

להלן השערותיו : $\alpha=5\%$, $H_1:\mu_D\neq 80,\qquad H_0:\mu_D=80$.

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

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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 מתפלגים נורמלית בנו רווח סמך למוצע 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.

תשובות סופיות

1) א. $-3.8 \le \mu_D \le 19$ ב. נכריע שיש אמת בפרסום.

2) א. $-8.5 \le \mu_X - \mu_Y \le 26.5$ ב. נכריע שאין הבדל.

3) ג.

4) א.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 13 - שאלות מסכמות בבדיקת השערות

תוכן העניינים

1. שאלות רב ברירה (אמריקאיות) 88

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) 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.

3) 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:

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

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

א. $H_1: \mu_d < 0, \quad H_0: \mu_d = 0$

ב. $H_1: \mu_X - \mu_Y < 0, \quad H_0: \mu_X - \mu_Y = 0$

ג. $H_1: \mu_X - \mu_Y < 0, \quad H_0: \mu_X - \mu_Y = 0$

ד. $H_1: \mu_d > 0, \quad H_0: \mu_d = 0$

4) 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 מהתושבות לא נכונה בהכרח.

© כל הזכויות שמורות לברק קנדל

5) 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.

א. מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
ב. מבחן T למדגם יחיד ; מבחן T למדגמים תלויים.
ג. מבחן T למדגם יחיד ; מבחן T למדגמים בלתי תלויים.
ד. מבחן T למדגמים בלתי תלויים ; מבחן T לממוצע יחיד.

6) 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.

7) 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 :

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

באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש :

א. ההתפלגות Z ללא דרגות חופש.
ב. ההתפלגות T ו-3 דרגות חופש.
ג. ההתפלגות T ו-6 דרגות חופש.
ד. ההתפלגות χ^2 ו-3 דרגות חופש.

© כל הזכויות שמורות לברק קנדל

8) 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'?

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. ללא נתונים נוספים אי אפשר לדעת מה יחליט.

9) חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?

א. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

ב. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

ג. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.

ד. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

10) 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?

א. טעות מסוג ראשון.

ב. טעות מסוג שני.

ג. טעות מסוג שלישי.

ד. לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחוניים.

© כל הזכויות שמורות לברק קנדל

11) 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.

12) 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.

13) 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.

14) 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?

א. מבחן T למדגמים בלתי תלויים עם 38 דרגות חופש.

ב. מבחן T למדגמים מזווגים, עם 39 דרגות חופש.

ג. מבחן T למדגמים בלתי תלויים עם 39 דרגות חופש.

ד. מבחן T למדגמים מזווגים עם 38 דרגות חופש.

© כל הזכויות שמורות לברק קנדל

15) 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:

א. שלושתם במבחן T למדגמים בלתי תלויים.
ב. עידו במבחן T למדגמים מזווגים, ורון ויואב במבחן T למדגמים בלתי תלויים.
ג. עידו במבחן T למדגמים מזווגים, רון במבחן T למדגמים בלתי תלויים ויואב במבחן T למדגם יחיד.
ד. עידו במבחן T למדגמים מזווגים, רון ויואב במבחן T למדגם יחיד.

16) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%.
מה תמיד נכון?

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.

17) 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:

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

18) 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?

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.

© כל הזכויות שמורות לברק קנדל

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

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

20) 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.

21) 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 מהתושבות לא נכונה בהכרח.

22) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

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

23) 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.

© כל הזכויות שמורות לברק קנדל

24) 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?

א. טעות מסוג ראשון.
ב. טעות מסוג שני.
ג. טעות מסוג שלישי.
ד. אין טעות במסקנתו.

25) בבדיקת השערות התקבל שה- $p-value=0.02$.
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.

26) 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.

27) בבדיקת השערות מסוימת התקבל $p\ value=0.0254$, 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 .

28) 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.

© כל הזכויות שמורות לברק קנדל

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

30) 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

31) 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	4.5	9.6	0.7	2.5
משקל במכשיר 2	3.5	6.9	1.7	0.5

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.

32) 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.

© כל הזכויות שמורות לברק קנדל

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

א. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

ב. תמיד לדחות את השערת האפס כאשר היא נכונה.

ג. לעולם לא לדחות את השערת האפס כאשר היא נכונה.

ד. תמיד לדחות את השערת האפס כאשר היא לא נכונה.

34) 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.

35) 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 מתפלגים 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.

36) 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.

© כל הזכויות שמורות לברק קנדל

37) חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות: $H_1:\mu\neq 10,\quad H_0:\mu=10$.
לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל: $t_{\bar{x}}=-2.63$. 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.

38) 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

39) 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 **<u>במציאות</u>** השערתה של החוקרת נכונה, <u>סביר</u> 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.

40) איזה מהמשפטים הבאים נכון תמיד?

א. $\text{POWER}+\alpha+\beta=1$

ב. $\text{POWER}=0.5-\beta$

ג. $\text{POWER}+\alpha=1$

ד. $\beta+\alpha=1$

ה. הכול לא נכון.

© כל הזכויות שמורות לברק קנדל

41) 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.

42) 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.

43) 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

44) 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.

45) מה מהבאים **<u>נכון</u>** 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.

© כל הזכויות שמורות לברק קנדל

46) לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית $H_0:\mu\geq 10$, $H_1:\mu<10$. 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.

47) לבדיקת ההשערה החד צדדית ימנית $H_0:\mu=55,\quad H_1:\mu=65$, 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: $H_0:\mu=55,\quad H_1:\mu=60$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $H_0:\mu=55,\quad H_1:\mu=65$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות: $H_0:\mu=55,\quad H_1:\mu=65$ 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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

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

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 14 - מבחני חי בריבוע

תוכן העניינים

1. מבחן לאי תלות. (ללא ספר)
2. מקדם המתאם של קרמר. 101
3. ניתוח פלטים במבחן אי תלות 103

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/נמנעים באיזשהו נושא. להלן טבלת השכיחויות המשותפת שהתקבלה.

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

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 $E_i = \frac{f(x)\cdot f(y)}{n}$.

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

שלב ב':

נחשב $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$.

שלב ג':

נחשב: $r_c = \sqrt{\frac{1}{n(L-1)}\chi^2}$.

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

© כל הזכויות שמורות לברק קנדל

שאלות

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

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

האם קיים קשר בין מין להשכלה? נמקו!

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. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.

ב. האם קיים קשר בין פעילות גופנית למצב בריאותי?
חשבו לפי מדד הקשר של קרמר.

תשובות סופיות

1) קיים קשר בעוצמה בינונית בין המין להשכלה מקדם המתאם של קרמר הוא 0.595.

2) א. להלן טבלה: ב. מדד קרמר 0.19 מעיד על קשר בעוצמה נמוכה.

y / x	תקין	לא תקין	$f(x)$
כן	50	10	60
לא	90	50	140
$f(y)$	140	60	200

© כל הזכויות שמורות לברק קנדל

פלטים על מבחן לאי תלות – רקע

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים.
שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

מבנה המבחן:

השערות:

אין תלות בין המשתנים: H_0.

יש תלות בין המשתנים: H_1.

דרגות חופש: $d.f=(r-1)(c-1)$.

כאשר: r - מספר הקטגוריות של המשתנה שבשורות.

c - ספר הקטגוריות של המשתנה שבעמודות.

סטטיסטי המבחן: $\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$.

O_i - השכיחות נצפית במדגם בתא i.

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

$$E_i=\frac{f(x)\cdot f(y)}{n}$$

הערה:

תנאי כדי לבצע את המבחן הוא $E_i\geq 5$ לכל צ i. במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.
תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ 20% מהתאים E קטן מ-5.

© כל הזכויות שמורות לברק קנדל

דוגמה (פתרון בהקלטה) :

במחקר רצו לבדוק את הקשר בין צבע שיער לבין צבע עיניים של אנשים.
הפלטים שהתקבלו מצורפים.

א. מהו המבחן הסטטיסטי המתאים?

ב. כמה קטגוריות יש לכל משתנה?

ג. רשמו את השערות המחקר.

ד. מה מספר דרגות החופש?

ה. כמה אנשים במדגם נמצאו עם שיער חום?

ו. כמה אנשים היית מצפה במדגם שיהיה להם שיער חום ועיניים ירוקות בהנחה ואין קשר בין צבע שיער לצבע עיניים?

ז. מתוך הבלונדינים מה אחוז בעלי עיניים כחולות במדגם?

ח. מתוך בעלי עיניים ירוקות מה אחוז הבלונדינים במדגם?

ט. מה ערכו של סטטיסטי המבחן ומהי מובהקות התוצאה?

י. מה מסקנת המחקר? $\alpha = 5\%$

© כל הזכויות שמורות לברק קנדל

להלן הפלטים שהתקבלו:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
hair_color * eye_color	78	100.0%	0	0.0%	78	100.0%

hair_color * eye_color Crosstabulation

			eye_color			Total
			brown	green	Blue	
hair_color	black	Count	13	7	7	27
		Expected Count	10.7	8.3	8.0	27.0
		% within hair_color	48.1%	25.9%	25.9%	100.0%
		% within eye_color	41.9%	29.2%	30.4%	34.6%
	brown	Count	12	12	6	30
		Expected Count	11.9	9.2	8.8	30.0
		% within hair_color	40.0%	40.0%	20.0%	100.0%
		% within eye_color	38.7%	50.0%	26.1%	38.5%
	blond	Count	6	5	10	21
		Expected Count	8.3	6.5	6.2	21.0
		% within hair_color	28.6%	23.8%	47.6%	100.0%
		% within eye_color	19.4%	20.8%	43.5%	26.9%
	Total	Count	31	24	23	78
		Expected Count	31.0	24.0	23.0	78.0
		% within hair_color	39.7%	30.8%	29.5%	100.0%
		% within eye_color	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.880[a]	4	.208
Likelihood Ratio	5.641	4	.228
Linear-by-Linear Association	2.682	1	.101
N of Valid Cases	78		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.19.

© כל הזכויות שמורות לברק קנדל

שאלות

1) בסקר שנעשה על ידי משרד ראש הממשלה נדגמו 60 אזרחים.
כל אזרח נשאל על מגדרו והאם הוא בעד הקמת מדינה פלסטינית.

א. מה ההשערות הנבדקות ומהו סטטיסטי המבחן?

ב. אם סטטיסטי המבחן היה גדל כיצד הדבר היה משפיע על SIG שבפלט.

ג. האם קיים קשר בין מגדר ודעה ברמת מובהקות של 5%?

ד. מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?

ה. איזה אחוז מהנשאלים שהיו בעד מדינה פלשתינית הם גברים?

