<u>סטודנטים יקרים</u>

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

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

הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי, לדוגמה <u>לחצו כאן</u>.

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

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

אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

גול זה בול. בִּשְׁבִילך

תוכן עניינים

פרק 1 - התפלגויות בדידות מיוחדות - התפלגות בינומית

רקע:

נגדיר את המושג ניסוי ברנולי:

ניסוי ברנולי הנו ניסוי שיש לו שתי תוצאות אפשריות: "הצלחה" ו"כישלון" כמו: מוצר פגום או תקין אדם עובד או מובטל עץ או פלי בהטלת מטבע וכדומה.

בהתפלגות בינומית חוזרים על אותו ניסוי ברנולי n פעמים באופן בלתי תלוי זה בזה.

מגדירים את X להיות מספר ההצלחות שהתקבלו בסך הכול.

נסמן ב p את הסיכוי להצלחה בניסוי בודד וב q את הסיכוי לכישלון בניסוי בודד.

ואז נגיד ש: $X \sim B(n, p)$.

פונקציית ההסתברות של X:

לכל $k = 0,1,2,.........,n \; ; P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

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

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

תוחלת: $E(X) = np$

שונות: $V(X) = npq$

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

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

<u>דוגמה :</u> (פתרון בהקלטה)

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

א. מהי ההסתברות שבדיוק ל-9 מהם יש רישיון נהיגה?

ב. מה ההסתברות שלפחות ל-9 מהם יש רישיון נהיגה?

ג. מהי התוחלת ומהי סטיית התקן של מספר התושבים שנדגמו ושיש להם רישיון נהיגה?

תרגילים:

1. במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסיה.
נגדיר את X להיות מספר המובטלים שהתקבלו במדגם.
א. מהי ההתפלגות של X?
ב. מה ההסתברות שיהיה בדיוק מובטל אחד?
ג. מה ההסתברות שכולם יעבדו במדגם ?
ד. מה ההסתברות ששלושה יעבדו במדגם ?
ה. מה ההסתברות שלפחות אחד יהיה מובטל ?
ו. מה תוחלת ומהי השונות של מספר המובטלים במדגם ?

2. על פי נתוני משרד התקשורת ל-70% מהאוכלוסייה יש סמארט-פון. נבחרו 10 אנשים באקראי. נגדיר את X כמספר האנשים שנדגמו עם סמארט-פון.

א. מהי ההתפלגות של X? הסבירו.
ב. מה ההסתברות שבמדגם ל-8 אנשים יש סמארט-פון?
ג. מה ההסתברות שבמדגם לפחות ל-9 יהיו סמארט-פון?
ד. מה התוחלת ומה סטיית התקן של מספר האנשים שנדגמו ולהם סמארט-פון?

3. במדינה מסוימת התפלגות ההשכלה בקרב האוכלוסייה מעל גיל 30 היא כזו:

השכלה	נמוכה	תיכונית	תואר I	תואר II ומעלה
פרופורציה	0.1	0.6	0.2	0.1

נבחרו 20 אנשים אקראיים מעל גיל 30 מהמדינה הנ"ל.
א. מה ההסתברות ש-5 מהם אקדמאים?
ב. מה התוחלת של מס' בעלי ההשכלה הנמוכה?

4. 5% מקו היצור פגום. המוצרים נארזים בתוך קופסת קרטון. בכל קופסא 10 מוצרים שונים. הקופסאות נארזות בתוך מכולה. בכל מכולה 20 קופסאות.

א. מה ההסתברות שבקופסא אקראית לפחות מוצר פגום אחד?

ב. מה התוחלת ומהי סטיית התקן של מספר הקופסאות במכולה בהן לפחות מוצר פגום אחד?

פתרונות :

שאלה 2:

ב. 0.2335

ג. 0.1493

ד. תוחלת : 7

סטיית תקן : 1.449

שאלה 3:

א. 0.1789

ב. 2

שאלה 4 :

א. 0.401

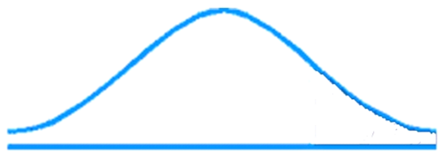
ב. תוחלת : 8.025

סטיית תקן : 2.193

פרק 2 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

רקע:

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



לעקומה זו קוראים גם עקומת גאוס ועקומה אחת נבדלת מהשנייה באמצעות הממוצע וסטיית התקן שלה. אלה הם הפרמטרים שמאפיינים את ההתפלגות.

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

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

כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלבנטים שמתחת לעקומה.
כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון.
התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת והיא תסומן באות Z.

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

תהליך התקנון מבוצע על ידי הנוסחה הבאה:

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

אחרי תקנון מקבלים ערך הנקרא ציון תקן.
ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.
לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי.

ובאופן כללי נתאר את הסכמה הבאה :

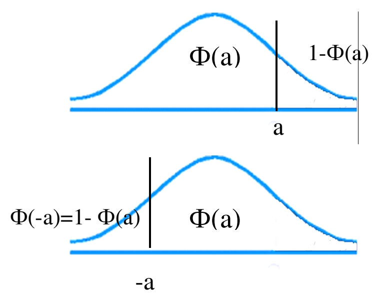
$$X \sim N(\mu, \sigma^2)$$

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

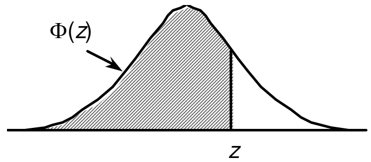
$$Z \sim N(0, 1^2)$$

שימוש בטבלה

P



טבלת ההתפלגות המצטברת הנורמלית סטנדרטית – ערכי Φ(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

<u>דוגמה</u> : (הפתרון בהקלטה)

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

א. מה אחוז חפיסות השוקולד ששוקלות מתחת ל- 110 גרם?

ב. מה אחוז חפיסות השוקולד השוקלות מעל 110 גרם?

ג. מה אחוז חפיסות השוקולד השוקלות מתחת ל 92 גרם?

ד. מהו המשקל ש90% מהחפיסות בקו הייצור שוקלים פחות מהם?

תרגילים:

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

2. נתון שהזמן שלוקח לתרופה מסוימת להשפיע מתפלג נורמלית עם ממוצע של 30 דקות ושונות של 9 דקות רבועות .
   א. מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?
   ב. מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?
   ג. מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?
   ד. מה שיעור המקרים שבהן ההשפעה של התרופה תסטה מ-30 דקות בפחות מ-3 דקות?

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

4. משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטיית תקן 400 גרם.
   א. מצאו את העשירון העליון.
   ב. מצאו את האחוזון ה95.
   ג. מצאו את העשירון התחתון.

5. ציוני מבחן אינטיליגנציה מתפלג נורמלית עם ממוצע 100 ושונות 225 .
   א. מה העשירון העליון של הציונים במבחן האינטיליגנציה?
   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?

6. נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ״ל, נתון ש33% מהבקבוקים הם עם נפח שעולה על 508.8 מ״ל.
   א. מה ממוצע נפח משקה בבקבוק ?
   ב. 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?
   ג. 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?

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

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

א. לאיזו התפלגות הממוצע הגבוה ביותר?

ב. במה מבין המדדים הבאים התפלגות 1 ו 2 זהות?

א. בעשירון העליון.

ב. בממוצע.

ג. בשונות.

ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?

א. 1

ב. 2

ג. 3

ד. אין לדעת.

9. הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטית תקן של 5 דקות.

א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?

ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00 . מה הסיכוי שיאחר לעבודתו?

ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משלושת רבעי השעה . מה ההסתברות שזמן הנסיעה הכולל יהיה פחות מ- 50 דקות?

ד. מה הסיכוי שבשבוע (חמישה ימי עבודה) בדיוק פעם אחת יהיה זמן הנסיעה לפחות שלושת רבעי השעה?

פתרונות :

שאלה 1

א. 89.25%

ב. 2.28%

ג. 0

ד. 50%

שאלה 3

א. 26.43%

ב. 89.44%

ג. 39.44%

ד. 0.383

ה. 100%

שאלה 5

א. 119.2

ב. 80.8

ג. 112.6

ד. 87.4

שאלה 7

א. 500

ב. 100

ג. 0.3446

ד. 733

ה. 267

שאלה 8

א. 3

ב. בממוצע.

ג. 1

שאלה 9

א. 0.1587

ב. 0.0228

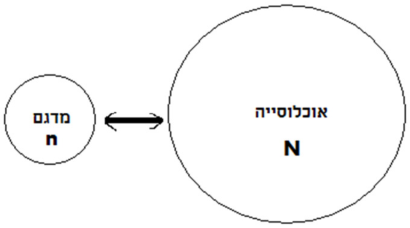
ג. 0.8563

ד. 0.3975

פרק 3 - הסקה סטטיסטית - הקדמה

רקע:

אוכלוסייה – קבוצה שאליה מפנים שאלה מחקרית.



למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.

מדגם – חלק מתוך האוכלוסייה.

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

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי , אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה.

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

סטטיסטי – גודל המחושב על המדגם.

פרמטר – גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים

למשל:

	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?

<u>תרגילים :</u>

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 מאוכלוסיה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.

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

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

3. נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.

א. מהי האוכלוסייה ?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. מהו הסטטיסטי?

פרק 4 - התפלגות הדגימה

ממוצע המדגם ומשפט הגבול המרכזי

רקע:

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

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

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים. להלן רשימה של פרמטרים החשובים לפרק זה :

ממוצע האוכלוסייה נסמן ב μ (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- σ^2 .

סטיית תקן של אוכלוסייה : σ .

א. תכונות התפלגות

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה :

$$E(\bar{x}) = \mu_{\bar{x}} = \mu$$

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב-n . תכונה זו נכונה רק במדגם מקרי :

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

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

$$\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

דוגמה: (פתרון בהקלטה)

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

א. מי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטית התקן של ממוצע המדגם?

ב. דגימה מהתפלגות נורמאלית

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

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

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

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?

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 נורמאלית $\bar{x} \sim\succ N(\mu, \frac{\sigma^2}{n})$.

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?

<u>תרגילים :</u>

1. מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.

   א. מי האוכלוסייה?

   ב. מה המשתנה?

   ג. מהם הפרמטרים?

   ד. מהו גודל המדגם?

   ה. מהו תוחלת ממוצע המדגם?

   ו. מהי טעות התקן?

2. להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים :

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

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

א. בנו את פונקצית ההסתברות של x.

ב. חשבו את התוחלת, השונות וסטיית התקן של x.

ג. אם נדגום 4 משפחות מהישוב עם החזרה מה תהיה התוחלת, מהי השונות ומהי סטיית התקן של ממוצע המדגם?

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

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

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

נתון כי ביום מסוים נולדו 4 תינוקות.

ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק״ג ?

ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק״ג?

ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?

ה. הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

5. הגובה של המתגייסים לצה״ל מתפלג נורמאלית עם תוחלת של 175 ס״מ וסטיית תקן של 10 ס״מ. ביום מסוים התגייסו 16 חיילים.

א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס״מ?

ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס״מ?

ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס״מ?

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

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

א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?

ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?

ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?

ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?

7. נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ״ק וסטיית תקן של 10 סמ״ק.

א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ״ק?

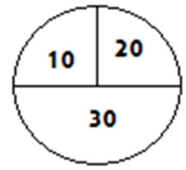
ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ755 סמ״ק?

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?

8. משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4 .
א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
ג. הסבר את ההבדל בתשובות של שני הסעיפים.