© כל הזכויות שמורות לברק קנדל

להלן הפלטים:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.973[a]	2	.373
Likelihood Ratio	1.987	2	.370
Linear-by-Linear Association	1.882	1	.170
N of Valid Cases	60		

a. 0 cells (.0%) have expected count less than 5.

b. The minimum expected count is 7.25.

gender * opinion Crosstabulation

			opinion			Total
			yes	now	no opinion	
gender	Male	Count	10	10	9	29
		Expected Count	12.6	9.2	7.3	29.0
		% within gender	34.5%	34.5%	31.0%	100.0%
		% within opinion	38.5%	52.6%	60.0%	48.3%
		% of Total	16.7%	16.7%	15.0%	48.3%
	female	Count	16	9	6	31
		Expected Count	13.4	9.8	7.8	31.0
		% within gender	51.6%	29.0%	19.4%	100.0%
		% within opinion	61.5%	47.4%	40.0%	51.7%
		% of Total	26.7%	15.0%	10.0%	51.7%
Total		Count	26	19	15	60
		Expected Count	26.0	19.0	15.0	60.0
		% within gender	43.3%	31.7%	25.0%	100.0%
		% within opinion	100.0%	100.0%	100.0%	100.0%
		% of Total	43.3%	31.7%	25.0%	100.0%

© כל הזכויות שמורות לברק קנדל

2) להלן פלט על סמך סקר שנעשה בקרב סטודנטים, בסקר נשאלו הסטודנטים על המוזיקה אותה הם מעדיפים וצורת הבילוי המועדפת עליהם.

Crosstab

Count

		בילוי			Total
		קריאה	ספורט	מועדון	
מוזיקה	רוק	0	0	11	11
	פופ	1	6	8	15
	קלאסי	5	6	9	20
Total		6	12	28	46

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.929[a]	?	.018
N of Valid Cases	46		

a. 5 cells (55.6%) have expected count less than 5.

b. The minimum expected count is 1.43.

א. בין אלו משתנים נבדק הקשר? כמה קטגוריות לכל משתנה?

ב. האם התנאים של המודל מתקיימים?

ג. מה מספר דרגות החופש במבחן הנ"ל?

ד. מה ההשערות של המבחן?

© כל הזכויות שמורות לברק קנדל

3) מחקר התעניין לבדוק את הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות. הפלטים מצורפים.

א. השלימו את שלושת המספרים החסרים בטבלה (היכן שיש סימני שאלה).

ב. מה ערכו של חי בריבוע הסטטיסטי.

ג. תנו הערכה למובהקות התוצאה לבדיקת הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות.

רמת_הכנסה * צרכן עגבניות Crosstabulation

			צרכן עגבניות		Total
			אורגני	לא אורגני	
רמת_הכנסה	הרבה מתחת לממוצע	Count	17	42	59
		% within רמת_הכנסה	28.8%	?	100.0%
		% within צרכן עגבניות	13.6%	33.6%	23.6%
	מתחת לממוצע	Count	27	22	49
		% within רמת_הכנסה	55.1%	44.9%	100.0%
		% within צרכן עגבניות	?	17.6%	19.6%
	ממוצע	Count	31	29	60
		% within רמת_הכנסה	51.7%	48.3%	100.0%
		% within צרכן עגבניות	24.8%	23.2%	24.0%
	מעל הממוצע	Count	44	26	70
		% within רמת_הכנסה	62.9%	37.1%	100.0%
		% within צרכן עגבניות	35.2%	20.8%	28.0%
	הרבה מעל הממוצע	Count	?	6	12
		% within רמת_הכנסה	50.0%	50.0%	100.0%
		% within צרכן עגבניות	4.8%	4.8%	4.8%
Total		Count	125	125	250

© כל הזכויות שמורות לברק קנדל

4) חוקר בדק את הקשר בין צבע השיער לבין צבע העיניים בעזרת מבחן חי בריבוע בקרב 52 נבדקים. תוצאות המבחן מוצגות בטבלה. בנוסף ידוע כי סטטיסטי המבחן שהתקבל מעיבוד הנתונים הוא 8.08.

א. מה תהיה מסקנת המחקר ברמת מובהקות של 1%?

ב. מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.

ג. מה יהיה בקירוב ערכו של מקדם המתאם של קרמר?

ד. מהי פרופורציית בעלי צבע השיער הבהיר מקרב בעלי העיניים הירוקות?

להלן הפלט:

צבע עיניים * **צבע שיער** Crosstabulation

			צבע שיער		Total
			כהה	בהיר	
צבע עיניים	כחול	Count			
		% within צבע עיניים	50.0%	50.0%	100.0%
		% within צבע שיער	21.6%	53.3%	30.8%
		% of Total	15.4%	15.4%	30.8%
	חום	Count			
		% within צבע עיניים	83.3%	16.7%	100.0%
		% within צבע שיער	27.0%	13.3%	23.1%
		% of Total	19.2%	3.8%	23.1%
	ירוק	Count			
		% within צבע עיניים	79.2%	20.8%	100.0%
		% within צבע שיער	51.4%	33.3%	46.2%
		% of Total	36.5%	9.6%	46.2%
Total		Count			
		% within צבע עיניים	71.2%	28.8%	100.0%
		% within צבע שיער	100.0%	100.0%	100.0%
		% of Total	71.2%	28.8%	100.0%

© כל הזכויות שמורות לברק קנדל

5) במחקר מסוים רצו לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש בשנה האחרונה. חוקר א' בדק האם קיים הבדל בתוחלות ההוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיים הבדל בהתפלגות ההוצאה בין גברים לנשים.
הקטגוריות חולקו לשלוש קבוצות הוצאה.

א. איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.

ב. מה מסקנתו של חוקר א'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים).

ג. איזו טעות יכולה להיות במסקנתו של חוקר א'?
נסחו את הטעות במונחי השאלה.

ד. מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים).

ה. איזו טעות יכולה להיות במסקנתו של חוקר ב'? נסחו זאת במונחי השאלה.

ו. כיצד ניתן ליישב את מסקנות שני החוקרים?

להלן פלט ראשון:

T-Test

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
dimension1	female	40	2.9000	1.15025	.18187
	male	40	2.6000	2.52982	.40000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
expese	Equal variances assumed	16.805	.000	.683	78	.497	.30000	.43941	-.57479	1.17479
	Equal variances not assumed			.683	54.464	.498	.30000	.43941	-.58078	1.18078

© כל הזכויות שמורות לברק קנדל

להלן פלט שני:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * category	80	100.0%	0	.0%	80	100.0%

gender * category Crosstabulation

			category			Total
			a	b	c	
gender	Female	Count	2	30	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	5.0%	75.0%	20.0%	100.0%
		% within category	9.1%	71.4%	50.0%	50.0%
	Male	Count	20	12	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	50.0%	30.0%	20.0%	100.0%
		% within category	90.9%	28.6%	50.0%	50.0%
Total		Count	22	42	16	80
		Expected Count	22.0	42.0	16.0	80.0
		% within gender	27.5%	52.5%	20.0%	100.0%
		% within category	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.442[a]	2	.000
Likelihood Ratio	25.064	2	.000
N of Valid Cases	80		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. 1.973 ב. קטן . ג. לא נדחה H_0 .
ד. 34.5% ה. 38.5%.

2) א. בילוי מועדף ומוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
ב. לא ג. 4.
ד. H_0 אין תלות בין בילוי מועדף למוזיקה מועדפת.
H_1 יש תלות בין בילוי מועדף למוזיקה מועדפת.

3) א. 71.2%, 21.6%, 6 ב. 15.8 ג. קטן מ-0.005.

4) א. לא נדחה H_0 . ב. 11.4 ג. 0.394
ד. 20.8%.

5) א. פלא א'- חוקר א', פלא ב'- חוקר ב'. ב. נקבל את H_0 .
ג. טעות מסוג שני- הכרענו שאין הבדל בין גברים לנשים למרות שיש במציאות הבדל.
ד. נקבל את H_1 .
ה. טעות מסוג ראשון- הכרענו שיש קשר בין מין להוצאה למרות שבמציאות אין קשר.
ו. כל חוקר פעל בשיטה סטטיסטית שונה ובמצב כזה יתכן מסקנות סותרות.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 15 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 114

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $\left(\mu_{1..},\mu_k\right)$ של k אוכלוסיות שונות.
לכן, בניתוח שונות, <u>השערות</u> המחקר הן:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ (התוחלות של כל האוכלוסיות שוות)

$H_1:$ אחרת (לפחות שתיים מהתוחלות שונות)

ההנחות הדרושות לביצוע התהליך:

1) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

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

3) המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor). משתנה זה הוא קטגוריאלי עם k רמות (levels).
כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה:
נסמן ב- n_i את גודל המדגם בקבוצה i.

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

$\overline{X}_1$ - ממוצע המדגם הראשון,....., $\overline{X}_k$ -ממוצע המדגם ה- k -י.

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

סכום ריבועים בין הקבוצות: $SS_B = \sum_{i=1}^{k} n_i \left[\overline{X}_i - \overline{X}\right]^2$

סכום ריבועים בתוך הקבוצות: $SS_W = \sum_{i=1}^{k} n_i \left[n_i - 1\right] \bullet \hat{S}_i^2$

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

$SST = SSB + SSW$

© כל הזכויות שמורות לברק קנדל

יש למלא את טבלת ניתוח השונות הבאה:

מקור השונות	סכום הריבועים SS	דרגות חופש df	ממוצע הריבועים MS	F
בין הקבוצות - B	SSB	$k-1$	$\frac{SSB}{k-1}$	$\frac{MSB}{MSW}$
בתוך הקבוצות - W	SSW	$n-k$	$\frac{SSW}{n-k}$	
סה"כ - T	SST	$n-1$		

$$F=\frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}}\sim F\left(k-1,n-k\right)$$

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

© כל הזכויות שמורות לברק קנדל

שאלות

1) מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש: קבוצה 1קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן".
כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר? מהו ה"גורם" וכמה רמות יש לו?

ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.

ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

2) בעיר מסוימת שלושה בתי ספר תיכון. ראש העיר התעניין לבדוק האם קיים הבדל בהצלחה של בתי הספר במקצוע מתמטיקה. לצורך כך הוא דגם מספר תלמידים שנבחנו במבחן הבגרות במתמטיקה ברמה של 3 יחידות בעירו ובדק עבור כל תלמיד מה ציון הבגרות שלו במתמטיקה.
להלן הציונים שהתקבלו:

"המתמיד"	"רבין"	"הס"
78	98	85
65	62	83
70	55	74
90	80	85
56		75

א. מהו המבחן הסטטיסטי המתאים? רשמו את ההשערות ואת ההנחות של המבחן.

ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (factor) כמה רמות יש לו?

ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.

ד. מלאו את טבלת ANOVA.

ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.

ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

3) מעוניינים לבדוק האם יש הבדל בהשפעה של שיטות טפול שונות על לחץ הדם הסיסטולי (SBP) באוכלוסייה של קשישים. נבדקו 4 שיטות שונות.
בטבלה המצורפת מרוכזים ממצאי המחקר.

השיטה	A	B	C	D
גודל המדגם	12	14	8	12
הממוצע	178	172	180	182
סטיית התקן	4	8	5	3

א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.

ב. מה מסקנת המחקר ברמת מובהקות של 5%?

ג. האם יש צורך לבצע השוואות מרובות?

© כל הזכויות שמורות לברק קנדל

4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.
האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות?
בדקו ברמת מובהקות של 5%.

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

5) להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות.
מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.317	3	3.439	1.361	.279
Within Groups	60.648	24	2.527		
Total	70.964	27			

6) להלן טבלת ANOVA בטבלה הושמטו חלקים.
השלימו את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	ב	ג	ה	.000
Within Groups	א	17	ד		
Total	522.950	19			

© כל הזכויות שמורות לברק קנדל

7) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים ל שלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה (dosage). המינונים שניתנו הם : 10 מ״ג, 20 מ״ג ו-30 מ״ג. לאחר שעה מזמן לקיחת התרופה ניבדק קצב פעימות הלב של כל אדם (pulse). הנתונים הוזנו לתוכנה סטטיסטית והתקבלו התוצאות הבאות:

ANOVA

pulse

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	414.400	2	207.200	19.733	.000
Within Groups	126.000	12	10.500		
Total	540.400	14			

pulse

Tukey HSD[a]

dosage	N	Subset for alpha = 0.05	
		1	2
30.00	5	71.0000	
20.00	5		80.2000
10.00	5		83.4000
Sig.		1.000	.299

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

Post Hoc Tests

Multiple Comparisons

pulse
Tukey HSD

(I) dosage	(J) dosage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
10.00	20.00	3.20000	2.04939	.299	-2.2675	8.6675
	30.00	12.40000*	2.04939	.000	6.9325	17.8675
20.00	10.00	-3.20000	2.04939	.299	-8.6675	2.2675
	30.00	9.20000*	2.04939	.002	3.7325	14.6675
30.00	10.00	-12.40000*	2.04939	.000	-17.8675	-6.9325
	20.00	-9.20000*	2.04939	.002	-14.6675	-3.7325

*. The mean difference is significant at the 0.05 level.

א. בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.

ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

ג. האם יש צורך במחקר בהשוואת מרובות. נמקו!

ד. לטבלת הANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת ״טוקי״. ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

© כל הזכויות שמורות לברק קנדל

8) בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו. ענו על הסעיפים הבאים:

א. כמה בתי ספר יש בעיר?

ב. כמה תלמידים השתתפו בסך הכול במחקר?

ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%

ד. בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

Oneway

ANOVA

grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7799.600	4	1949.900	13.586	.000
Within Groups	2870.400	20	143.520		
Total	10670.000	24			

© כל הזכויות שמורות לברק קנדל

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	5.40000	7.57681	.971	-20.2543	31.0543
	3.00	36.80000*	7.57681	.003	11.1457	62.4543
	4.00	36.40000*	7.57681	.003	10.7457	62.0543
	5.00	-2.60000	7.57681	.998	-28.2543	23.0543
2.00	1.00	-5.40000	7.57681	.971	-31.0543	20.2543
	3.00	31.40000*	7.57681	.011	5.7457	57.0543
	4.00	31.00000*	7.57681	.013	5.3457	56.6543
	5.00	-8.00000	7.57681	.888	-33.6543	17.6543
3.00	1.00	-36.80000*	7.57681	.003	-62.4543	-11.1457
	2.00	-31.40000*	7.57681	.011	-57.0543	-5.7457
	4.00	-.40000	7.57681	1.000	-26.0543	25.2543
	5.00	-39.40000*	7.57681	.001	-65.0543	-13.7457
4.00	1.00	-36.40000*	7.57681	.003	-62.0543	-10.7457
	2.00	-31.00000*	7.57681	.013	-56.6543	-5.3457
	3.00	.40000	7.57681	1.000	-25.2543	26.0543
	5.00	-39.00000*	7.57681	.001	-64.6543	-13.3457
5.00	1.00	2.60000	7.57681	.998	-23.0543	28.2543
	2.00	8.00000	7.57681	.888	-17.6543	33.6543
	3.00	39.40000*	7.57681	.001	13.7457	65.0543
	4.00	39.00000*	7.57681	.001	13.3457	64.6543

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

grade

Scheffe[a]

school	N	Subset for alpha = 0.05	
		1	2
3.00	5	45.0000	
4.00	5	45.4000	
2.00	5		76.4000
1.00	5		81.8000
5.00	5		84.4000
Sig.		1.000	.888

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 5.000.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. משתנה בלתי תלוי: סוג התרופה.
משתנה תלוי: הזמן עד להשפעת התרופה בדקות.

ב. ניתוח שונות חד כיווני

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_1: otherwise$$

ג. 1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

2) א. המבחן לניתוח שונות חד כיוונית.

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_1: otherwise$$

הנחות:
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונות.

ב. גודל המדגם: 14. משתנה ב"ת: בית הספר, בעל 3 רמות.

ג. $\hat{X} = 12.93, \hat{S} = 71.8$, $\hat{X} = 19.29, \hat{S} = 73.75$, $\hat{X} = 5.46, \hat{S} = 80.4$.

ד. להלן טבלה:

מקור השונות	SS	df	MS	F
B	200.6	2	100.3	
W	1904.75	11	173.2	
סה"כ	2105.35	13		0.58

ה. $F > 3.98$.

ו. נקבל את H_0.

3) א.
$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_1: otherwise$$

ב. נדחה את H_0.

ג. כן.

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

© כל הזכויות שמורות לברק קנדל

4) נקבל את H_0 : נכריע שאין הבדל מובהק בין האופים מבחינת תוחלת זמן הכנה.

5) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1 : otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

נקבל את H_0 : לסוג סוללה אין השפעה של תוחלת החיים ברמת ביטחון של 10%.

6) להלן טבלה:

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450 ב	2	178.725 ג	18.36 ה	.000
Within Groups	165.5 א	17	9.735 ד		
Total	522.950	19			

7) א. $H_0 : \mu_1 = \mu_2 = \mu_3$
$H_1 : otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונות.

נדחה את H_0 : ברמת ביטחון של 5% קיים הבדל במינונים השונים מבחינת תוחלת הדופק.

ב. ראה וידאו. ג. כן. ד. $\mu_{20} = \mu_{10} > \mu_{30}$.

8) א. 5 ב. 25

ג. נדחה את H_0 : יש לפחות שני בתי ספר בעיר עם תוחלת רמת ציונים שונה.

ד. $(\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3)$.

© כל הזכויות שמורות לברק קנדל

הסקה וניתוח נתונים

פרק 16 - ניתוח שונות דו כיווני

תוכן העניינים

1. הקדמה..... 123
2. אפקטים פשוטים , עיקריים ואינטראקציה..... 133
3. תהליך ניתוח שונות דו כיווני..... 145

ניתוח שונות דו-כיווני - הקדמה

רקע

ראשית, נחזור על עיקרי ניתוח השונות החד-כיווני (חד-גורמי).

בניתוח שונות חד-כיווני יש משתנה תלוי יחיד, שהוא כמותי, ומשתנה בלתי תלוי יחיד, שהוא משתנה קטגוריאלי (משתנה שהערכים שלו שייכים למספר סופי של קטגוריות). המשתנה הקטגוריאלי נקרא לעתים גם גורם (פקטור), והקטגוריות שלו נקראות רמות. המטרה בניתוח שונות חד-כיווני היא לבדוק האם לגורם יש השפעה מובהקת על המשתנה התלוי. השערת האפס של המחקר בניתוח שונות חד-כיווני היא שבכל הקטגוריות יש אותה התוחלת, והשערת המחקר טוענת שיש לפחות שתי קטגוריות שבהן התוחלות שונות.

<u>דוגמה: (פתרון בהקלטה)</u>

נבדקו שלושה סוגי דיאטות על אנשים בעלי משקל עודף. נבחרו 30 מטופלים בעלי משקל עודף, והם חולקו באקראי לשלוש קבוצות שוות בגודלן, כך שכל קבוצה קיבלה דיאטה נחקרת אחרת. כעבור שלושה חודשים בדקו את מספר הקילוגרמים שהפחית כל מטופל ממשקלו בתקופה זו. מטרת המחקר הייתה לבדוק האם קיים הבדל בין הדיאטות מבחינת ההפחתה במשקל.

- מהו המשתנה התלוי במחקר?
- מהו המשתנה הבלתי תלוי במחקר? כמה רמות יש לו?
- מה הן השערות המחקר?
- מהו המבחן הסטטיסטי המתאים?

בניתוח שונות דו-כיווני אנו מוסיפים עוד משתנה בלתי תלוי למחקר, כלומר עוד גורם שאנו רוצים לבדוק איך הוא משפיע על המשתנה התלוי.
לכן בניתוח שונות דו-כיווני יש משתנה תלוי כמותי יחיד ושני משתנים בלתי תלויים שכל אחר מהם קטגוריאלי. כזכור, למשתנים הבלתי תלויים אנו קוראים גם גורמים (פקטורים), ומספר הקטגוריות של כל גורם נקרא גם מספר הרמות שלו.
ניתוח שונות רב-כיווני או רב-גורמי הוא ניתוח שונות שבו יש יותר מגורם אחד, כלומר יותר ממשתנה בלתי תלוי קטגוריאלי אחד. בניתוח שונות דו-כיווני יש שני גורמים, בניתוח שונות תלת-גורמי יש שלושה גורמים וכו'. ככל שנוסיף גורמים, הניתוח הסטטיסטי יהיה מורכב יותר וייידרשו יותר תצפיות למחקר אבל כיוון שהוא יקטין את שונות הטעויות (שונות מקרית) וייתן יותר הסבר לשונות הכללית, כך שהמבחן יהיה עוצמתי יותר.

© כל הזכויות שמורות לברק קנדל

המשך הדוגמה:

מבין 30 המטופלים שבמחקר 15 היו גברים ו-15 היו נשים. המטופלים חולקו כך שבכל דיאטה השתתפו 5 גברים ו-5 נשים.

מה הם המשתנים הבלתי תלויים? כמה רמות יש לכל משתנה?

בניתוח שונות דו-כיווני אנו בעצם רוצים לבדוק סימולטנית שלוש שאלות מחקר על אוכלוסיית כבדי המשקל:

- האם יש הבדלים משמעותיים בין שיעורי הפחתת המשקל של מטופלים כבדי משקל כתוצאה משימוש בדיאטות שונות?
- האם יש הבדלים משמעותיים בין שיעורי הפחתת המשקל של מטופלים כבדי משקל כתוצאה ממגדר שונה?
- האם יש השפעה משולבת (אינטראקציה) של שני הגורמים הנבדקים על הפחתת המשקל של מטופלים כבדי משקל, כלומר האם צירוף של דיאטה מסוימת ומגדר מסוים מביא להפחתת משקל גדולה יותר או קטנה יותר מצירופים אחרים?

נסמן גורם אחד ב-a ואת מספר הרמות שלו ב-A. באותו האופן הגורם האחר יסומן ב-b, ואת מספר הרמות שלו נסמן ב-B. מספר הקבוצות הכולל שאנו יוצרים הוא $A\cdot B$.

המשך הדוגמה:

- בחרו גורם אחד להיות a וגורם אחר להיות b. מהו A ומהו B?
- כמה קבוצות שונות נוצרו במחקר?

נסמן ב- m את מספר התצפיות בכל תא (בהנחה שהוא יהיה מספר קבוע). תא הוא שילוב של רמה מסוימת של גורם a עם רמה מסוימת של גורם b.

המשך הדוגמה:

- כמה תאים (קבוצות) יש במחקר?
- מה ערכו של m?
- מהו הקשר המתמטי בין m לבין n, גודל המדגם?

© כל הזכויות שמורות לברק קנדל

נסמן ב- a_1 את הרמה הראשונה של a, ב- a_2 את הרמה השנייה שלו וכך הלאה.