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



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.
א. בנו את פונקצית ההסתברות של סכום הזכייה במשחק בודד.
ב. מה התוחלת ומה השונות של סכום הזכייה?
ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?
ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה"כ יזכה ב-1050 ₪ ומעלה?

10. לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪. מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ- 8500 ₪?

11. מטילים קובייה 50 פעמים בכל פעם מתבוננים בתוצאה של הקובייה. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72 ב- 50 ההטלות?

12. אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ .
א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?
ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?
ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. העזר במשפט הגבול המרכזי.

13. נתון משתנה מקרי בדיד בעל פונקצית ההסתברות הבאה:

X	8	6	4	2
P(X)	¼	¼	¼	¼

מתוך התפלגות זו נלקח מדגם מקרי בגודל 50 . מה הסיכוי שממוצע המדגם יהיה קטן מ- 5?

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

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

א. 0

ב. 0.5

ג. 1

ד. לא ניתן לדעת.

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

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

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

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

ג. $\bar{X} \sim N(\mu,\sigma^2)$

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

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

(בחר בתשובה הנכונה)

א. μ ו- $\bar{X}$ יהיו משתנים מקריים.

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

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

ד. μ ו $\bar{X}$ יהיו קבועים.

17. משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם . החפיסות נארזות בקרטון המכיל 36 חפיסות שוקולד אקראיות. ההסתברות שהמשקל הממוצע של חפיסות השוקולד בקרטון יהיה מעל 99 גרם הוא 0.9932.
   א. מהי סטיית התקן של משקל חפיסת שוקולד בודדת?
   ב. מה הסיכוי שמתוך 4 קרטונים בדיוק קרטון אחד יהיה עם משקל ממוצע לחפיסה הנמוך מ-100 גרם?

18. משתנה מקרי כלשהו מתפלג עם סטיית תקן של 20. מה הסיכוי שאם נדגום 100 תצפיות בלתי תלויות מאותה התפלגות אזי ממוצע המדגם יסטה מתוחלתו בפחות מ-2?

פתרונות:

שאלה 2

א.

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

ב. $\mu = 2.05\ \sigma^2 = 0.9475\ \sigma = 0.973$

ג. $\mu_{\bar{x}} = 2.05\ \sigma_{\bar{x}}^2 = 0.2369$

$\sigma(\overline{X}) = 0.486$

שאלה 3

$\mu_{\bar{x}} = 3.5$

$\sigma(\overline{X}) = 1.21$

שאלה 4

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

שאלה 6

א. 0.0465

ב. 27.71

ג. 0.2628

שאלה 7

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

<u>שאלה 8</u>

א. 0.5468

ב. 0.6826

<u>שאלה 9</u>

א.

	10	20	30
P(x)	0.25	0.25	0.5

ב. התוחלת : 22.5

השונות : 68.75

ג. התוחלת : 22.5

השונות : 13.75

ד. 0.8997

<u>שאלה 10</u>

0.0475

<u>שאלה 11</u>

0.1814

<u>שאלה 12</u>

א. 0.9772

ב. 0.0228

ג. 271

<u>שאלה 14</u>

התשובה ב

<u>שאלה 15</u>

התשובה ד

<u>שאלה 16</u>

התשובה ג

<u>שאלה 17</u>

א. 2.429

ב. 0.25

התפלגות סכום תצפיות המדגם ומשפט הגבול המרכזי

רקע:

כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם $T=\sum_{i=1}^{n} X_i$

כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה.

כלומר, היו $X_1,...,X_n$ - משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2 אזי:

א. התוחלת והשונות של סכום התצפיות:

$$E(T)=n\mu$$
$$V(T)=n\sigma^2$$

ב. דגימה מתוך התפלגות נורמלית:

אם $X\sim N(\mu,\sigma^2)$ אזי

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

ג. משפט הגבול המרכזי :

אם x מתפלג כלשהו וידוע
$$E(X)=\mu$$
$$V(X)=\sigma^2$$

אזי עבור מדגם מספיק גדול (לפחות 30)

$$T \sim\succ N(n\mu,n\sigma^2)$$

דוגמה: (פתרון בהקלטה)

בעיר מסוימת המשכורת הממוצעת של עובד הינה 8000 ₪. עם סטיית תקן של 2000 ₪. נדגמו 100 עובדים מהעיר שמפקידים את משכורותיהם לסניף בנק.

א. מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?
ב. מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע"י אותם עובדים? (0.1587)

תרגילים:

1. המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק"ג וסטיית תקן של 10 ק"ג.
   א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק"ג?
   ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק"ג?
   ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק"ג?

2. נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ"ל וסטיית תקן של 20 מ"ל. אדם קנה מארז של 4 בקבוקי יין.
   א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?
   ב. את היין שבמארז האדם מזג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?
   ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א הייתה משתנה? האם התשובה לסעיף ב הייתה משתנה?

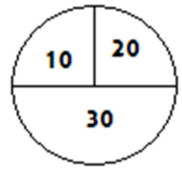
3. בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב 4 דקות עם סטיית תקן של 1 דקות.
   א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?
   ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?

4. במגדל נבנו 40 יחידות דיור. כמו כן נבנו 135 מקומות חנייה לבניין. להלן פונקצית ההסתברות של מספר המכוניות ליחידת דיור:

x	1	2	3	4	5
$P(X=x)$	0.1	0.2	0.3	0.25	0.15

נניח שמספר המכוניות ליחידת דיור בלתי תלויות זו בזו ועם אותה פונקצית הסתברות לכל יחידת דיור (אין צורך בתיקון רציפות).
א. מהי ההסתברות שיהיה מקום בחניון המגדל לכל מכוניות הבניין ?
ב. בהינתן ויש מקום במגדל לכל המכוניות , מה הסיכוי שבפועל מספר המכוניות נמוך מ-130?

5. בקזינו ישנה רולטה עליה מסומנים המספרים הבאים:



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.

א. אם האדם משחק את המשחק 50 פעמים מה ההסתברות שבסך הכול יזכה בסכום של 1050 שקלים ומעלה?

ב. האדם מגיע בכל יום לקזינו ומשחק את המשחק 50 פעם עד אשר מגיע היום בו הוא יזכה ב- 1050 שקלים ומעלה. מה התוחלת ומהי השונות של מספר הימים שיבלה בקזינו?

פתרונות:

שאלה 1

א. 0.6915

ב. 0.8413

ג. 0.5

שאלה 2

א. תוחלת 3000 מ"ל וסטיית תקן 40 מ"ל

ב. 0.0062.

שאלה 4

א. 0.883

שאלה 5

א. 0.8997

ב. תוחלת : 1.111 שונות 0.1239

התפלגות מספר ההצלחות במדגם - הקרוב הנורמלי להתפלגות הבינומית

רקע:

תזכורת על התפלגות בינומית

בפרק זה נדון על התפלגות מספר ההצלחות במדגם אקראי (תצפיות בלתי תלויות זו בזו).

מספר ההצלחות במדגם נסמן ב –Y.

מחלקים כל תצפית במדגם להצלחה או כישלון.

כעת מה שמשתנה מתצפית לתצפית הוא משתנה דיכוטומי (משתנה שיש לו שני ערכים).

תצפית

הצלחה כישלון

הסיכוי להצלחה יסומן עם הפרמטר p וכישלון יסומן ע"י הפרמטר $q = 1 - p$.

מבצעים מדגם אקראי בגודל n .

$Y \sim B(n, p)$

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

תוחלת : $E(y) = np$

שונות : $V(y) = npq$

<u>קירוב נורמלי עבור התפלגות בינומית</u>

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

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

אז :

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

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

<u>תיקון רציפות:</u>

כאשר משתמשים בקירוב הנורמלי להתפלגות הבינומית יש לבצע תיקון רציפות .

הסיבה שעוברים כאן מהתפלגות בדידה להתפלגות נורמלית שהיא התפלגות רציפה.

על פי הכללים הבאים :

1. $p(Y=a) \cong p(a-\frac{1}{2} \leq Y \leq a+\frac{1}{2})$
2. $P(Y \leq a) \cong P(Y \leq a+0.5)$
3. $P(Y \geq a) \cong P(Y \geq a-0.5)$

__הערות:__

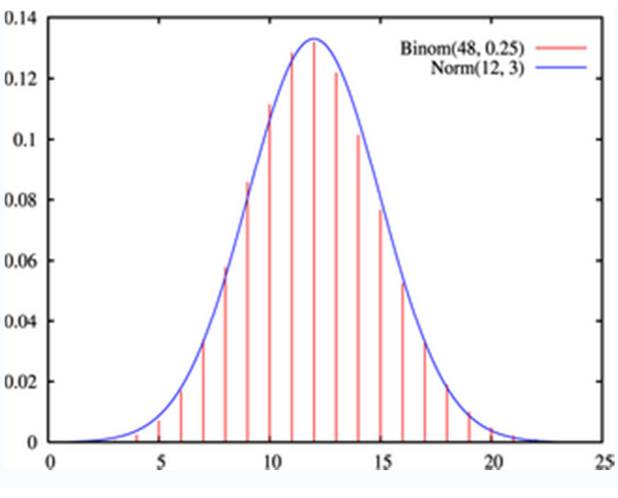
- התנאים למעבר מבינומי לנורמלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n\cdot p\geq 5$
  2. $n\cdot(1-p)\geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n\cdot p\geq 10$
  2. $n\cdot(1-p)\geq 10$
- וישנם מרצים שפשוט התנאי שהם נותנים הוא: $(n\geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור מהתפלגות בינומית לנורמלית.
- הערה נוספת היא לגבי תיקון רציפות. ישנם מרצים שלא מחייבים לבצע תיקון רציפות שהמדגמים גדולים (בדרך כלל מעל 100 תצפיות) אני בפתרונות שאציג תמיד אבצע תיקון רציפות במעבר מבינומי לנורמלי כיוון שכך הפתרון יהיה יותר מדויק (בכל מקרה שהמדגמים גדולים העניין זניח).

__דוגמה:__ (הפתרון בהקלטה)

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

א. מה הסיכוי שבדיוק 14 מתוכם יהיו זקוקים למשקפיים?

ב. מה הסיכוי שלכל היותר 13 מתוכם זקוקים למשקפיים?



תרגילים:

1. נתון ש-20% מאוכלוסייה מסוימת אקדמאית. נבחרו באקראי 10 אנשים באותה אוכלוסייה.

א.מה ההסתברות ששלושה מהם אקדמאים?

ב. מה ההסתברות שלכל היותר אחד מהם אקדמאי?

ג. מה התוחלת ומהי סטיית התקן של מספר האקדמאים במדגם?

2. במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.

א.מה ההסתברות שנדגמו לפחות 6 מוצרים פגומים?

ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?

3. ציוני פסיכומטרי בקרב הנרשמים למוסד מסוים מתפלגים נורמאלית עם ממוצע 500 וסטיית תקן 100. למוסד מסוים הוחלט לקבל אך ורק סטודנטים שקיבלנו מעל 600 בפסיכומטרי. 100 סטודנטים אקראיים נרשמו למוסד. מה ההסתברות שלפחות 20 יתקבלו?

4. מטילים מטבע 50 פעמים.
א. מה ההסתברות לקבל לכל היותר 30 עצים?
ב. מה ההסתברות לקבל 28 עצים לפי התפלגות הבינומית ולפי הקירוב הנורמאלי?