נסמן ב- b_1 את הרמה הראשונה של b, ב- b_2 את הרמה השנייה שלו וכך הלאה.

נסמן ב- $\mu_{i.}$ את התוחלת ברמה a_i. נסמן ב- $\mu_{.j}$ את התוחלת ברמה b_j. נסמן ב- μ_{ij} את התוחלת של תא ij.

<u>המשך הדוגמה:</u>

- מה המשמעות של $\mu_{1.}$ ושל $\mu_{.2}$?
- מה המשמעות של μ_{12} ושל μ_{21}?

<u>השערות המחקר בניתוח שונות דו-כיווני</u>

את השערות המחקר בניתוח שונות דו-כיווני אפשר לרשום בצורות רבות:

לגורם a אין השפעה על המשתנה התלוי: H_0

אחרת: H_1

לגורם b אין השפעה על המשתנה התלוי: H_0

אחרת: H_1

אין אינטראקציה בין שני הגורמים: H_0

אחרת: H_1

© כל הזכויות שמורות לברק קנדל

דרך אחרת היא שימוש בתוחלות:

$H_0: \mu_{1.} = \mu_{2.} = \cdots = \mu_{A.}$

$H_1:$ אחרת

$H_0: \mu_{.1} = \mu_{.2} = \cdots = \mu_{.B}$

$H_1:$ אחרת

$H_0:$ אין אינטראקציה בין שני הגורמים

$H_1:$ אחרת

<u>המשך הדוגמה:</u>

אם אנחנו מעוניינים לבצע ניתוח שונות דו-כיווני, מה הן ההשערות הנחקרות?

© כל הזכויות שמורות לברק קנדל

שאלות

1) בחברת טקסטיל בחנו 4 סוגי בדים שונים מבחינת חוזקם.
דגמו 5 חתיכות בד מכל סוג ובדקו את חוזק הקריעה של כל סוג בד.

א. מהו המשתנה התלוי במחקר?

ב. כמה משתנים בלתי תלויים יש במחקר? מה הם?

ג. מהו המבחן הסטטיסטי המתאים במקרה זה?

2) במחקר בתחום הפסיכולוגיה נדגמו אנשים הסובלים מחרדות מסוגים שונים. כל מטופל סווג כסובל מאחד מסוגי החרדות הבאים: חרדה חברתית, חרדה כללית או אגורפוביה. במחקר השתתפו 6 מטופלים מכל סוג חרדה שצוין. המטופלים במחקר חולקו כך שכל מטופל היה צריך לעבור במשך שנה את אחד מהטיפולים הבאים: טיפול קוגניטיבי התנהגותי (CBT), טיפול קבוצתי או טיפול דיאלקטי התנהגותי (DBT). בכל סוג טיפול השתתפו 2 מטופלים מכל סוג חרדה. בסוף השנה נבדקו כל המטופלים וקיבלו ציון כמותי על השיפור במצבם הנפשי (משתנה כמותי). מטרת המחקר הייתה לבדוק האם סוג החרדה, סוג הטיפול והשילוב ביניהם משפיעים על המצב הנפשי של המטופלים.

א. מהו גודל המדגם?

ב. מהו המשתנה התלוי במחקר הזה ומה הם המשתנים הבלתי תלויים?

ג. כמה קטגוריות יש לכל משתנה בלתי תלוי?

ד. כמה קבוצות שונות יש במערך המחקרי?

ה. מהו המבחן הסטטיסטי המתאים במערך מחקרי זה?

© כל הזכויות שמורות לברק קנדל

3) מחקר שיווקי בדק את השפעת גובה המדף בסופרמרקט והשפעת החומר שממנו עשוי הבקבוק (זכוכית או פלסטיק) על היקף המכירות של משקאות קלים. נבדקו שני סופרמרקטים. בכל סופרמרקט נבחן כל צירוף אפשרי של גובה המדף וחומר הבקבוק, ועבור כל צירוף כזה נבדק מספר בקבוקי המשקה הקל שנמכרו באותו סופרמרקט ביום מסוים. הנה התוצאות שהתקבלו:

סוג בקבוק / גובה המדף	זכוכית	פלסטיק
נמוך	23 32	59 63
בינוני	47 55	88 90
גבוה	40 48	51 56

א. מהו המבחן הסטטיסטי המתאים? נמקו.

ב. מהו מספר הרמות של כל גורם מחקרי?

ג. מה יהיו השערות המחקר אם יתבצע ניתוח שונות דו-כיווני?

ד. מהו ערכו של m ומהו ערכו של n?

© כל הזכויות שמורות לברק קנדל

4) יצרן של נוזל כביסה מעוניין לבחון שני נוזלי ניקוי מבחינת יעילותם בהסרת כתמים בשלוש רמות טמפרטורה. בכל אחד משש הצירופים של סוג נוזל וטמפרטורה נבחנה יכולת הסרת הכתמים מבדים דומים, וניתן ציון בין 0 ל-15 (הציון הטוב ביותר).

מספר סידורי	סוג הנוזל	טמפרטורה במעלות צלזיוס	ציון הסרת כתמים
1	C	30	4
2	C	30	5
3	C	30	4
4	C	30	6
5	C	40	6
6	C	40	6
7	C	40	7
8	C	40	6
9	C	60	9
10	C	60	8
11	C	60	7
12	C	60	10
13	w	30	9
14	w	30	9
15	w	30	9
16	w	30	10
17	w	40	12
18	w	40	13
19	w	40	11
20	w	40	11
21	w	60	14
22	w	60	14
23	w	60	15
24	w	60	13

© כל הזכויות שמורות לברק קנדל

א. כמה משתנים יש במחקר?
ב. לגבי כל משתנה קבעו האם הוא משתנה תלוי או בלתי תלוי.
ג. כמה רמות יש לכל גורם?
ד. אם נבצע ניתוח שונות דו-כיווני, מה יהיו השערות המחקר?
ה. רכזו את נתוני המחקר בטבלה שבה בשורות גורם אחד, בעמודות גורם שני ובתאים התוצאות שהתקבלו למשתנה התלוי.

5) קבעו לגבי כל אחד מהבאים האם הוא משתנה קטגוריאלי:
א. מספר הניתוחים שעבר אדם בחייו.
ב. אחוז האבטלה בישראל בחודש זה.
ג. סוג הדם של חולה.
ד. שונות הציונים בבחינת הבגרות באנגלית במועד האחרון.
ה. משקל חבילה בדואר בגרמים.
ו. היבשת שאירחה את משחקי המונדיאל.

בשאלות הבאות יש לבחור את התשובה הנכונה ביותר:

6) משרד החינוך רוצה לבדוק עד כמה שיטת הוראה (יש 3 שיטות הוראה מקובלות) ומגדר משפיעים על ציוני הבגרות בהיסטוריה. מהו המבחן הסטטיסטי המתאים למחקר זה?
א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. מבחן T לתוחלת אחת.

7) מחלקת שירות הלקוחות של חברת החשמל דגמה עובדים כדי לבחון האם ככל שמספר שנות הוותק של נותן השירות גדול יותר גדל גם מספר הלקוחות שבו הוא מטפל במהלך משמרת. מהו המבחן הסטטיסטי שיכול לבדוק זאת?
א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. אף אחת מהאפשרויות שלעיל.

© כל הזכויות שמורות לברק קנדל

8) האיחוד האירופי המשותף דגם 10 עובדים מתחום ההוראה בכל אחת מהמדינות הבאות: הולנד, צרפת, בלגיה, גרמניה ואוסטריה. לכל עובד בדקו את גובה המשכורת החודשית שלו ביורו. אם נרצה להשוות בין המדינות הללו מבחינת תוחלת השכר של עובדי ההוראה במדינה, מה יהיה המבחן הסטטיסטי המתאים?

א. מבחן T להשוואת תוחלות
ב. ניתוח שונות חד-כיווני
ג. ניתוח שונות דו-כיווני
ד. אף אחת מהאפשרויות שלעיל

9) בקו ייצור 2 סוגים של מכונות ו-3 רמות ותק של מפעיל המכונה (עד שנתיים במפעל, בין שנתיים ל- חמש שנים במפעל, יותר מחמש שנים במפעל). מנהל הייצור רוצה לבדוק אם קיימת השפעה של סוג המכונה והוותק של המפעיל על מספר המוצרים הפגומים שיוצאים מהמכונה. מה יהיה המבחן הסטטיסטי המתאים במקרה זה?

א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. ניתוח שונות תלת-כיווני.

10) במחקר נאספו הנתונים הבאים על קבוצת נחקרים:
1. כמה כוסות קפה הנחקר שותה ביום: לא שותה / 1-2 כוסות/ יותר מ-2 כוסות.
2. מין הנחקר: גבר/אישה.
3. דופק (מספר פעימות בדקה) שעתיים אחרי הקימה.

מטרת המחקר הייתה לבדוק האם מספר כוסות הקפה שאדם שותה ביום משפיע על הדופק אצל גברים אחרת מאשר אצל נשים.
מה יהיה המבחן הסטטיסטי המתאים במקרה זה?

א. מבחן T להשוואת תוחלות.
ב. ניתוח שונות חד-כיווני.
ג. ניתוח שונות דו-כיווני.
ד. ניתוח שונות תלת-כיווני.

© כל הזכויות שמורות לברק קנדל

11) במחקר יש משתנה כמותי אחד ושני גורמים שלכל אחד מהם שתי רמות. אילו מהמשפטים הבאים אינו נכון?

א. אפשר מבחינה טכנית לבדוק כיצד כל גורם בנפרד משפיע על המשתנה התלוי באמצעות ניתוח שונות חד-כיווני שייערך לכל גורם בנפרד.

ב. אפשר מבחינה טכנית להשוות בין התוחלות של כל רמה בגורם הראשון על ידי מבחן T להשוואת תוחלות.

ג. אפשר מבחינה טכנית לבצע ניתוח שונות דו-כיווני במערך מחקרי זה.

ד. כיוון שבמחקר יש בסך הכול שלושה משתנים, אפשר מבחינה טכנית לבצע ניתוח שונות תלת-כיווני.

תשובות סופיות

1) א. חוזק הקריעה. ב. אחד. סוג הבד.
ג. ניתוח שונות חד גורמי.

2) א. 18
ב. המשתנה התלוי: ציון במצב הנפש. המשתנם הב"ת: סוג חרדה, סוג הטיפול.
ג. 3,3 ד. 9
ה. ניתוח שונות דו גורמי.

3) א. ניתוח שונות דו גורמי. ב. 3,2
ג. H_0: אין אינטראקציה, H_1: יש אינטראקציה. ד. $m=2,\ n=12$

4) א. 3
ב. משתנים ב"ת: סוג הנוזל, טמפרטורה. משתנה תלוי: ציון הסרת כתמים.
ג. 3,2
ד. H_0: אין אינטראקציה בין הגורמים, H_1: אחרת.
ה. עיין בסרטון הוידאו.

5) א. כן. ב. לא.
ג. כן. ד. לא.
ה. תלוי. ו. כן.