5. במטוס מקום ל-400 נוסעים. נרשמו לטיסה 430 אנשים (overbooking). מנתונים סטטיסטיים ידוע שהסיכוי שאדם שנרשם לטיסה אכן יגיע הוא 0.9.
א. מה ההסתברות שלא יהיו מקומות ישיבה לכל האנשים שהגיעו לטיסה?
ב. מה צריך להיות גודל המטוס כדי שבסיכוי שלפחות 95% המטוס יספיק לכמות הנרשמים?

6. מפעל לייצור ארטיקים טוען ש הסיכוי שארטיק שהוא מייצר יהיה פגום הוא 0.01 . מוכר הזמין 1000 ארטיקים מהמפעל . מה ההסתברות שהמוכר יקבל לפחות 980 ארטיקים תקינים אם טענת המפעל מוצדקת ?

7. מהמר מטיל קובייה הוגנת 100 פעמים. בכל הטלה, אם מתקבל תוצאה זוגית בקובייה המהמר זוכה בשקל. אחרת, המהמר משלם שקל. המהמר הטיל את הקובייה 100 פעמים מה הסיכוי שהרווח של המהמר יהיה לכל היותר 10 ?

פתרונות:

שאלה 1

א. 0.201

ב. 0.3758

ג. התוחלת: 2, סטיית התקן: 1.2649

שאלה 2

א. 0.9332

ב. 0.6915

שאלה 3

0.1611

שאלה 4

א. 0.9406

שאלה 5

א. 0.015

שאלה 6

0.8544

שאלה 7

0.8643

<u>התפלגות פרופורציית ההצלחות במדגם</u>

<u>רקע:</u>

בפרק זה נדון על התפלגות הדגימה של פרופורציית המדגם.

Y- מספר ההצלחות במדגם (למשל , מספר המובטלים במדגם)

$\widehat{p}=\frac{y}{n}$ - פרופורציית ההצלחות במדגם (למשל , שיעור המובטלים במדגם)

למשל,

$n=200$

מספר המובטלים : $Y=20$

פרופורציית המובטלים במדגם $\widehat{p}=\frac{20}{200}=0.1$

נסמן ב- p את שיעור ההצלחה באוכלוסייה וב- q את שיעור הכישלונות באוכלוסייה.

נבצע מדגם מקרי (הנחה שהתצפיות בלתי תלויות זו בזו) ונתבונן בהתפלגות של פרופורציית המדגם.

<u>התוחלת , השונות וסטיית התקן של פרופורציית המדגם:</u>

$$E(\hat{P})=p$$
$$V(\hat{P})=\frac{pq}{n}$$
$$\sigma(\hat{p})=\sqrt{\frac{pq}{n}}$$

<u>משפט הגבול המרכזי עבור הפרופורציה המדגמית :</u>

אם $np\geq 5\,\&\,nq\geq 5$ אזי $\widehat{p}\sim\succ N(p,\frac{pq}{n})$

$$Z_{\widehat{p}}=\frac{\widehat{p}-p}{\sqrt{\frac{pq}{n}}}$$

<u>הערות:</u>

- התנאים לקרוב הנורמאלי הם נזילים, כלומר משתנים ממרצה אחד לשני. התנאי שהצגתי כאן הוא הפופולרי ביותר:
  1. $n\cdot p\geq 5$
  2. $n\cdot(1-p)\geq 5$
- ישנם מרצים שנותנים את התנאי המחמיר הבא:
  1. $n\cdot p\geq 10$
  2. $n\cdot(1-p)\geq 10$
- וישנם מרצים המשתמשים בתנאי: $(n\geq 30)$.
- תאלצו לבדוק מהו התנאי שנתנו לכם בכיתה כדי לעבור לנורמלית.
- כיוון שפרופורציה אינה חייבת להיות מספר שלם בהכרח לא נהוג לבצע כאן תיקון רציפות.

<u>דוגמה:</u> (פתרון בהקלטה)

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

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

<u>תרגילים:</u>

1. במדינה מסוימת 10% מכלל האוכלוסייה הינם מובטלים. נדגמו באקראי 140 אנשים מהמדינה.
   א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
   ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
   ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?

2. נניח כי 30% מהאוכלוסייה תומכים בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
   א. לפחות 35% יתמכו בהצעת החוק במדגם.
   ב. לכל היותר 25% יתמכו בהצעת החוק במדגם.
   ג. יותר מ – 27% יתמכו בהצעת החוק במדגם.

3. לפי נתוני משרד התקשורת 40% מהאוכלוסייה מחזיקים בטלפון נייד מסוג "סמארטפון". נדגמו 400 אנשים מהאוכלוסייה.
   א. מה ההסתברות שבמדגם לכל היותר ל 40% יש סמארטפון?
   ב. מה ההסתברות שבמדגם לרוב יש סמאטרפון?
   ג. מה ההסתברות שפרופורציית בעלי הסמרטפון במדגם תסטה מהפרופורציה באוכלוסייה בלא יותר מ-4%?
   ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם הינו מגדילים את גודל המדגם?

4. נתון כי 80% מבתי האב מחוברים לאינטרנט. נדגמו 400 בתי אב אקראיים.
   א. מה ההסתברות שלפחות 340 מהם מחוברים לאינטרנט?
   ב. מה ההסתברות שפרופורציית המחוברים לאינטרנט במדגם תסטה מהפרופורציה האמתית ביותר מ-4%?
   ג. כמה בתי אב יש לדגום כדי שהסטייה בין הפרופורציה המדגמית לפרופורציה האמתית לא תעלה על 3% בהסתברות של 90%?
   ד. מהו העשירון התחתון של התפלגות פרופורציית המדגם?

5. נתון שציוני פסיכומטרי מתפלגים נורמלית עם תוחלת 500 וסטיית תקן 100. ל"מועדון ה-700" נכללים נבחנים שמקבלים ציון מעל 700 בפסיכומטרי. מה הסיכוי שבמועד בו נבחנו 2000 נבחנים אקראיים יהיו לפחות 3% המשתייכים למועדון?

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

א. הוכיחו ש :

$$E(\hat{P}) = p$$
$$V(\hat{P}) = \frac{P(1-P)}{n}$$

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

פתרונות:

שאלה 1

א. התוחלת: 0.1, השונות: 0.00064

ב. 0.5

ג. 0.3446

שאלה 2

א. 0.0618

ב. 0.0618

ג. 0.8238

שאלה 3

א. 0.5

ב. 0

ג. 0.8968

ד. גדלה

שאלה 4

א. 0.0062

ב. 0.0456

פרק 5 - מושגים בסיסיים באמידה

רקע:

כזכור מהמפגש הקודם פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת.

כמו ממוצע הגבהים בקרב מתגייסים לצה"ל- μ .

כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה - p .

בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת , ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

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

דוגמה: (פתרון בהקלטה)

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

מה הפרמטר בדוגמה זו?

מהי טעות האמידה של ערוץ 10?

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

<u>להלן פרמטרים מרכזיים והאומדים שלהם:</u>

<u>ממוצע האוכלוסייה</u>: 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: $\sigma(\bar{x})=\frac{\sigma}{\sqrt{n}}=SE$

<u>פרופורציה באוכלוסייה</u>: 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: $\sigma(\hat{P})=\sqrt{\frac{p\cdot(1-p)}{n}}$

<u>שונות האוכלוסייה</u>: 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.

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

<u>הערה:</u> אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

<u>דוגמה: (פתרון בהקלטה)</u>

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה. להלן התוצאות שהתקבלו:

2,1,3,2,1,4,5,2,1,3

אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

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

תרגילים:

1. מתוך 500 טירונים נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
   א. מהי האוכלוסייה המוצגת בשאלה ? מהם הפרמטרים שלה?
   ב. מהי טעות התקן של 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. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.

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

   ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.

4. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו:
   $\sum_{i=1}^{20} X_i = 162$ $\sum_{i=1}^{20} X_i^2 = 1502.2$.

   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.

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, $\sum_{i=1}^{n}(x_i-\overline{x})^2$, 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. $X_1, X_2, \ldots\ldots, X_{16}$ 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.

פתרונות:

שאלה 3

א. 177.9
ב. 64.1
ג. 0.4

שאלה 4

א. 8.1
ב. 3.16

שאלה 5

התשובה היא ד.

שאלה 6

התשובה היא ג.

שאלה 7

התשובה היא ד.

שאלה 8

התשובה היא ד.

שאלה 9

התשובה היא ג.

פרק 6 - רווח סמך לתוחלת (ממוצע האוכלוסייה)

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%.

מהי אוכלוסיית המחקר?

מה המשתנה באוכלוסייה?

מה הפרמטר שהחוקר רצה לאמוד?

מהו רווח הסמך?

מה אורך רווח הסמך?

מהי רמת הביטחון של רווח הסמך?

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

הפרמטר שנרצה לאמוד : μ

האומד נקודתי : $\bar{x}$

התנאים לבניית רווח הסמך :

1 $X \sim N$ או $n \geq 30$

2 σ^2 (שונות האוכלוסייה) ידועה

הנוסחה 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.

<u>שגיאת האמידה המקסימלית:</u>

$$\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

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

<u>דוגמה</u> : (פתרון בהקלטה)

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

<u>קשרים מתמטיים ברווח הסמך:</u>

- 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 $z_{1-\alpha/2}$ יותר גבוה, ורווח הסמך יותר ארוך.

תרגילים :

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. בנה 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. הסבר 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 הינו רוצים להקטין את רווח הסמך ב50%?
   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?
5. בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (99,105). שחזרו את :
   א. ממוצע המדגם.
   ב. שגיאת האמידה המקסימאלית.
   ג. רמת הסמך.

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%, מי מהמשפטים הבאים אינו 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.

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? (בחר בתשובה הנכונה)

א.רמת הביטחון.

ב. סטיית התקן באוכלוסייה.

ג. מספר המשתתפים.

ד. סטיית התקן במדגם.

פתרונות :

שאלה 2

$4920.6 < \mu < 4979.4$

שאלה 3

א. $223.42 < \mu < 236.58$

ב. $222.16 < \mu < 237.84$

שאלה 5

א. 102

ב. 3

ג. 0.9544

שאלה 6

א. $3.58 < \mu < 4.42$

ב. יקטן פי 2

ג. גדל

שאלה 7

א. 87

ב. 5

ג. 0.9544

שאלה 8

א. 139

ב. $21 < \mu < 25$

שאלה 9

התשובה היא : ב

שאלה 10

התשובה היא : ג

קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה

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)

תרגילים:

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 $-\mu$
ברמת ביטחון של 0.95 כך שהאורך של הרווח יהיה 0.5σ . מהו גודל המדגם הנדרש?

<u>פתרונות :</u>

<u>שאלה 1</u>

780

<u>שאלה 2</u>

א. 139

ב. הדבר יקטין את ε פי 2.

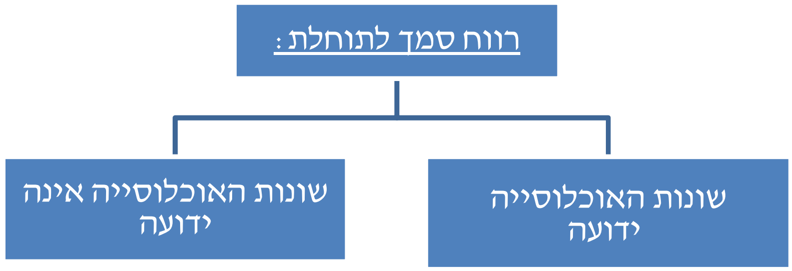
<u>שאלה 3</u>

$n = 62$

רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה

רקע:

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



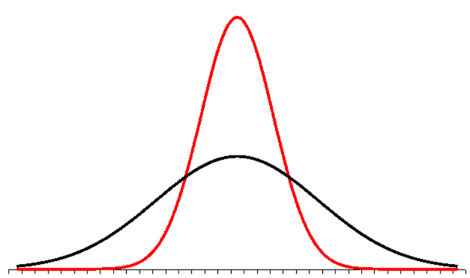
בפרק זה נעסוק במקרה ששונות האוכלוסייה(σ^2) אינה ידועה לנו .מקרה יותר פרקטי .