6) ג.
7) ד
8) ב
9) ג
10) ג
11) ד

© כל הזכויות שמורות לברק קנדל

אפקטים פשוטים, עיקריים ואינטראקציה

רקע

בניתוח שונות דו-כיווני אנו דנים במשתנה כמותי תלוי יחיד ובשני משתנים בלתי תלויים (גורמים) המחולקים כל אחד למספר רמות.
מטרת המחקר היא לבדוק שלוש השערות שונות:

H_0: לגורם a אין השפעה על המשתנה התלוי

H_1: אחרת

H_0: לגורם b אין השפעה על המשתנה התלוי

H_1: אחרת

H_0: אין אינטראקציה בין שני הגורמים

H_1: אחרת

נרצה להבין מה בדיוק כל השערה בודקת לגבי האוכלוסייה הנחקרת.

<u>אפקט עיקרי</u>: אם יש שתי קטגוריות (רמות) לפחות של גורם מסוים שהתוחלות שלהן שונות, נאמר שלגורם זה יש השפעה על המשתנה התלוי. השפעה זאת נקראת "אפקט עיקרי". למשל, אם יימצאו לפחות שתי תרופות נוגדות דיכאון שונות שמביאות לתוחלות שונות במצב הנפשי, נגיד שלסוג התרופה יש השפעה על המצב הנפשי, כלומר יש אפקט עיקרי. כמות האפקטים העיקריים שאפשר למצוא היא ככמות הגורמים במחקר.

אפקט אינטראקציה: מצב שבו גורם אחד משפיע על המשתנה התלוי באופן שונה בקטגוריות שונות של הגורם השני. למשל, תרופה נוגדת דיכאון אחת מביאה את הגברים למצב רוח טוב יותר מאשר את הנשים לעומת תרופה אחרת שמביאה דווקא את הנשים למצב רוח טוב יותר מאשר את הגברים. אפקט האינטראקציה הוא יחיד, כלומר נאמר אם יש או אין אינטראקציה. כמו כן הוא אפקט סימטרי: אם קיימת אינטראקציה בין מגדר לסוג התרופה, יש גם אינטראקציה בין סוג התרופה למגדר.

אפקט פשוט: אפקט פשוט מתייחס להשפעת גורם אחד על המשתנה התלוי בתוך קטגוריה מסוימת של הגורם השני. למשל, נרצה לבדוק רק בקטגוריה של הגברים האם קיים הבדל בין התרופות נוגדות הדיכאון. אם נמצא הבדלים כאלה נאמר שיש

© כל הזכויות שמורות לברק קנדל

אפקט פשוט של סוג התרופה בקרב אוכלוסיית הגברים. כמות האפקטים הפשוטים שאפשר למצוא היא סכום מספר הקטגוריות (רמות) של כל גורם. למשל, אם יש שלושה סוגי תרופות ושתי אפשרויות למגדר, בסך הכול נוכל לבדוק 5 אפקטים פשוטים.

<u>דוגמה</u>

נבדקו שלושה סוגי דיאטות על אנשים בעלי משקל עודף. כעבור שלושה חודשים בדקו כמה קילוגרמים הפחית כל מטופל ממשקלו באותה התקופה. נניח שאנו יודעים את תוחלת הפחתת המשקל של כל דיאטה בחלוקה למגדרים.

נתאר כמה מצבים אפשריים לגבי האוכלוסייה הנחקרת וננתח כל מצב מבחינת ההשפעה של כל גורם על תוחלת המשתנה התלוי ומבחינת אפקט האינטראקציה.

שימו לב שהמצבים שנתאר להלן מתייחסים לתוחלות האמיתיות. בניתוח שונות אין לנו נתוני אמת, אלא רק נתוני מדגם, ונרצה לבדוק האם האפקטים שהתקבלו במדגם הם מובהקים, כנדרש בכל תהליך של הסקה סטטיסטית.

אם התוצאות שלנו יהיו ממוצעי מדגם ולא תוחלות, נוכל לבדוק אם קיימים אפקטים במדגם, אך אין זה אומר שקיימים אפקטים באוכלוסייה, כלומר לא נוכל לדעת אם האפקטים במדגם הם מובהקים. כדי לבדוק אם האפקטים הם מובהקים נצטרך לעשות את מבחן ניתוח השונות.

<u>מצב א:</u>

הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק״ג לכל קבוצה:

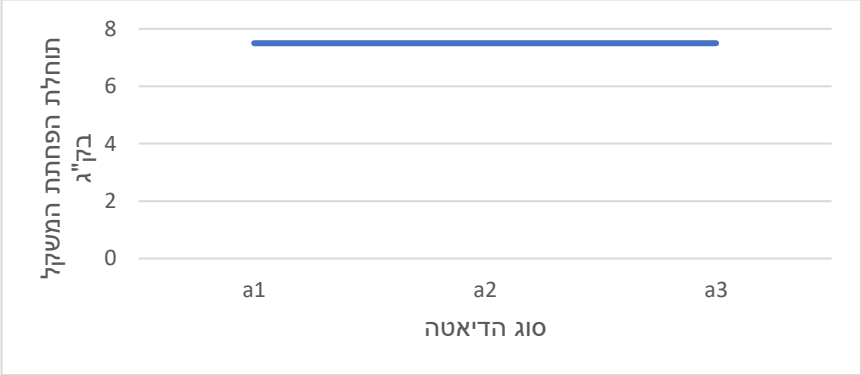
	גברים	נשים
a_1	10	5
a_2	10	5
a_3	10	5

<u>תיאור גרפי לבדיקת אפקט למגדר</u>

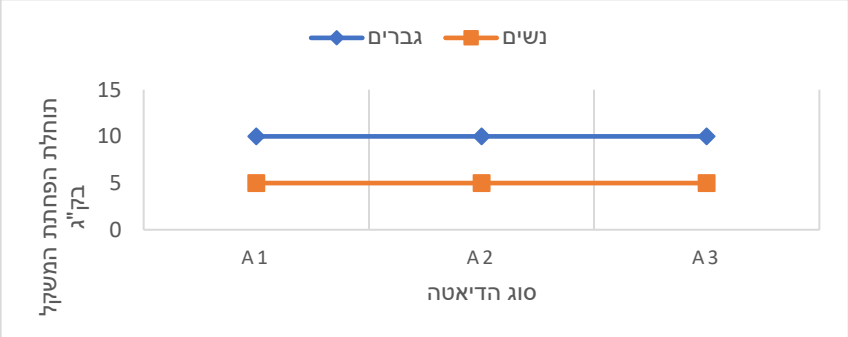


© כל הזכויות שמורות לברק קנדל

תיאור גרפי לבדיקת אפקט לסוג הדיאטה



גרף אפקטים פשוטים



ניתוח המצב: למגדר יש אפקט, לסוג הדיאטה אין אפקט, אין אפקט אינטראקציה.

הערה: אם הקווים הנוצרים בגרף האפקטים הפשוטים מקבילים או מתלכדים, אנו אומרים שאין אפקט אינטראקציה.

© כל הזכויות שמורות לברק קנדל

<u>מצב ב</u>

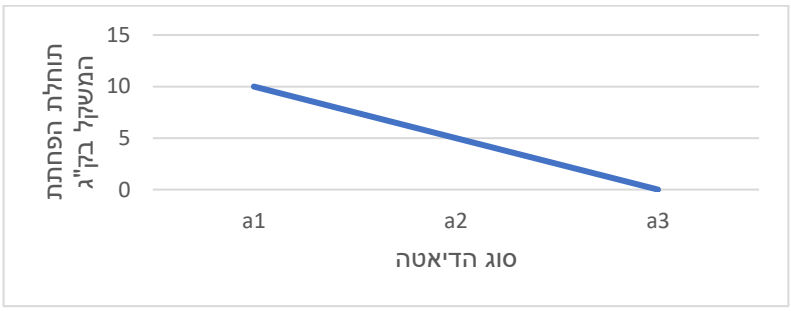
הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק"ג לכל קבוצה:

	גברים	נשים
a_1	10	10
a_2	5	5
a_3	0	0

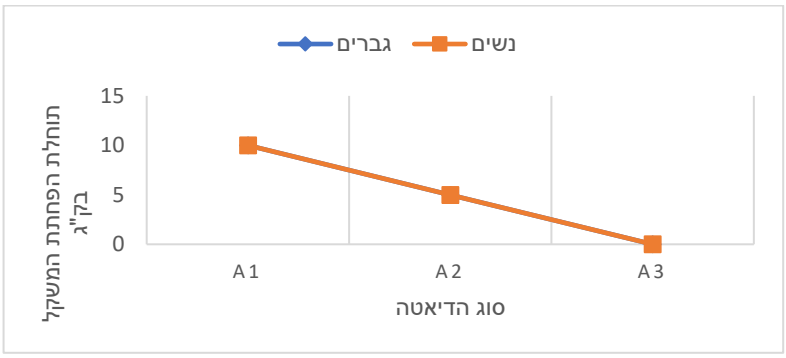
<u>תיאור גרפי לבדיקת אפקט למגדר</u>



<u>תיאור גרפי לבדיקת אפקט לסוג הדיאטה</u>



<u>גרף אפקטים פשוטים</u>



<u>ניתוח המצב</u>: למגדר אין אפקט, לסוג הדיאטה יש אפקט, אין אפקט אינטראקציה.

© כל הזכויות שמורות לברק קנדל

מצב ג

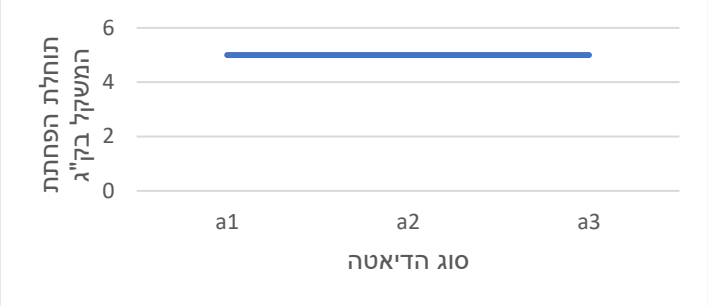
הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק"ג לכל קבוצה:

	גברים	נשים
a_1	10	0
a_2	5	5
a_3	0	10

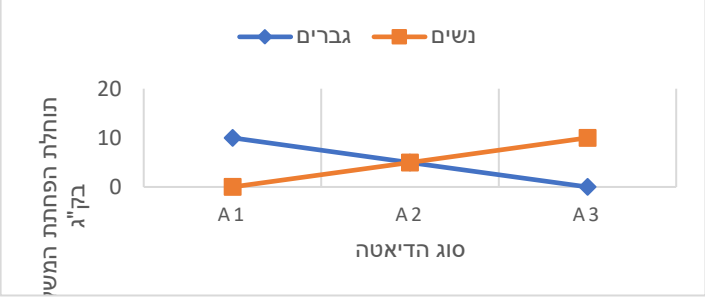
תיאור גרפי לבדיקת אפקט למגדר



תיאור גרפי לבדיקת אפקט לסוג הדיאטה



גרף אפקטים פשוטים



ניתוח המצב: למגדר אין אפקט, לסוג הדיאטה אין אפקט, יש אפקט אינטראקציה.

© כל הזכויות שמורות לברק קנדל

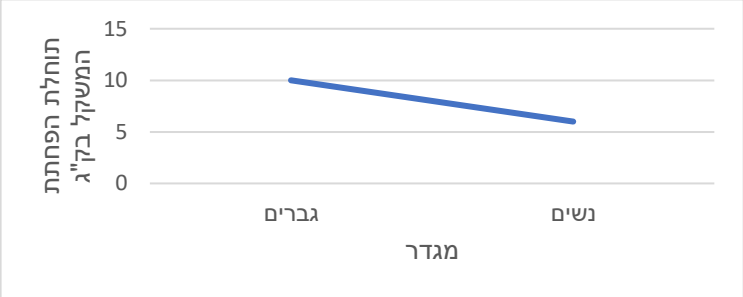
אינטראקציה דיסאורדינלית (נקראת גם "אינטראקציה מהותית"):
אפשר לזהות מצב של אינטראקציה כזו באמצעות גרף של אפקטים פשוטים, כאשר נוצרים קווים נחתכים שאחד מהם עולה והאחר יורד. המשמעות היא שגורם אחד משפיע על המשתנה התלוי ברמה מסוימת של הגורם השני באופן הפוך משהוא משפיע על המשתנה התלוי ברמה אחרת של הגורם השני.
במצב זה אין להתייחס לאפקטים עיקריים. יש להתייחס רק לאפקטים הפשוטים.

<u>מצב ה</u>

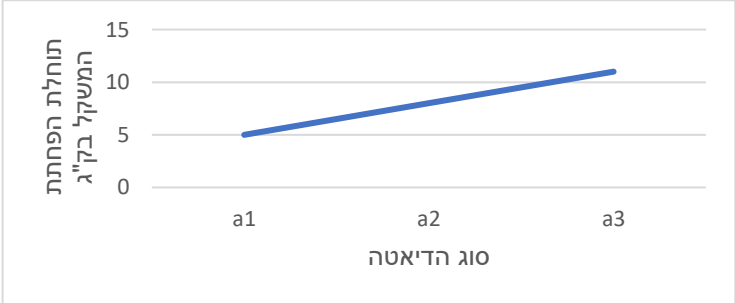
הטבלה הבאה מתארת את תוחלת הפחתת המשקל בק"ג לכל קבוצה:

	גברים	נשים
a_1	5	5
a_2	10	6
a_3	15	7

<u>תיאור גרפי לבדיקת אפקט למגדר</u>

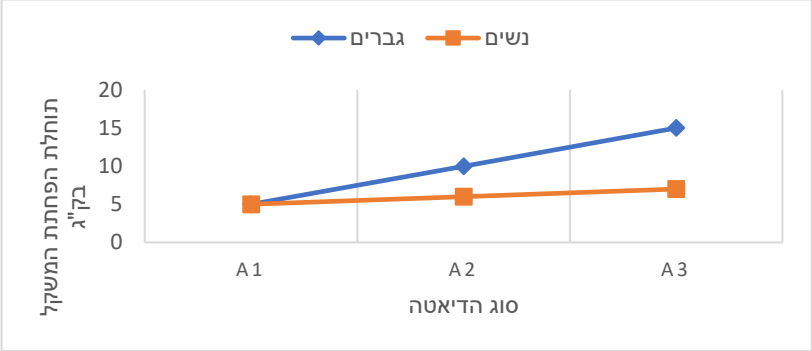


<u>תיאור גרפי לבדיקת אפקט לסוג הדיאטה</u>



© כל הזכויות שמורות לברק קנדל

גרף אפקטים פשוטים



ניתוח המצב: למגדר יש אפקט, לסוג הדיאטה יש אפקט, יש אפקט אינטראקציה.

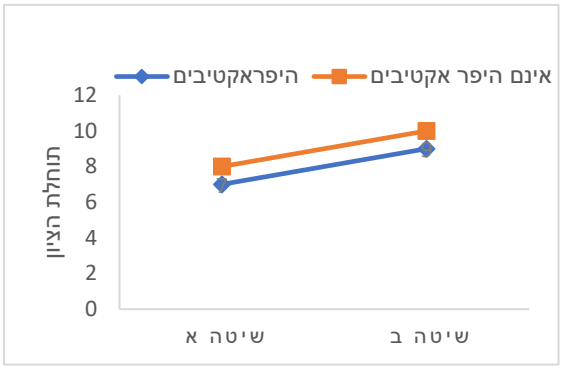
אינטראקציה אורדינלית (נקראת גם "אינטראקציה לא מהותית"): אפשר לזהות מצב של אינטראקציה כזו כאשר בגרף האפקטים הפשוטים נוצרים קווים נחתכים עם אותו הכיוון (כולם עולים או כולם יורדים אבל לא באותו השיפוע). המשמעות היא שבמעבר של גורם אחד מרמה אחת לרמה אחרת שלו הוא משפיע על המשתנה התלוי באותו אופן בכל רמה של המשתנה האחר אבל עם גודל אפקט שונה.

© כל הזכויות שמורות לברק קנדל

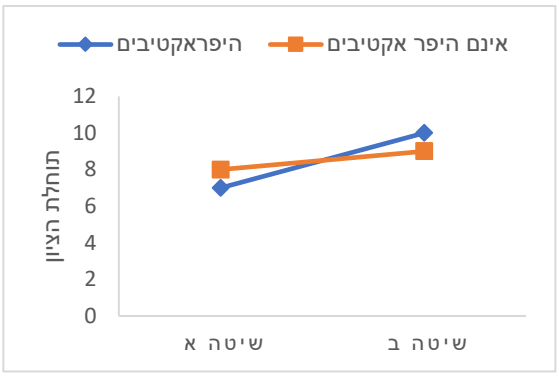
שאלות

1) בגני החובה יש שתי שיטות הוראה. שיטות אלו נוסו על ילדים היפראקטיביים וילדים שאינם היפראקטיביים. בתרשימים הבאים מיוצגים גרפים שמתארים את תוחלת הציון במבחן אוצר המילים שניתן לילדים בסוף השנה.
בכל אחד מהמקרים יש לקבוע האם קיימת אינטראקציה בין שני הגורמים. אם קיימת אינטראקציה, יש לקבוע האם היא אינטראקציה אורדינלית או דיסאורדינלית.

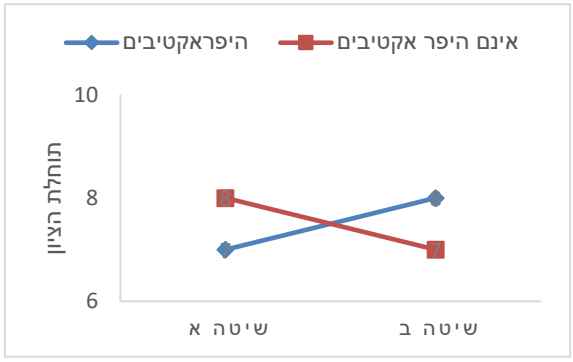
א.



ב.

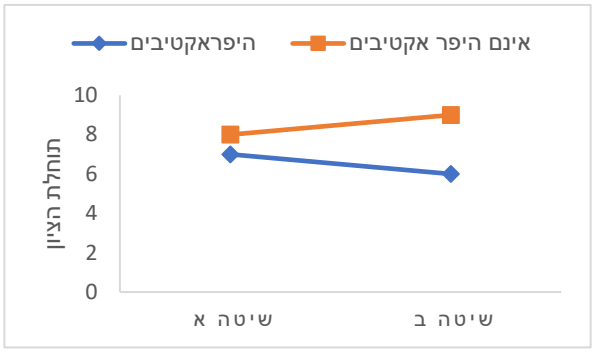


ג.

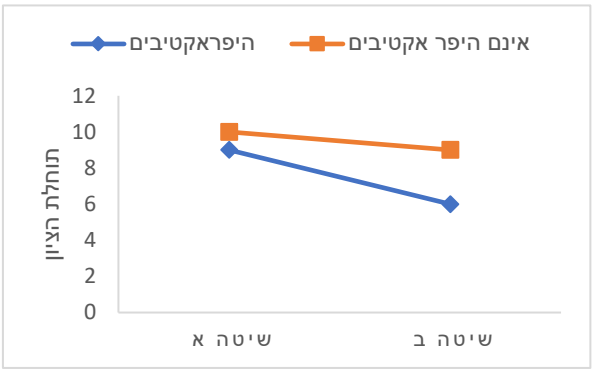


© כל הזכויות שמורות לברק קנדל

ד.



ה.



2) משרד האוצר פרסם נתונים על המחיר הממוצע של דירות גן ודירות גג של 4 חדרים ב-3 ערים בארץ. מחיר הדירות נמדד במיליוני שקלים.
להלן התוצאות שהתקבלו:

	דירות גן	דירות גג
הרצליה	4	3
אשדוד	2	1
חולון	3	2

א. מהו המשתנה התלוי ומה הם המשתנים הבלתי תלויים?

ב. האם קיים אפקט לעיר? היעזרו בגרף מתאים.

ג. האם קיים אפקט לסוג הדירה? היעזרו בגרף מתאים.

ד. האם קיימת אינטראקציה בין הגורמים? אם כן, מהו סוג האינטראקציה? היעזרו בגרף מתאים.

ה. האם יש אפקט פשוט לעיר עבור דירות גן?

ו. האם יש אפקט פשוט לעיר עבור דירות גג?

ז. האם יש אפקט פשוט לסוג הדירה בהרצליה?

ח. האם יש אפקט פשוט לסוג הדירה באשדוד?

ט. האם יש אפקט פשוט לסוג הדירה בחולון?

© כל הזכויות שמורות לברק קנדל

3) משרד החינוך פרסם נתונים על תוחלת הציונים בבחינת הבגרות באנגלית לפי עיר וסוג בית הספר (עיוני או מקצועי). להלן התוצאות שהתקבלו:

	עיוני	מקצועי
רעננה	85	70
תל אביב	75	75
פתח תקווה	70	85

א. תארו את הנתונים באמצעות גרף אפקטים פשוטים.

ב. האם קיימת אינטראקציה בין הגורמים? אם כן, מה סוג האינטראקציה?

ג. באילו ערים קיים אפקט פשוט לסוג בית הספר?

4) משרד התחבורה פרסם נתונים על תוחלת מספר עבירות התנועה לבעלי רישיון נהיגה לפי עיר ולפי מגדר. להלן התוצאות שהתקבלו:

	גבר	אישה
חיפה	2	1
אשקלון	2	1
רמת גן	2	1

א. האם קיים אפקט עיקרי לעיר?

ב. האם קיים אפקט עיקרי למגדר?

ג. האם יש אפקט פשוט לעיר אצל הגברים?

ד. האם קיימת אינטראקציה בין הגורמים? אם כן, מהו סוג האינטראקציה?

5) המשרד לאיכות הסביבה פרסם נתונים על תוחלת רמת זיהום האוויר בערים שונות בארץ בחורף ובקיץ. להלן התוצאות שהתקבלו:

	קיץ	חורף
חיפה	20	20
ירושלים	10	10
באר שבע	15	15

א. האם קיים אפקט עיקרי לעיר?

ב. האם קיים אפקט עיקרי לעונה?

ג. האם קיימת עיר שבה יש אפקט פשוט לעונה?

ד. האם קיימת אינטראקציה בין הגורמים? אם כן, מה סוג האינטראקציה?

© כל הזכויות שמורות לברק קנדל

6) המשרד לאיכות הסביבה פרסם נתונים על תוחלת רמת זיהום האוויר בערים שונות בארץ בחורף ובקיץ. להלן התוצאות שהתקבלו:

	קיץ	חורף
רמת גן	10	10
גבעתיים	10	10
בת ים	10	10

האם קיים אפקט עיקרי לגורם כלשהו? האם קיימת אינטראקציה?