התנאי : $X \sim N$ או שהמדגם גדול

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

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

התפלגות T :



הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן df=n-1. ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

<u>דוגמה :</u> (פתרון בהקלטה)

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

במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח' . להלן התוצאות שהתקבלו בדקות : 4.7,5.2,4.6,5.3.

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

<u>פתרון :</u>

$4.39 < \mu < 5.51$

תרגילים:

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: 27,34,32,40,30 .
א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים ?

4. ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.
ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.

5. נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$ $\sum_{i=1}^{60} X_i^2 = 643.19$. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

6. נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו.
להלן התוצאות שהתקבלו: $\bar{x} = 13.8$, $S = 2$
בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.

7. שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר **μ**. לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10 . שניהם קבעו אותה רמת סמך.

סטטיסטיקאי א : הניח $\sigma = 20$

סטטיסטיקאי ב : חישב לפי המדגם וקיבל **$S = 20$**

למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר? (בחר בתשובה הנכונה)

א. סטטיסטיקאי א

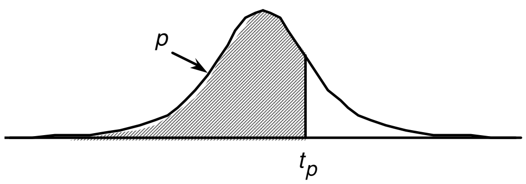
ב. סטטיסטיקאי ב

ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

8. נתון ש : $X \sim N(\mu,\sigma^2)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח הסמך שהתקבל הוא: 8.765 . מהי רמת הביטחון של רווח הסמך?

נספח : טבלת התפלגות T



P

דרגות חופש	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.000	3.078	6.314	12.709	31.821	63.657	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.291

פתרונות:

שאלה 1

א. $79.88 < \mu < 89.72$

שאלה 4

א. $96.63 < \mu < 107.37$

ב. $96.90 < \mu < 107.10$

שאלה 5

$3.149 < \mu < 3.351$

שאלה 8

90%

פרק 7 - רווח סמך לפרופורציה

רקע:

מטרה: לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי: $\hat{p} = \frac{y}{n}$ (Y- מספר ההצלחות שבמדגם)

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

התנאי לבנות את רווח הסמך הינו מדגם של לפחות 30 תצפיות(לעיתים נותנים תנאי של מספר הצלחות ומספר כשלונות לפחות 5 או לפחות 10)

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

מתקיים ש: $\hat{P} = \frac{A+B}{2}$ $\qquad L = 2\varepsilon$

דוגמה: (פתרון בהקלטה)

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

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

ב. מהו האומד לטעות התקן?

פתרון:

א. 7.5%<p<16.5%

ב. 2.29%

<u>תרגילים:</u>

1. נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.
א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.
ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?
ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95 % למספר הדירות בחיפה עם ממ"ד
בפועל.

2. במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.
א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי היי-טק).
ב. כיצד רווח הסמך של סעיף א היה משתנה אם היינו מקטינים את רמת הסמך?
ג. כיצד רווח הסמך היה משתנה אם הינו מגדילים את גודל המדגם?

3. במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים: $0.08 < p < 0.18$
א. כמה נהגים במדגם היו נהגים חדשים?
ב. מהי רמת הסמך של רווח הסמך שנבנה?

4. במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו.
510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של $\pm 3\%$ מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?

5. במדגם של 300 נשים בגילאי 35-40 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.
א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.
ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?

6. ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95% . אורכו הינו 8.3156%.
מהו גודל המדגם שנלקח?

פתרונות:

שאלה 3

א. 52

ב. 0.997

שאלה 5

א. $22.5\% < p < 30.9\%$

ב. $45.91\% < p < 60.72\%$

שאלה 6

200

קביעת גודל מדגם באמידת פרופורציה

רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת:

החוקר קובע מראש את רמת הסמך הרצויה: $1-\alpha$.

החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: ε (או את אורך רווח הסמך).

$L=2\varepsilon$ - אורך רווח הסמך.

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

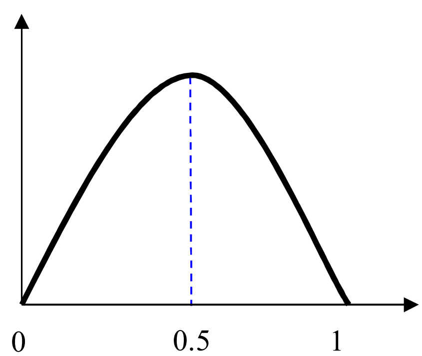
$\varepsilon = z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

נקבל ש: $n \ge \left(\frac{2\cdot Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\hat{p}\left(1-\hat{p}\right)}}{L}\right)^2$

הבעיה שאין אנו יודעים את $\hat{p}$.

נתבונן בביטוי: $\hat{p}\left(1-\hat{p}\right)$:



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

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

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

<u>דוגמה:</u> (פתרון בהקלטה)

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

א. מהו גודל המדגם המינימאלי שיש לקחת?

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

<u>פתרון:</u>

א. 423

ב. 271

תרגילים:

1. הממשלה אומדת מדי חודש את אחוז התמיכה בה. מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?

2. משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.
א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?
ב. חזרו על סעיף א. אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.

3. ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים. המטרה שבביטחון של 95% הסטייה המרבית בין האומד לרייטינג האמתי לא תעלה על 4%.
א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?
ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?

4. השאלות הבאות מתייחסות לסעיף 4 :
א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?
ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.
ג. על סמך סעיף ב'. האם תקבל את הטענה שמיעוט האוכלוסייה תומך הממשלה?

5. משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.
א. כמה מחוסנים יש לדגום ?
ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבת בסעיף הקודם וקיבל ש 15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.
ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98% ? מדוע הוא קטן מ3% ?

פתרונות:

שאלה 1

1068

שאלה 3

א. 601

ב. 108000 ₪.

פרק 8 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

כששונויות האוכלוסייה ידועות

רקע:

מטרה: 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. $\sigma^2{}_1,\sigma^2{}_2$ ידועות.

2. $X_1,X_2\sim N$ או $n_1,n_1>30$

3. שני מדגמים בלתי תלויים.

רווח סמך:

$$(\bar{x}_1-\bar{x}_2)\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}$$

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

דוגמה: (פתרון בהקלטה)

נדגמו 100 תושבים מאזור a והמשכורת הממוצעת הייתה שם 9200 ₪.

כמו כן נדגמו 120 תושבים מאזור b וממוצע המשכורות שהתקבל שם 8700 ₪.

לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ₪.

אמדו ברמת סמך של 90% את הפרש השכר הממוצע בין אזור a לאזור b .

תרגילים:

1. מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?

2. ציוני .I.Q מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99.

א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה"ב בממוצע הציונים במבחן ה-IQ.
ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?

3. חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25 . עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

פתרונות :

שאלה 1

(-20,90)

שאלה 3:

רמות בטחון הגבוהות מ : 0.9476

כששונויות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים

רקע:

מטרה: 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:

$$(\bar{x}_1-\bar{x}_2)\pm t_{1-\frac{\alpha}{2}}^{n_1+n_2-2}\cdot\sqrt{\frac{S^2{}_p}{n_1}+\frac{S^2{}_p}{n_2}}$$

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.
פתרון : (1357,1643)

תרגילים:

1. נדגמו 15 ישראלים ו-15 אמריקאים.

כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

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

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

X	22	20	21	25
Y	18	25	17	12

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

פרק 9 - רווח סמך לתוחלת ההפרש במדגם מזווג

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:

- $x, y \sim N$
- המדגם מזווג

נוסחת רווח הסמך:

$$\bar{D} \pm t^{n-1}_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

כאשר דרגות החופש: $d.f = n-1$

<u>דוגמה:</u> (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.

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

תרגילים:

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

סמסטר א	סמסטר ב
74	80
68	84
90	87
75	76
82	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. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

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

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

פרק 10 - רווח סמך ליחס שונויות

רקע:

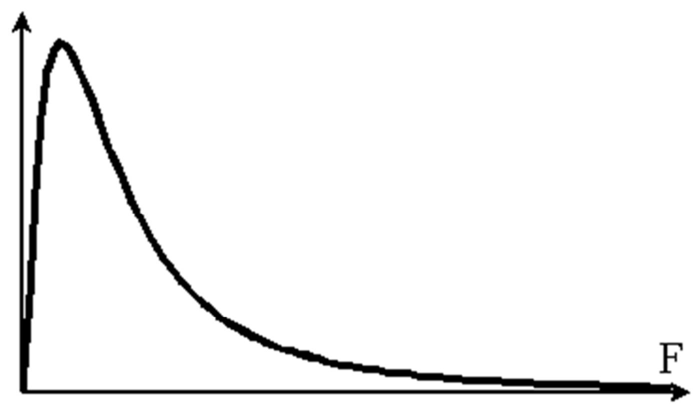
נרצה לאמוד את ההבדל בין שתי שונויות משתי אוכלוסיות שונות.

הפרמטר יהיה : $\frac{\sigma_1^{\,2}}{\sigma_2^{\,2}}$: כלומר היחס בין השונויות.

התנאים :

- $X_1, X_2 \sim N$ או מדגמים גדולים.
- מדגמים בלתי תלויים.

רווח הסמך יבנה על סמך התפלגות הנקראת התפלגות F. התפלגות זו היא אסימטרית חיובית ומושפעת משתי דרגות החופש זו של המונה וזו של המכנה.



$$df_1 = n_1 - 1$$
$$df_2 = n_2 - 1$$

רווח הסמך יהיה :

$$\frac{s_1^{\,2}}{s_2^{\,2}}\frac{1}{F_{1-\frac{\alpha}{2}}\left(n_1-1,n_2-1\right)} \le \frac{\sigma_1^{\,2}}{\sigma_2^{\,2}} \le \frac{s_1^{\,2}}{s_2^{\,2}} F_{1-\frac{\alpha}{2}}\left(n_2-1,n_1-1\right)$$

<u>דוגמה:</u> (פתרון בהקלטה)

מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות:

במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומד חוסר הטיה לשונות ההוצאה החודשית על בילויים 10,000. כמות הסטודנטים שנדגמו 16.

במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומד חסר הטיה לשונות ההוצאה החודשית על בילויים 490,000.

נניח שההוצאה החודשית לבילוי בכל קבוצת גיל מתפלג נורמאלית.

בנו רווח סמך ברמת סמך של 95% ליחס בין השונויות.

תרגילים:

1. בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים הבדל בין שני סוגי המתכות מבחינת שונות החוזק שלהן. דגמו מספר יחידות. מתכת מכל סוג והתקבלו התוצאות הבאות:

סוג המתכת	A	B
N	8	10
$\sum X_i$	16	30
$\sum X_i^2$	60	198

יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.

א. בנו רווח סמך ליחס השונויות של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.

ב. בנו רווח סמך ליחס סטיות התקן של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.

ג. האם בבטחון של 90% ניתן לומר שסטיות התקן של שני סוגי המתכות שונות?

2. מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת.
במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה:

$$\sum (y_i - \overline{y})^2 = 204$$

במדגם של 13 נשים התקבלו התוצאות הבאות:

$$\sum (x_i - \overline{x})^2 = 200$$

אמוד ברמת בטחון של 95% פי כמה גדולה השונות של הגברים באוכלוסייה מהשונות של הנשים.
מה יש להניח לצורך פתרון?

טבלת התפלגות F. ערכי החלוקה F_p של התפלגות $F(m, n)$

m — דרגות חופש המונה; n — דרגות חופש המכנה

		m										
p	n	1	2	3	4	5	6	7	8	9	10	12
.95	1	161	200	216	225	230	234	237	239	241	242	244
.975		648	800	864	900	922	937	948	957	963	969	977
.99		4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106
.995		16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426
.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41
.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41
.99		98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42
.995		198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39	199.39	199.42
.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74
.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34
.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05
.995		55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.39
.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
.975		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75
.99		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37
.995		31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.70
.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
.975		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52
.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89
.995		22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38
.95	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00
.975		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37
.99		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72
.995		18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03
.95	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57
.975		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67
.99		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47
.995		16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18
.95	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
.975		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20
.990		11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67
.995		14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01

m

	p	*n*	1	2	3	4	5	6	7	8	9	10
.95	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07
.975		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87
.99		10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11
.995		13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23
.95	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91
.975		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62
.99		10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71
.995		12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66
.95	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69
.975		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28
.99		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16
.995		11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91
.95	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48
.975		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96
.99		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67
.995		10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25
.95	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28
.975		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68
.99		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23
.995		9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68
.95	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09
.975		5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41
.99		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84
.995		9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18
.95	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92
.975		5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17
.99		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50
.995		8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74
.95	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83
.975		5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05
.99		6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34
.995		8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54
.95	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75
.975		5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94
.99		6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18
.995		7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36

פרק 11 - תרגול מסכם ברווחי סמך

1. 200 אנשים נשאלו כמה פעמים ביום הם שותים כוס קפה. להלן התפלגות התשובות:

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

א. תנו רווח סמך לממוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום. $\alpha = 0.05$

ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז "המכורים לקפה" $\alpha = 0.1$

2. חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה. רווח הסמך שהתקבל הוא $81 < p < 91$ רווח הסמך הנ"ל התבסס על מדגם של 500 איש.

א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?

ב. באיזו רמת סמך נבנה רווח הסמך?

ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 95% על סמך תוצאות המדגם

3. ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן ה-IQ התקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} x_i = 2040$$

$$\sum_{i=1}^{20} x_i^2 = 210740$$

א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה- IQ בישראל – מהי ההנחה הדרושה לפתרון?

ב. על סמך רווח הסמך של סעיף א האם תקבלו את הטענה שבישראל ממוצע הציונים שונה מארה"ב?

ג. מה היה קורה לרווח הסמך אם הינו מגדילים את רמת הסמך שלו?

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

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

ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

פתרונות:

שאלה 1

א. $1.21 \le \mu \le 1.65$

ב. $10.85\% \le p \le 19.15\%$

שאלה 2

א. 70

ב. 99.88%

ג. $83\% \le p \le 89\%$

שאלה 3

א. $97.4 \le \mu \le 106.6$

ב. לא

ג. יגדל

שאלה 4

א. $-372 \le \mu_A - \mu_B \le 772$

ב. $6467 \le \mu \le 7133$

פרק 12 - בדיקת השערות על פרמטרים

הקדמה

רקע:

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

שלב א: נזהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.

השערת האפס המסומנות ב- H_0

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

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

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

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

שלב ד: נרשום את כלל ההכרעה .

בתהליך של בדיקת השערות יוצרים כלל שניקרא **כלל הכרעה :**
הכלל יוצר אזור שניקרא **אזור דחייה** (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) **ואזור קבלה** (קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי .

אזור 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.

דוגמה: (פתרון בהקלטה)

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

$n = 20$

$\overline{X} = 3120$

$S = 280$

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

<u>תרגילים:</u>

1. ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.
   א. מהי אוכלוסיית המחקר?
   ב. מה המשתנה הנחקר?
   ג. מה הפרמטר הנחקר?
   ד. מהן השערות המחקר?

2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק . אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
   א. מהי אוכלוסיית המחקר?
   ב. מה המשתנה הנחקר?
   ג. מה הפרמטר הנחקר?
   ד. מהן השערות המחקר?

3. במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.
   א. מהי אוכלוסיית המחקר?
   ב. מה המשתנה הנחקר?
   ג. מה הפרמטר הנחקר?
   ד. מהן השערות המחקר?

4. בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.
   א. מהי אוכלוסיית המחקר?
   ב. מה המשתנה הנחקר?
   ג. מה הפרמטר הנחקר?
   ד. מהן השערות המחקר?

<u>טעויות בבדיקת השערות</u>

<u>רקע:</u>

בתהליך של בדיקת השערות יוצרים כלל שניקרא **<u>כלל הכרעה</u>** :

הכלל יוצר אזור שניקרא **<u>אזור דחייה</u>** (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) **ו<u>אזור קבלה</u>** (קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי .

בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

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

<u>הגדרת הטעויות:</u>

<u>טעות מסוג ראשון</u>- להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.

<u>טעות מסוג שני</u>- להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.

<u>דוגמה</u> : (פתרון בהקלטה)

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

<u>תרגילים:</u>

1. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק . אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
   א. רשמו את השערות המחקר.
   ב. מה מסקנת המחקר?
   ג. איזו סוג טעות יתכן וביצעו במחקר?

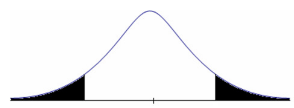
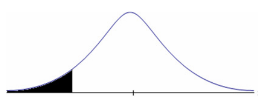
2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
   א. האם ניתן לדעת אם בוצע טעות במחקר?
   ב. מה סוג הטעות האפשרית?

3. לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
   א. מהי אוכלוסיית המחקר?
   ב. מה המשתנה הנחקר?
   ג. מה הפרמטר הנחקר?
   ד. מה השערות המחקר?
   ה. מה מסקנת המחקר?
   ו. מהי סוג הטעות האפשרית במחקר?

פרק 13 - בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסיה 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 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 :

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

חלופה אחרת לכלל הכרעה :

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

<u>דוגמה :</u> (פתרון בהקלטה)

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

תרגילים:

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

2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק . אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
   א. מה מסקנתכם ברמת מובהקות של 2.5%?
   ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?

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

4. המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.

5. לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ.
   קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.

6. במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחר בתשובה הנכונה.
   א. הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
   ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.
   ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
   ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

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

אם החוקר היה עורך מבחן צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח: (בחר בתשובה הנכונה)

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

8. שני סטטיסטיקאים בדקו השערות $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$ עבור שונות ידועה ובאותה רמת מובהקות. שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.
   א. אם חוקר א' החליט לדחות את H_0, מה יחליט חוקר ב'? נמקו.
   ב. אם חוקר א' יחליט לא לדחות את H_0, מה יחליט חוקר ב'? נמקו.

<u>פתרונות :</u>

<u>שאלה 1:</u>

נקבל H_0

<u>שאלה 2:</u>

נדחה H_0

<u>שאלה 3:</u>

נדחה H_0

<u>שאלה 4:</u>

נדחה H_0

<u>שאלה 5:</u>

נקבל H_0

<u>שאלה 6:</u>

ב

<u>שאלה 7:</u>

ג

<u>שאלה 8:</u>

א. אותה מסקנה

ב. לא ניתן לדעת.

סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה

רקע:

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

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות)

$\boldsymbol{\alpha}=P($לדחות את H_0 | H_0 נכונה$)= P_{H_0}($לדחות $H_0)$

הסיכוי לבצע טעות מסוג 2:

$\boldsymbol{\beta}=P($לקבל את H_0 | H_1 נכונה$)=P_{H_1}($לקבל $H_0)$

רמת בטחון:

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

עוצמה :

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

<u>התהליך לחישוב סיכוי לטעות מסוג שני:</u>

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:	3. σ ידועה 4. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0:	$\bar{X} > \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$
חישוב β:	$P_{H_1}(\mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$

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

התקנון: $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

דוגמה : (פתרון בהקלטה)

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

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

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

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

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

תרגילים:

1. נתון ש $X \sim N(\mu,\sigma^2=1)$

להלן השערות של חוקר לגבי הפרמטר μ :

$$H_0: \mu=5$$
$$H_1: \mu=7$$

מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.

א. עבור אילו ערכים של X שידגם נדחית השערת H_0 ?

ב. מה הסיכוי לבצע טעות מסוג שני?

ג. אם במדגם התקבל ש $X=6.9$ מה תהיה המסקנה ומה הטעות האפשרית?

2. לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.

א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.

ב. בהמשך לסעיף א מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?

ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א?

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

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

א. רשום כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.

ב. בהמשך לסעיף א מהי העצמה אם התוחלת שווה ל- 195?

ג. הסבר ללא חישוב איך העצמה תשתנה אם רמת המובהקות תהייה 5%?

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

5. להלן השערות של מחקר

$$H_0 : \mu = 50$$
$$H_1 : \mu = 58$$

מעוניינים לדגום 100 תצפיות . ידוע שסטיית התקן של ההתפלגות הינה 20.
   א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10% . מהי רמת המובהקות?
   ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו) ?
      1. סטיית התקן הייתה יותר גדולה .
      2. הסיכוי לטעות מסוג שני גדול יותר.

<u>השאלות שלהלן הן שאלות רב בררתיות. בחר בכל שאלה את התשובה הנכונה ביותר :</u>

6. אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי :
   א. הסיכוי לטעות מסוג ראשון גדל.
   ב. העוצמה של המבחן גדלה.
   ג. הסיכוי לטעות מסוג שני גדל.
   ד. תשובות א ו-ב נכונות.

7. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :
   א. השערת האפס נכונה.
   ב. השערת האפס נדחתה.
   ג. השערת האפס לא נדחתה.
   ד. אף אחת מהתושבות לא נכונה בהכרח.

8. מה המצב הרצוי לחוקר המבצע בדיקת השערה:

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

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

א. הן α, והן $(1 - \beta)$, יקטנו.

ב. α יישאר ללא שינוי ואילו $(1 - \beta)$ יגדל.

ג. α יגדל ואילו $(1 - \beta)$ יקטן.

ד. הן α והן $(1 - \beta)$ יגדלו.

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

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

פתרונות :

שאלה 1:

א. מעל 6.645

ב. 0.3632

שאלה 2:

א. נדחה H_0 אם $\bar{X} < 2.24$

ב. נדחה H_0

ג. 1

שאלה 3:

א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$

ב. 0.8051

ג. תקטן.

שאלה 4:

א. נדחה H_0 אם $\bar{X} < 48.9$

ב. 0.0885

ג. תקטן.

ד. תקטן.

שאלה 6:

ד

שאלה 7:

ג

שאלה 8:

ג

שאלה 9:

א

שאלה 10:

ב

קביעת גודל מדגם כששונות האוכלוסיה ידועה

רקע:

השערות המחקר הן :

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu = \mu_1$$

סטיית התקן של האוכלוסייה ידועה σ ומעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β.

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

$$n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

דוגמה: (פתרון בהקלטה)

משרד החינוך מפעיל בגן חובה שיטת חינוך שפותחה בשנת 1995. לפי שיטת חינוך זו תוחלת הציון במבחן אוצר מילים לגיל הרך הוא 70. אנשי חינוך החליטו לבדוק שיטת חינוך שפותחה בהולנד הנותנת שם תוחלת ציון אוצר מילים של 80.

נניח שציוני מבחן זה מתפלגים נורמאלית עם $\sigma = 17$.

כדי לבדוק האם גם בישראל הפעלת שיטת החינוך ההולנדית תעבוד בגנים, רוצים לבנות מחקר ברמת מובהקות של 5%. כמו כן, מעוניינים שאם בהפעלת השיטה ההולנדית תוחלת הציונים תעלה לכדי 80, המחקר יגלה זאת בסיכוי של 90%. כמה ילדי גן חובה דרושים למחקר?

תרגילים:

1. במבחן אינטליגנציה הציונים מתפלגים נורמאלית עם סטיית תקן 8 וממוצע 100. פסיכולוג מעוניין לבדוק את הטענה שבאוכלוסיות במצב סוציו אקונומי נמוך תוחלת הציונים היא 95. אם מעוניינים לגלות את הטענה בהסתברות של לפחות 99% כשרמת המובהקות היא 5% מהו גודל המדגם הדרוש?

2. משרד התקשורת טוענים שאדם מדבר בממוצע 180 דקות בחודש בטלפון הסלולרי. חברות הטלפון הסלולרי טוענות שאינפורמציה זו אינה נכונה ואדם מדבר בממוצע פחות : כ-160 דקות. לצורך פתרון נניח שסטיית התקן של זמן השיחה החודשי ידוע ושווה ל-60 דקות. כמה אנשים יש לדגום כך שאם טענת משרד התקשורת נכונה נדחה אותה בסיכוי של 5% (איך קוראים להסתברות זאת?) כמו כן אם טענת חברות הטלפון הסלולרית נכונה המחקר יגלה זאת בסיכוי של 90% (איך קוראים להסתברות זאת?)

3. השערות המחקר הן :
$$H_0: \mu = \mu_0$$
$$H_1: \mu = \mu_1$$

כמו כן נתון שהמשתנה מתפלג נורמלית עם סטיית התקן ידועה σ מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β.

הוכח שגוגל המדגם הרצוי לכך יהיה :

$$n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

פתרונות :

שאלה 1:

41

שאלה 2 :

78

שאלה 3:

הוכחה

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

רקע:

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

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v .

את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.

המסקנה של המחקר תקבע לפי העיקרון הבא :

אם $p_v \le \alpha$ דוחים את H_0

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

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

אם ההשערה היא דו צדדית :

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

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:	5. σ ידועה 6. $X \sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \ge \bar{x})$ אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \ge \bar{x})$

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

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

<u>דוגמה:</u> (פתרון בהקלטה)

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

א. מהי מובהקות התוצאה?

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

תרגילים:

1. לפניך השערות של מחקר:

$$H_0 : \mu = 70$$
$$H_1 : \mu > 70$$

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

$$n = 100$$
$$\bar{x} = 74$$

מהי מובהקות התוצאה?

2. השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪ . מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?

3. אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ – 100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטית 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 מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון ? לא נכון? נמק.

6. בבדיקת השערות התקבל שה- $p\text{-}value=0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.
א. יקבל את השערת האפס בכל מקרה.
ב. ידחה את השערת האפס מקרה.
ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
ד. לא ניתן לדעת כי אין מספיק נתונים.

7. מובהקות התוצאה (PV) היא גם : (בחר בתשובה הנכונה)
א. רמת המובהקות המינימאלית לדחות השערת האפס.
ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

8. בבדיקת השערות מסוימת התקבל p value=0.0254 לכן (בחר בתשובה הנכונה) :
א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0.
ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0.
ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0.
ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0.

פתרונות :

שאלה 1:

0.0228

שאלה 2:

עבור כל רמת מובהקות סבירה.

שאלה 3:

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

שאלה 4:

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיול.

שאלה 5:

נכון

שאלה 6:

תשובה :א

שאלה 7:

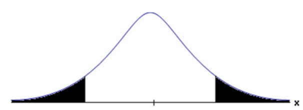
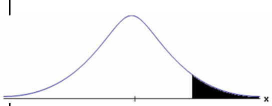
תשובה : א

שאלה 8:

תשובה : ג

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

רקע:

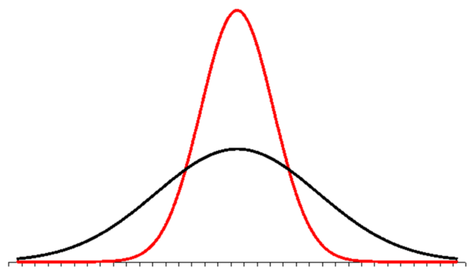
השערת האפס: **השערה אלטרנטיבה:**	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
תנאים:	7. σ אינה ידועה 8. $X \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0:	$t_{\bar{x}} > t^{(n-1)}_{1-\frac{\alpha}{2}}$ או $t_{\bar{x}} < -t^{(n-1)}_{1-\frac{\alpha}{2}}$  $-t_{1-\frac{\alpha}{2},n-1}$ $t_{1-\frac{\alpha}{2},n-1}$ ■ - 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:

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

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

<u>התפלגות T:</u>



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.

<u>דוגמה:</u> (פתרון בהקלטה)

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

א. מהן השערות המחקר?
ב. מה ההנחה הדרושה לצורך פתרון?
ג. בדוק ברמת מובהקות של 5%.

תרגילים:

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

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

$$n = 20$$
$$\bar{x} = 3120$$
$$S = 280$$

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

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

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

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

נתון שההתפלגות היא נורמלית.

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

5. ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם.

ליאור השתמש בטבלה של התפלגות Z.

רוני השתמשה בטבלה של התפלגות t.

מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.

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

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

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

ד. לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

6. נתון ש $X \sim N(\mu,\sigma^2)$ כמו כן נתונות ההשערות הבאות:

$$H_0 : \mu = \mu_0$$
$$H_0 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.

בחר בתשובה הנכונה:

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

פתרונות:

שאלה 1:

נדחה H_0

שאלה 2:

נדחה H_0

שאלה 3:

נקבל H_0

שאלה 4:

נקבל H_0

שאלה 5:

התשובה היא : ב

שאלה 6:

התשובה היא : ג

מובהקות התוצאה (p-value) כאשר שונות האוכלוסייה לא ידועה

רקע:

נזכיר שהמסקנה של המחקר תקבע לפי העיקרון הבא :

אם $p_v \le \alpha$ דוחים את H_0

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

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

אם ההשערה היא דו צדדית :

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס : **השערה אלטרנטיבה:**	$H_0 : \mu = \mu_0$ $H_1 : \mu \ne \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$
תנאים:	9. σ אינה ידועה 10. $X \sim N$ או מדגם מספיק גדול		
p-value	אם $\bar{x} > \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \ge \bar{x})$ אם $\bar{x} < \mu_0 \Leftarrow 2 \cdot P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \le \bar{x})$	$P_{H_0}(\bar{X} \ge \bar{x})$

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

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

$$d.f = n-1$$

<u>דוגמה :</u> (פתרון בהקלטה)

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

א. רשום את השערות המחקר.

ב. מצא חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5% ?

תרגילים :

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 :

   1008,1024,996,1005,997

   א. רשמו את השערות המחקר.

   ב. מהי מובהקות התוצאה? הצג חסמים.

   ג. מה המסקנה ברמת מובהקות של 5%?

2. חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות.

   מהי ה- α המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר ?

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

   $$\bar{x} = 176.2$$

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

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

פתרונות :

שאלה 3:

נקבל H_0

הקשר בין רווח סמך לבדיקת השערות על תוחלת

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: (פתרון בהקלטה)

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו :

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

החוקר בנה רווח סמך ברמה של 90% וקיבל : $79 < \mu < 84$.

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

תרגילים :

1. חוקר רצה לבדוק את ההשערות הבאות:

$$H_0 : \mu = 90$$
$$H_1 : \mu \neq 90$$

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

2. חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ״ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4מ״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? הסבר.

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 של 5 מ״ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.

א. בנה רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.

ב. בדוק ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

פתרונות :

שאלה 1:

1. נקבל השערת H_0

שאלה 2:

א. $112.87 \le \mu \le 118.13$

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

שאלה 3:

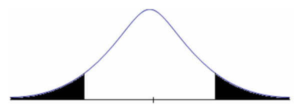
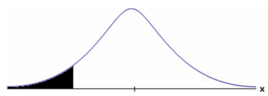
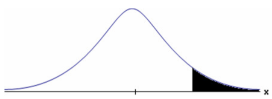
א. $191.8 \le \mu \le 200.2$

ב. נכריע שיש אמת בפרסום.

פרק 14 - בדיקת השערות על פרופורציה

התהליך

רקע:

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

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה:

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

<u>דוגמה:</u> (פתרון בהקלטה)

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

תרגילים:

1. במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. ברמת מובהקות של 5% האם השנה הקשו על תנאי הקבלה?

2. במדגם של 300 אזרחים 57% מתנגדים להצעת חוק מסוימת. לאור נתונים אלה האם רוב האזרחים מתנגדים להצעת החוק ? בדקו ברמת מובהקות של 10%.

3. הטילו מטבע 50 פעמים וקיבלו 28 פעמים עץ. האם המטבע הוגן ברמת מובהקות של 5%?

4. קפיטריה במכללה מסוימת מעריכה כי אחוז הסטודנטים שקונים קפה בקפיטריה הינו 20%. נערך סקר אשר כלל 200 סטודנטים. התברר כי 33 מהם רוכשים קפה בקפיטריה. מטרת הסקר הייתה לבדוק את אמיתות הערכה של הקפיטריה.
   א. רשמו את ההשערות.
   ב. בדוק את ההשערות ברמת מובהקות של 10%.
   ג. מה תהיה המסקנה אם נקטין את רמת המובהקות?

5. חבר כנסת רוצה להעביר חוק. לצורך כך הוא דוגם 400 אזרחים במטרה לבדוק האם רוב האזרחים תומכים בחוק. במדגם התקבל ש-276 אזרחים תומכים בחוק.
   א. מה מסקנתכם ברמת מובהקות של 5%?
   ב. האם ניתן לדעת מה תהיה המסקנה אם רמת המובהקות תהיה גדולה יותר? הסבירו.

6. שני חוקרים בדקו את ההשערות הבאות:

$$H_0 : p = p_0$$
$$H_1 : p > p_0$$

חוקר א השתמש ברמת מובהקות α_1 וחוקר ב ברמת מובהקות α_2 החוקר הראשון דחה את H_0 ואילו החוקר השני קיבל את H_0. שניהם התבססו על אותם תוצאות של מדגם.

בחר בתשובה הנכונה:

א. $\alpha_1 = \alpha_2$

ב. $\alpha_1 > \alpha_2$

ג. $\alpha_1 < \alpha_2$

ד. המצב המתואר לא אפשרי.

פתרונות :

שאלה 1:

נדחה H_0

שאלה 2:

נדחה H_0

שאלה 3:

נקבל H_0

שאלה 4:

ב. נקבל H_0

ג. המסקנה לא תשתנה.

שאלה 5:

א. נדחה H_0

ב. המסקנה לא תשתנה.

שאלה 6:

התשובה היא : ג.

סיכוי לטעויות ועוצמה

רקע:

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

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה :

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

<u>התהליך לחישוב סיכוי לטעות מסוג שני:</u>

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

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

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

דוגמה: (פתרון בהקלטה)

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

א. רשמו את ההשערות וכלל הכרעה ברמת מובהקות של 10%.