בשאלות הבאות יש לבחור את התשובה הנכונה ביותר:

7) במחקר נדגמו 5 אנשים מכל אחת מ-4 הקבוצות הבאות: 1. מתעמלים באופן קבוע ושומרים על תזונה בריאה; 2. מתעמלים באופן קבוע ולא שומרים על תזונה בריאה; 3. לא מתעמלים באופן קבוע ושומרים על תזונה בריאה; 4. לא מתעמלים באופן קבוע ולא שומרים על תזונה בריאה. להלן טבלה המסכמת את ממוצע הטריגליצרידים בדם (מ״ג לדציליטר) שנמצא בכל מדגם:

	תזונה בריאה	לא תזונה בריאה
מתעמלים	90	100
לא מתעמלים	100	160

א. קיים אפקט עיקרי מובהק לגורם ההתעמלות.

ב. קיים אפקט עיקרי מובהק לגורם התזונה.

ג. קיים אפקט אינטראקציה מובהק בין שני הגורמים במחקר.

ד. אי אפשר לדעת אם קיים אפקט מובהק כלשהו על סמך תוצאות המדגם בלבד ללא ביצוע מבחן מתאים וללא קביעת רמת המובהקות של המחקר.

8) במחקר בדקו 3 טיפולים שונים לחולי פסוריאזיס. המחקר השווה גם בין גברים לנשים ובדק את זמן התגובה לטיפול. מסקנת המחקר הייתה שאצל גברים נמצאו הבדלים מובהקים בין הטיפולים השונים מבחינת תוחלת זמן התגובה. לאיזה סוג אפקט המסקנה מתייחסת?

א. אפקט אינטראקציה.

ב. אפקט עיקרי של גורם המין.

ג. אפקט עיקרי של גורם סוג הטיפול.

ד. אפקט פשוט.

© כל הזכויות שמורות לברק קנדל

9) במחקר בדקו 3 טיפולים שונים לחולי פסוריאזיס. המחקר השווה גם בין גברים לנשים ובדק את זמן התגובה לטיפול. במדגם היה ממוצע זמן התגובה של הגברים שונה מממוצע זמן התגובה של הנשים.

א. אפשר להגיד שבמדגם קיים אפקט עיקרי, אך אי אפשר לדעת אם האפקט העיקרי מובהק.

ב. אפשר להגיד שבמדגם קיימת אינטראקציה, אך אי אפשר לדעת אם האינטראקציה מובהקת.

ג. אפשר להגיד שקיים אפקט עיקרי מובהק.

ד. אפשר להגיד שקיימת אינטראקציה מובהקת.

10) במחקר בדקו 3 טיפולים שונים לחולי פסוריאזיס. המחקר השווה גם בין גברים לנשים ובדק את זמן התגובה לטיפול. אחת המסקנות של המחקר הייתה שהטיפולים השונים משפיעים במידה משמעותית יותר על זמן התגובה של הגברים מאשר על זה של הנשים, אם כי באותו האופן.

א. המסקנה היא שאין אינטראקציה בין הגורמים במחקר.

ב. המסקנה היא שיש אינטראקציה אורדינלית בין הגורמים במחקר.

ג. המסקנה היא שיש אינטראקציה דיסאורדינלית בין הגורמים במחקר.

ד. המסקנה היא שיש אפקט עיקרי של המגדר.

תשובות סופיות

1) א. אין אינטראקציה. ב. אינטראקציה אורדינלית.
ג. אינטראקציה דיסאורדנלית. ד. אינטראקציה דיסאורדנלית.
ה. אינטראקציה אורדינלית.

2) א. המשתנים הב"ת: העיר, סוג הדירה. המשתנה התלוי: מחיר.
ב. קיים. ג. קיים.
ד. לא קיים. ה. קיים.
ו. קיים. ז. קיים.
ח. קיים. ט. קיים.

3) א. עיין בסרטון הוידאו. ב. אינטראקציה דיסאורדינלית.
ג. רעננה ופתח תקווה.

4) א. לא. ב. כן.
ג. לא. ד. לא.

5) א. כן. ב. לא.
ג. לא. ד. לא.

6) לא, לא.

7) ד

8) ד

9) א

10) ב

© כל הזכויות שמורות לברק קנדל

תהליך ניתוח שונות דו כיווני – הליך מבחן

רקע

כפי שכבר ציינו, ניתוח שונות דו-כיווני נעשה כאשר יש שני גורמים מחקריים ומשתנה כמותי תלוי אחד. מטרת המחקר היא לבדוק האם הגורמים משפיעים על המשתנה התלוי. מערך מחקר זה נקרא "מערך מחקר פקטוריאלי", כיוון שאנו בונים את המחקר לפי גורמים. מערך דו-גורמי יסומן כמערך מסוג $A\times B$, כאשר A מייצג את מספר הרמות של גורם a, ו-B מייצג את מספר הרמות של גורם b.

במערך מחקרי תלת-גורמי נסמן את סוג המערך $A\times B\times C$, וכך הלאה.

<u>דוגמה</u>

נבדקו שלושה סוגי דיאטות על אנשים בעלי משקל עודף. נבחרו 18 מטופלים בעלי משקל עודף, 9 מהם גברים ו-9 נשים. המטופלים חולקו כך שבכל דיאטה השתתפו 3 גברים ו-3 נשים. כעבור שלושה חודשים מתחילת הדיאטה נשקלו כלל המטופלים ונבדק המשקל בק"ג שהם הפחיתו. הטבלה הבאה מסכמת את המשקל שכל מטופל במדגם הפחית כעבור שלושה חודשים.

סוג הדיאטה / מין	b_1	b_2	b_3	סה"כ
נשים	8 4 0	6 8 10	4 6 8	54
גברים	6 10 14	0 2 4	9 12 15	72
סה"כ	42	30	54	126

מטרת המחקר היא לבדוק האם יש השפעה של סוג הדיאטה, המין והשילוב ביניהם על ההפחתה במשקל.

- באיזה סוג מערך מחקרי מדובר?
- מהו המבחן הסטטיסטי המתאים לבדיקת ההשערות?
- מה הן השערות המחקר?

© כל הזכויות שמורות לברק קנדל

בדומה לניתוח שונות חד-כיווני גם התהליך של ניתוח שונות דו-כיווני דורש הנחות. ההנחות הן:

1. $A\times B$ הקבוצות שנוצרות בלתי תלויות זו בזו.
2. בכל $A\times B$ האוכלוסיות המשתנה התלוי מתפלג נורמלית.
3. בכל $A\times B$ האוכלוסיות אותה שונות, σ^2.

<u>הערה</u>: ניתוח שונות הוא מבחן רובסטי, כלומר יש לו רגישות נמוכה להנחות. התיאוריה הסטטיסטית שפותחה התבססה על ההנחות האלה, אבל הלכה למעשה השיטה תעבוד טוב גם אם ההנחות הללו לא יתקיימו במדויק במלואן. זו הסיבה שהשיטה הזו נפוצה כל כך בעולם הסטטיסטיקה.

<u>בהמשך לדוגמה</u>

רשמו את כל ההנחות הדרושות לביצוע ניתוח השונות.

© כל הזכויות שמורות לברק קנדל

<u>הליך המבחן</u>

בניית טבלת ממוצעים

נבנה טבלת ממוצעים לכל רמה ולכל תא:

$\bar{X}_{i.}$ – ממוצע המדגם ברמה i של גורם a

$\bar{X}_{.j}$ – ממוצע המדגם ברמה j של גורם b

$\bar{X}_{ij}$ – ממוצע המדגם בתא ij

<u>בהמשך לדוגמה</u>

- מלאו את טבלת הממוצעים הבאה:

סוג הדיאטה / מין	b_1	b_2	b_3	$\bar{X}_{i.}$
נשים				
גברים				
$\bar{X}_{.j}$				

- שרטטו גרפים מתאימים לבדיקת אפקטים עיקריים ולבדיקת אינטראקציה במדגם. האם אפשר להגיד שיש אפקט מובהק?

© כל הזכויות שמורות לברק קנדל

<u>בניית טבלת ריבועי הפרשים מהממוצעים</u>

נמלא את הטבלה הבאה. בתוך תא ij נחשב: $(\bar{X}_{ij}-\bar{X}_{i.}-\bar{X}_{.j}+\bar{X})^2$

<u>בהמשך לדוגמה</u>

- מלאו את טבלת הפרשי הממוצעים:

סוג הדיאטה / מין	b_1	b_2	b_3	$(\bar{X}_{i.}-\bar{X})^2$
נשים				
גברים				
$(\bar{X}_{.j}-\bar{X})^2$				

© כל הזכויות שמורות לברק קנדל

<u>חישוב סכום ריבועי הסטיות מהממוצע</u>

מתוך טבלת ריבועי הסטיות מהממוצע נחשב את סכום ריבועי הסטיות מהממוצע הבאים:

הסימון SS הוא ראשי התיבות של "sum of squares" (סכום הריבועים).

סכום ריבועי הסטיות מהממוצע של גורם a: $SS_a = m \cdot B \sum_{i=1}^{A} (\bar{X}_{i.} - \bar{X})^2$

סכום ריבועי הסטיות מהממוצע של גורם b: $SS_b = m \cdot A \sum_{j=1}^{B} (\bar{X}_{.j} - \bar{X})^2$

סכום ריבועי הסטיות של האינטראקציה: $SS_{ab} = m \sum_{i=1}^{A} \sum_{j=1}^{B} (\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})^2$

סכום ריבועי השגיאות (סכום ריבועי הסטיות של התצפיות בתא מהממוצע בתא):

$$SS_W = \sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{k=1}^{m} (X_{ijk} - \bar{X}_{ij})^2 = (m-1) \sum_{i=1}^{A} \sum_{j=1}^{B} S_{ij}^2$$

סכום ריבועי הסטיות של כלל התצפיות מהממוצע הכללי:

$$SS_T = \sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{k=1}^{m} (X_{ijk} - \bar{X})^2 = (n-1) \cdot S^2$$

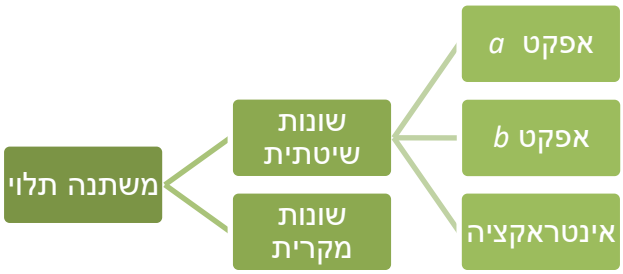
הקשר המתמטי בין סכום הריבועים הללו הוא:

$$SS_T = SS_a + SS_b + SS_{ab} + SS_W$$

לכן אין אנו צריכים לחשב את כל חמשת המרכיבים הללו.

החלק הזה של הנוסחה מתייחס לשונות השיטתית: $SS_a + SS_b + SS_{ab}$. השונות השיטתית היא שונות שמקורה בגורמים עצמם.