ב. מהי עוצמת המבחן אם מסתבר ש 60% מהאוכלוסייה אינם מבקרים אצל רופא שיניים באופן קבוע.

תרגילים:

1. משרד הבריאות פרסם ש 10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.
   א. רשמו את השערות המחקר , וצרו מבחן ברמת מובהקות של 5% לבדיקתן.
   ב. מהי עצמת המבחן של סעיף א׳ בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?
   ג. כיצד תשנה התשובה לסעיף ב׳ אם מסתבר שבחיפה 18% סובלים מאסטמה?
   ד. בהמשך לסעיף א׳ האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?

2. אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
   א. נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10% מהי עצמת המבחן עבור רמת מובהקות של 5%?

3. בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.
   א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.
   ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?

4. מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת.
חוקר א׳ קבע רמת מובהקות של 5% וחוקר ב׳ החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?

5. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן (בחר בתשובה הנכונה)
   א. השערת האפס נכונה.
   ב. השערת האפס נדחתה.
   ג. השערת האפס לא נדחתה.
   ד. אף אחת מהתשובות לא נכונה בהכרח.

6. קבע אם הטענה הבאה נכונה:
״בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני״

פתרונות:

שאלה 1:

ב. 0.9015

ג. תגדל

ד. טענה לא נכונה.

שאלה 2:

0.4404

שאלה 3:

א. 0.1446

ב. תקטן.

שאלה 4:

חוקר א.

שאלה 5:

התשובה הנכונה היא ג.

שאלה 6:

נכונה.

קביעת גודל מדגם

רקע:

השערות המחקר הן :
$$H_0 : p = p_0$$
$$H_1 : p = p_1$$

מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β.

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

$$n \geq \left(\frac{Z_{1-\alpha}\sqrt{p_0 q_0} + Z_{1-\beta}\sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$$

דוגמה: (פתרון בהקלטה)

רוצים לבדוק האם אחוז האנשים השוהים בשמש ללא הגנה ירד בעקבות הפרסומת על נזקי השמש. בעבר 60% מהאוכלוסייה שהתה בשמש ללא הגנה. מה גודל המדגם המינימלי שיש לקחת כדי לבדוק שהאחוז הנ"ל ירד ל48% אם מעוניינים שהסיכוי לטעות מסוג ראשון יהיה 5% והסיכוי לטעות מסוג שני יהיה 1%?

תרגילים:

1. משרד התמ"ת פרסם שאחוז האבטלה במשק היום עומד על 8%. לעומתו, משרד הפנים טוען שחלה עלייה בשיעור האבטלה עד לכדי 11%.
כדי לבדוק מי מביניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:
   - אם משרד התמ"ת צודק, נדחה את טענתו בסיכוי של 10%.
   - אם משרד הפנים צודק, נדחה את טענתו בסיכוי של 4%.

2. מפעיל קזינו מפרסם שהסיכוי לזכות במכונת מזל הינו 0.42.
אדם טוען שהסיכויים לזכות במשחק נמוכים יותר. כמה פעמים יש לשחק את המשחק כדי שאם טענת מפעיל הקזינו נכונה נקבל את טענת האדם בסיכוי של 1% ואם במציאות הסיכוי לזכות במכונה הוא 0.3 נקבל את מפעיל הקזינו בסיכוי של 8%.

פתרונות:

שאלה 1:

891

שאלה 2:

224

מובהקות התוצאה

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.

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

אם ההשערה היא דו צדדית:

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: **השערה אלטרנטיבית:**	$H_0: p = p_0$ $H_1: p \ne p_0$	$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$
תנאים:	$np_0 \ge 5 \,\&\, n(1-p_0) \ge 5$		
p-value	אם $\hat{p} > p_0 \Leftarrow 2 \cdot P_{H_0}(\hat{P} \ge \hat{p})$ אם $\hat{p} < p_0 \Leftarrow 2 \cdot P_{H_0}(\hat{P} \le \hat{p})$	$P_{H_0}(\hat{P} \le \hat{p})$	$P_{H_0}(\hat{P} \ge \hat{p})$

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

והתקנון:

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

דוגמה: (פתרון בהקלטה)

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

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

תרגילים:

1. במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. רוצים לבדוק האם השנה הקשו על תנאי הקבלה.
   א. מהי מובהקות התוצאה?
   ב. מה תהיה המסקנה ברמת מובהקות של 1% וברמת מובהקות של 5%?

2. נהוג לחשוב ש 60% מהילדים בגיל שלוש קמים מהמיטה במהלך הלילה לפחות פעם אחת. ישנה טענה שללא שנת צהריים פחות מ60% מהילדים בגיל זה יקומו לפחות פעם אחת במהלך הלילה. נדגמו 80 ילדים בגיל 3 אשר אינם ישנים בצהריים מתוכם התקבל ש 41 קמו במהלך הלילה.
   א. מהי רמת המובהקות המינימלית עבורה תתקבל הטענה במחקר?
   ב. מהי רמת המובהקות המקסימלית עבורה לא תתקבל טענת המחקר?
   ג. עבור אילו רמות מובהקות נקבל את טענת המחקר?
   ד. מה תהיה מסקנת המחקר ברמת מובהקות של 6%?

3. במטרה לבדוק האם מטבע הוא הוגן מטילים אותו 80 פעמים. התקבל ש 60 מההטלות הראו עץ. רשמו את השערות המחקר, חשבו את מובהקות התוצאה והסיקו מסקנה ברמת מובהקות של 5%.

4. בבדיקת השערות על פרופורציה התקבל שה- $p\text{-}value=0.02$.
   מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 5% : (בחר בתשובה הנכונה)
   א. יקבל את השערת האפס
   ב. ידחה את השערת האפס.
   ג. לא ניתן לדעת כי אין מספיק נתונים.

5. קבע אם הטענה הבאה נכונה :
"במבחן לבדיקת השערות חד-צדדי התקבל ערך *p-value* של 3% לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי) היינו מקבלים ערך *p-value* של 6%"

6. במפעל 10% מהעובדים נפגעים לפחות פעם אחת בשנה מתאונות עבודה. לאור זאת, המפעל החליט לצאת בתוכנית לצמצום שיעור הנפגעים. תכנית זו נוסתה על 100 עובדים. מתוכם 12 נפגעו בתאונות עבודה במשך השנה. מהי רמת המובהקות הקטנה ביותר עבורה יוחלט שהתכנית יעילה?

פתרונות :

שאלה 1:

א. 0.0455

שאלה 2:

א. 0.0548

ב.0.0548

ג. מעל 0.0548

ד. נכריע לטובת טענת המחקר.

שאלה 3:

$p_v = 0$

שאלה 4:

התשובה הנכונה : ב

שאלה 5:

הטענה נכונה

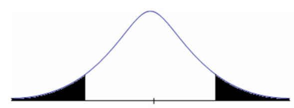
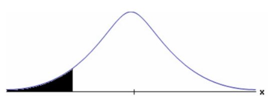
שאלה 6:

0.7486

פרק 15 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונויות של האוכלוסייה ידועות

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. $\sigma_1 , \quad \sigma_2$ ידועות 3. $X_1, X_2 \sim N$ או מדגמים מספיק גדולים		
כלל ההכרעה: אזור הדחייה של H_0 :	$Z_{\bar{x}_1-\bar{x}_2} > Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1-\bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $\quad Z_{1-\frac{\alpha}{2}}$ ■ - דוחים את H_0	$Z_{\bar{x}_1-\bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ ■ - דוחים את H_0	$Z_{\bar{x}_1-\bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ ■ - דוחים את H_0

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה:

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

התפלגות הפרש הממוצעים :

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

התקנון :

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

<u>דוגמה</u> : (פתרון בהקלטה)

בשנת 2004 הפער בין השכר הממוצע של הגברים לנשים היה 3000 ₪ לטובת הגברים.

מעוניינים לבדוק האם כיום הצטמצם הפער בין הגברים לנשים מבחינת השכר הממוצע.

נדגמו 100 עובדים גברים. שכרם הממוצע היה 9,072 ₪. נדגמו 80 עובדות, שכרן הממוצע היה 7809 ₪. לצורך פתרון נניח שסטיות התקן של השכר ידועות ושוות ל-2000 ₪ באוכלוסיית הנשים ו-3000 ₪ באוכלוסיית הגברים. מה המסקנה ברמת מבוהקות של 5%?

תרגילים :

1. מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלו נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה.
במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות.
במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות.
לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.

2. ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.

3. במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.
א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.
ב. מה תהיה המסקנה ברמת מובהקות של 8%?

4. במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע.
מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו 40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים. ו-12 ס"מ אצל הגברים.
א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?
ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

פתרונות:

שאלה 1:

נדחה H_0

שאלה 2:

לא נדחה את H_0

שאלה 3:

א. 0

ב. נדחה H_0

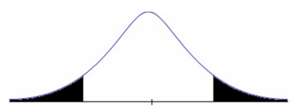
שאלה 4:

א. נדחה H_0 אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס"מ.

ב. 0.6331

כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות

רקע:

השערת האפס: **השערת 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:	4. מדגמים בלתי תלויים 5. $\sigma_1 ,\ \ \sigma_2$ לא ידועות אך שוות 6. המשתנים בכל אוכלוסייה מתפלגים נורמלית		
אזור הדחייה של H_0:	$t_{\bar{x}_1-\bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1-\bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2},n_1+n_2-2}$ $\quad t_{1-\frac{\alpha}{2},n_1+n_2-2}$ ■ - דוחים את H_0	$t_{\bar{x}_1-\bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha,n_1+n_2-2}$ ■ - דוחים את H_0	$t_{\bar{x}_1-\bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha,n_1+n_2-2}$ ■ - דוחים את H_0

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:

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

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%.

תרגילים:

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

2012	120	112	95	130	90	94	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	2-60W	1-100W
$\bar{x}$	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 60דולקות בממוצע יותר מאשר נורות מסוג W 100?

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

פתרונות:

<u>שאלה 1:</u>

לא נדחה H_0

<u>שאלה 2:</u>

<u>שאלה 3:</u>

א. נדחה H_0

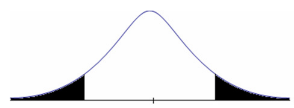
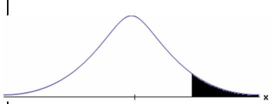
ב. רמות מובהקות של לפחות 5%

ג. לא נדחה H_0

פרק 16 - 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:	7. σ_D אינה ידועה 8. $D \sim N$ או מדגם מספיק גדול		
כלל ההכרעה: אזור הדחייה של H_0 :	$t_{\bar{D}} < -t^{(n-1)}_{1-\frac{\alpha}{2}}$ או $t_{\bar{D}} > t^{(n-1)}_{1-\frac{\alpha}{2}}$  $-t_{1-\frac{\alpha}{2},n-1}$ $t_{1-\frac{\alpha}{2},n-1}$ ■ - דוחים את H_0	$t_{\bar{D}} < -t^{(n-1)}_{1-\alpha}$  $-t_{1-\alpha,n-1}$ ■ - דוחים את H_0	$t_{\bar{D}} > t^{(n-1)}_{1-\alpha}$  $t_{1-\alpha,n-1}$ ■ - 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 :

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

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

דוגמה : (פתרון בהקלטה)

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

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

להלן המחירים :

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

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

תרגילים:

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

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

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

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

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

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

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

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

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

א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ 5 נקודות?

ב. על סמך הסעיף הקודם , מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ- 5 נקודות?

ג. לאור זאת, מה המסקנה ברמת מובהקות של 10% ?

4. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית.