החלק הזה של הנוסחה מתייחס לשונות המקרית: SS_W. השונות המקרית היא שונות שנקראת גם "שונות טעויות" או "שונות בתוך הקבוצות". זוהי שונות בין התצפיות שאינה נובעת מהגורמים הנחקרים. האות W מייצגת את המילה "Within", כלומר שונות **בתוך** התאים.



© כל הזכויות שמורות לברק קנדל

<u>בהמשך לדוגמה</u>

- חשבו את ריבועי הסטיות הבאים:

$SS_a =$

$SS_b =$

$SS_{ab} =$

$SS_T =$

$SS_W =$

© כל הזכויות שמורות לברק קנדל

חישוב ממוצע ריבועי הסטיות וסטטיסטי המבחן

MS הוא הסימון של ממוצע ריבועי הסטיות (Mean Square) שמהווה אומד לשונות של כל גורם. החישוב ייעשה על ידי חלוקת ה-SS המתאים בדרגות החופש המתאימות. לאחר מכן נחשב שלושה סטטיסטי מבחן, בהתאם לשלוש ההשערות הנבדקות.

נרכז את כלל החישובים הללו בטבלה הנקראת טבלת ניתוח שונות, ANOVA (Analysis of Variance).

מקור השונות Source of Variation	דרגות החופש Degrees of Freedom	סכום ריבועי הסטיות מהממוצע Sum of Squares	ממוצע ריבוע הסטייה Mean Square	*F*
a	$A-1$	SS_a	MS_a	$F_a = MS_a / MS_W$
b	$B-1$	SS_b	MS_b	$F_b = MS_b / MS_W$
ab	$(A-1)(B-1)$	SS_{ab}	MS_{ab}	$F_{ab} = MS_{ab} / MS_W$
Within	$AB(m-1)$	SS_W	MS_W	
Total	$n\text{-}1=ABm-1$	SS_T		

© כל הזכויות שמורות לברק קנדל

בהמשך לדוגמה: מלאו את טבלת ניתוח השונות

מקור השונות Source of Variation	דרגות החופש Degrees of Freedom	סכום ריבועי הסטיות מהממוצע Sum of Squares	ממוצע ריבוע הסטייה Mean Square	*F*
a				
b				
ab				
Within				
Total				

כללי ההכרעה לבדיקת ההשערות

הסטטיסטי F_a מייצג את היחס בין השונות המדגמית של גורם a ובין השונות המקרית. לכן ככל שהערכים שלו גבוהים יותר, נרצה להגיד שלגורם a יש השפעה גדולה יותר על המשתנה התלוי. F_a יקבל ערכים גבוהים אם השונות המדגמית של גורם A תגדל או אם השונות המדגמית המקרית תקטן.
הסטטיסטי מתפלג התפלגות F, ואזור הדחייה שלו יהיה בצד ימין.

- כלל ההכרעה לבדיקת המובהקות של גורם a:

  דחה את השערת H_0 ברמת מובהקות של α אם

$$F_a > F_{1-\alpha}\left(df_a, df_W\right)$$

לפי אותו עיקרון שאר כללי ההכרעה יהיו:

- כלל ההכרעה לבדיקת המובהקות של גורם b:

  דחה את השערת H_0 ברמת מובהקות של α אם

$$F_b > F_{1-\alpha}\left(df_b, df_W\right)$$

- כלל ההכרעה לבדיקת המובהקות של האינטראקציה:

  דחה את השערת H_0 ברמת מובהקות של α אם

$$F_{ab} > F_{1-\alpha}\left(df_{ab}, df_W\right)$$

© כל הזכויות שמורות לברק קנדל

בהמשך לדוגמה

רשמו את כל כללי ההכרעה המתאימים והסיקו מסקנות מתאימות ברמת מובהקות של 5%.

הערות

1. אם מכריעים שקיימת אינטראקציה מובהקת, יש לבדוק האם היא אורדינלית או דיסאורדינלית. אם האינטראקציה דיסאורדינלית, יש לבדוק האם האפקטים העיקריים נמצאו מובהקים. אם לפחות אחד מהם נמצא מובהק נאמר שהוא אינו משמעותי כיוון שהוא נובע מהאינטראקציה בין הגורמים ולא מהגורם עצמו.
2. אם אחד מהאפקטים נמצא מובהק, אין זה אומר אילו רמות שונות זו מזו בתוחלת. למשל, אם נמצא הבדל מובהק בין סוגי הטיפולים, לא נוכל לדעת לפי זה איזה טיפול שונה מאחר באופן מובהק. לכן יש להמשיך בתהליך של השוואות מרובות כדי להסיק ממה נובע השוני.

בהמשך לדוגמה

האם יש סיבה לבצע השוואות מרובות במחקר?

© כל הזכויות שמורות לברק קנדל

שאלות

1) מחקר שיווקי בדק את השפעת גובה המדף בסופרמרקט והשפעת החומר שממנו עשוי הבקבוק (זכוכית או פלסטיק) על היקף המכירות של משקאות קלים. נבדקו שני סופרמרקטים. בכל סופרמרקט נבחן כל צירוף אפשרי של גובה המדף וחומר הבקבוק, ועבור כל צירוף כזה נבדק מספר בקבוקי המשקה הקל שנמכרו באותו סופרמרקט ביום מסוים. הנה התוצאות שהתקבלו:

סוג בקבוק / גובה המדף	זכוכית	פלסטיק
נמוך	23 32	59 63
בינוני	47 55	88 90
גבוה	40 48	51 56

בצעו ניתוח שונות דו-כיווני על נתוני מחקר זה ברמת מובהקות של 5%. סכמו את המסקנות מתוך ניתוח השונות שביצעתם. מה הן ההנחות הדרושות לביצוע המבחן?

© כל הזכויות שמורות לברק קנדל

2) במחקר בתחום החקלאות נדגמו 8 חלקות אדמה: 4 חלקות בנגב ו-4 בעמק יזרעאל. בכל חלקה ההשקיה הייתה או באמצעות ממטרות או באמצעות טפטפות. בדקו את יבול העגבניות (בטונה לדונם) בכל חלקה.
להלן התוצאות שהתקבלו:

מספר חלקה	מיקום החלקה	שיטת השקיה	יבול העגבניות
1	נגב	ממטרות	12
2	נגב	ממטרות	10
3	נגב	טפטפות	15
4	נגב	טפטפות	17
5	עמק יזרעאל	ממטרות	12
6	עמק יזרעאל	ממטרות	14
7	עמק יזרעאל	טפטפות	17
8	עמק יזרעאל	טפטפות	19

א. רשמו את כלל המשתנים במחקר וציינו לגבי כל אחד מהם האם הוא משתנה תלוי או בלתי תלוי.

ב. הציגו את נתוני המחקר באמצעות גרפים מתאימים. האם נראה שבמדגם יש אפקט עיקרי לכל גורם? האם יש אינטראקציה בין הגורמים במדגם? האם האפקטים מובהקים?

ג. בדקו ברמת מובהקות של 5% האם האפקט העיקרי של כל גורם הוא מובהק והאם האינטראקציה היא מובהקת. מה הן ההנחות הדרושות?

© כל הזכויות שמורות לברק קנדל

3) חברה לייצור מוצרי שיער פיתחה נוסחה חדשנית לצבע לשיער שאינו דורש תוספת חמצן בעת תהליך הצביעה. החברה השוותה את צבע השיער החדש לצבע השיער הרגיל מבחינת כושר הכיסוי וזאת על שלושה סוגי שיער: בהיר, כהה ושיבה. ציון רמת הכיסוי הוא משתנה שמתפלג נורמלית עם שונות קבועה לכל סוג שיער ולכל סוג צבע. לכל קבוצה של סוג צבע וסוג שיער נדגמו 4 צביעות שנוסו על אנשים שונים, וניתן ציון מספרי על רמת הכיסוי.
להלן סיכום תוצאות המדגם שהתקבלו:

הקבוצה	ממוצע	שונות
צבע רגיל על שיער בהיר	62	40
צבע רגיל על שיער כהה	51	44
צבע רגיל על שיער שיבה	45	42
צבע חדש על שיער בהיר	60	46
צבע חדש על שיער כהה	54	40
צבע חדש על שיער שיבה	44	42

בצעו ניתוח שונות דו-כיווני על הנתונים ברמת מובהקות של 5%. סכמו את כל המסקנות המתקבלות.

4) בוצע ניתוח שונות על נתונים. במערך המחקרי לגורם a יש 4 רמות ולגורם b יש 3 רמות. נערכו 3 תצפיות לכל אחת מ-12 הקבוצות שנוצרו. להלן טבלת ניתוח שונות דו-גורמי שבוצע:

מקור השונות	*df*	SS	MS	F
a	?	318	?	?
b	?	?	?	?
אינטראקציה	?	190	?	?
W	?	156	?	
T	?	674		

א. מלאו את כל התאים בטבלה המסומנים בסימני שאלה.

ב. בצעו את הבדיקות הבאות ברמת מובהקות של 5%:

i. האם האינטראקציה מובהקת?

ii. האם גורם a משפיע על המשתנה התלוי הנחקר?

iii. האם לגורם b יש לפחות שתי רמות עם תוחלות שונות?

© כל הזכויות שמורות לברק קנדל

5) במחקר בדקו האם ארץ מוצא ומגדר של אדם משפיעים על שנות ההשכלה שלו. הנתונים סוכמו בטבלת ניתוח שונות:

מקור השונות	df	SS	MS	F
ארץ מוצא	4	34		
מגדר			2	
אינטראקציה		18	4.5	
W	10	12		
T				

א. כמה ארצות מוצא נבדקו במחקר זה?

ב. מהו גודל המדגם הכולל במחקר זה?

ג. חשבו את ערכי F הסטטיסטי עבור ארץ המוצא, המגדר והאינטראקציה.

ד. מה הם האפקטים המובהקים במחקר זה ברמת מובהקות של 5%?

© כל הזכויות שמורות לברק קנדל

6) בטבלה הבאה מסוכמים הממוצעים של מערך מחקרי דו-גורמי עם משתנה כמותי תלוי:

	b_1	b_2	b_3
a_1	8	14	11
a_2	6	13	16

מספר התצפיות בכל תא הוא 5.

הטבלה הבאה היא טבלה מסכמת של ניתוח השונות על סמך נתוני מחקר זה:

מקור השונות	df	SS	MS	F
a				
b		281.7		
ab		71.7		
W		190.1		
T				

א. מלאו את טבלת ניתוח השונות.

ב. הסיקו מסקנות ברמת מובהקות של 5%.

ג. שרטטו גרף אינטראקציות והסבירו את משמעות הממצאים.

תשובות סופיות

1) עיין בסרטון הוידאו.

2) א. משתנים ב"ת: מיקום החלקה, שיטת השקיה. משתנה תלוי: יבול בטונה לדונם.
ב. עיין בסרטון הוידאו. ג. עיין בסרטון הוידאו

3) עיין בסרטון הוידאו.

4) א. עיין בסרטון הוידאו. ב. i. כן. ii. כן. iii. לא.

5) א. 4 ב. 20 ג. עיין בסרטון הוידאו.
ד. עיין בסרטון הוידאו.

6) א. עיין בסרטון הוידאו. ב. עיין בסרטון הוידאו. ג. עיין בסרטון הוידאו.

© כל הזכויות שמורות לברק קנדל