המבחן שיש לבצע כאן הוא:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

משקל במכשיר 1	5.4	6.9	7.0	5.2
משקל במכשיר 2	5.3	6.9	7.1	5.0

נניח שהמשקלים מתפלגים נורמלית.

המבחן שיש לבצע כאן הוא:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

6. כדי להשוות בין שני אצים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים.

המבחן שיש לבצע כאן הוא:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

<u>פתרונות:</u>

<u>שאלה 1:</u>

לא נדחה H_0

<u>שאלה 2:</u>

לא נדחה H_0

<u>שאלה 3:</u>

א. $0.1 \le pv \le 0.25$

ב. 0.25

ג. לא נדחה H_0

<u>שאלה 4:</u>

התשובה היא ד.

<u>שאלה 5:</u>

התשובה היא ד.

<u>שאלה 6:</u>

התשובה היא ג.

פרק 17 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע:

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

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

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 :

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

החוקר בנה רווח סמך ברמה של 90%

$$78<\mu_D<83$$

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

תרגילים:

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

סטטיסטיקה א	סטטיסטיקה 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 אמריקאיות:

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

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

אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:

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

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

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

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

4. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ״ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו :

$-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%.

לכן מסקנת המחקר היא :

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.

פתרונות:

שאלה 1:

א. $-3.8 \le \mu_D \le 19$

ב. נכריע שיש אמת בפרסום.

שאלה 2:

א. $-8.5 \le \mu_X - \mu_Y \le 26.5$

ב. נכריע שאין הבדל.

שאלה 3:

התשובה היא ג.

שאלה 4

התשובה היא א.

פרק 18 - בדיקת השערות - מבחנים פרמטרים

בדיקת השערות על שתי שונויות.

1. מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה:

$$\sum (y_i - \overline{y})^2 = 204$$

במדגם של 13 נשים התקבלו התוצאות הבאות:

$$\sum (x_i - \overline{x})^2 = 200$$

בדקו ברמת מובהקות של 2% האם קיים הבדל בין השונויות? מה יש להניח?

2. ציוני בחינת הבגרות במתמטיקה מתפלגים נורמלית עם שונות 150. במדגם של 16 נבחנים מתל אביב התקבלה שונות חסרת הטיה- 190. במדגם של 25 ירושלמים התקבלה שונות חסרת הטיה- 118.

א. בדקו ברמת מובהקות של 5% האם שונות הציונים במתמטיקה בקרב נבחני תל אביב גבוהה מהשונות בכלל הארץ.

ב. בדקו ברמת מובהקות של 5% האם שונות ציונים במתמטיקה בקרב תלמידי תל אביב גבוהה מאשר בקרב תלמידי ירושלים.

פתרונות:

בדיקת השערות לשתי שונויות

שאלה 1	שאלה 2
נקבל H_0	א. נקבל H_0 ב. נקבל H_0

פרק 19 - מבחני חי בריבוע

א. מבחן טיב התאמה

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

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

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

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

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

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

ב. מבחן אי תלות

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

	לילה	ערב	יום
פגומים	70	60	50
תקינים	800	700	600

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

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

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

צבע	שחור	לבן	אדום	כחול
מספר הפריטים	15	20	15	50

כמו כן בדקו את התפלגות הצבעים בחנות שכנה B :

צבע	שחור	לבן	אדום	כחול
מספר הכדורים	60	20	10	20

א. בדוק ברמת מובהקות של 5% האם התפלגות הצבעים בחנות A היא ביחס של 1:1:1:3 לטובת הכחול.

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

<u>תשובות סופיות - מבחני חי בריבוע</u>

<u>פרק א׳- מבחן טיב התאמה</u>

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

<u>פרק ב׳ – מבחן לאי תלות</u>

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

פרק 20 - מדדי קשר - מדד הקשר הלינארי (פירסון)

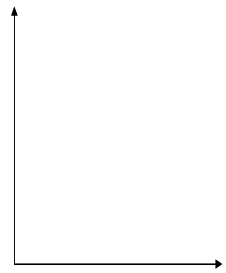
רקע:

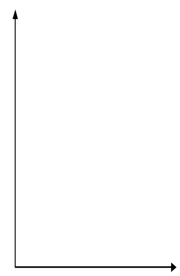
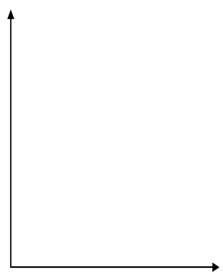
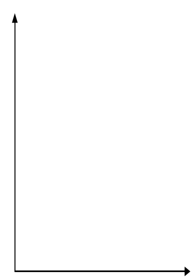
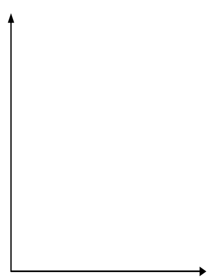
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים.
מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה.
בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי).למשל,
נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X– מסבירה את ההכנסה שלו Y.
במקרה זה שנות ההשכלה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

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

נשרטט מנתונים הללו דיאגרמת פיזור :

נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :



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

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

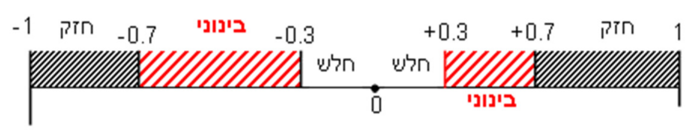
מקדם מתאם זה מקבל ערכים בין 1- ל 1.

מקדם מתאם1- או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה : $y = bx + a$.

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

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

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



מקדם המתאם יסומן באות – r .

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

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

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

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

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

תרגילים:

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

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

א. שרטט דיאגראמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר ביו מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?

ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?

ג. הסבר ללא חישוב כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

2. במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון X בדם החולה לרמת ההורמון Y שלו. לצורך כך מדדו את רמת ההורמונים ההלו עבור חמישה חולים.

להלן התוצאות שהתקבלו:

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

א. מה הממוצע של כל רמת הורמון?

ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

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

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

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

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

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

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

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

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

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

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

חשב את מדד הקשר הלינארי בין X ל-Y.

5. במוסד אקדמי ציון ההתאמה מחושב כך : מכפילים את הציון הממוצע בבגרות ב- 3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה2.
מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם ?

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

א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.

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

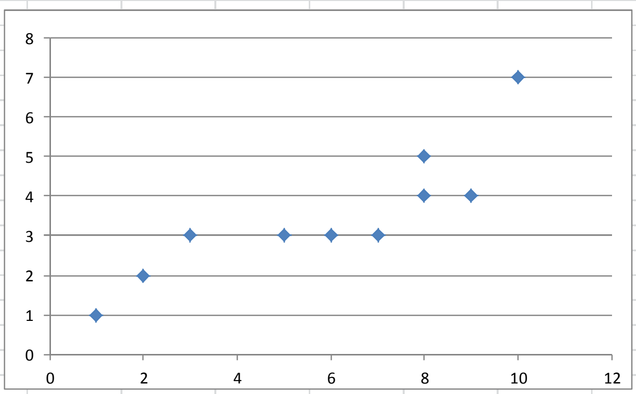
ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

<u>שאלות אמריקאיות:</u>

7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :
   א. הדבר מעיד שהציונים בכתה היו שליליים.
   ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
   ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
   ד. אף אחת מהתשובות לא נכונה.

8. נלקחו 20 מוצרים וניבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה - 4.2 ₪) מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?
   א. 1
   ב. 0
   ג.4.2
   ד. לא ניתן לדעת.

9. להלן דיאגראמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים?

   א. 1
   ב. 0.85
   ג. 0.15
   ד. 0

פתרונות:

שאלה 1:

א. בהקלטה

ב. -0.9325

שאלה 2:

א. $\bar{y} = 16$ $\bar{x} = 15.4$

ב. $r_{xy} = 0.96$

שאלה 3:

א: 0.8

שאלה 4:

0.8

שאלה 5:

1

שאלה 6:

א. נכון

ב. לא נכון

ג. נכון

שאלה 7:

התשובה: ג

שאלה 8:

התשובה: א

שאלה 9:

התשובה: ב

פרק 21 - מדדי קשר - רגרסיה ליניארית

רקע:

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

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

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

b - הוא שיפוע הקו נותן בכמה בעצם Y משתנה כאשר X גדל ביחידה אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה:

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

$$b = r\frac{S_Y}{S_x}$$

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

אם נרצה לבנות קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

תרגילים:

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

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

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

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

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

ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבר את משמעות הפרמטרים של קו הרגרסיה.

ג. משפחת כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

2. נסמן ב-X את ההשכלה של אדם בשנות למוד. נסמן ב-Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_Y = 5 \qquad S_X = 2$$

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

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

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

ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?

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

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

א. על פי משוואת הרגרסיה שעת הכנה נוספת משפרת את ציון המבחן ב?

ב. על פי משוואת הרגרסיה תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון ?

ג. מהו קו הרגרסייה לניבוי הציון לפי שעות ההכנה?

4. נתונים 2 משתנים Y,X . כמו כן נתון: X ממוצע = 1.5, שונות X = שונות Y = 4,וכן שקו הרגסיה של Y על בסיס X הינו $Y= -0.2X + 0.5$. חשב מהו מקדם המתאם בין X ל $-Y$?

פתרונות:

שאלה1:

א. 0.8

ב. $\tilde{Y} = 0.8X + 0.4$

ג. 12.4

שאלה 2:

א. 0.75

ב. 4.25 אלפי ש"ח

ג. 14.6 שנים

שאלה 3:

א. 1.2

ב. 29

ג. y=1.2x+29

שאלה 4:

-0.2

פרק 22 - מדדי קשר - רגרסיה - שונות מוסברת ושונות לא מוסברת

רקע:

המטרה ברגרסיה הנה להסביר את השונות של המשתנה התלוי.
למשל, להסביר את השונות של המשכורות באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

r^2 -נותן בעצם איזה חלק מהשונות של המשתנה התלוי מוסבר.
השונות המוסברת נקראת גם שונות ניבויים.
השונות הלא מוסברת נקראת גם שונות טעויות.

תרגילים :

1. נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן נתון שסטיית התקן של מחירי הדירות הינה 200.

א. איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?
ב. איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?
ג. מהי השונות המוסברות ומהי השונות הלא מוסברת של מחירי הדירות?

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

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

שאלות אמריקאיות:

בשאלות הבאות יש לבחור בתשובה הנכונה.

3. בקשר בין שני משתנים התקבל $r^2 = 0.64$ לכן :

א. ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.
ב. 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.
ג. הקשר בין שני המשתנים הוא בעוצמה של 0.64.
ד. כל התשובות נכונות.

4. אם מגדילים את r^2 מה ניתן לומר?

א. אחוז השונות המוסברת יקטן
ב. אחוז השונות המוסברת יגדל
ג. אחוז השונות המוסברת יישאר ללא שינוי.
ד. סטיית התקן משתנה
ה. לא ניתן לדעת

5. בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים : מבחן בסוף סימסטר א (X) ומבחן בסוף סימסטר ב (Y) . כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב לפי הציון במבחן סוף סמסטר א התקבלה שונות טעויות של 80 , ושונות ניבויים של 20 . לפי נתונים אלו מקדם המתאם בין הציון במבחן סוף סמסטר א לבין הציון במבחן סוף סמסטר ב הוא :

א. 0.44 .

ב. 0.44 - .

ג. עוצמת ההקשר הלינארי היא 0.44 , אך אין אפשרות לדעת את סימנה.

ד. אין אפשרות לחשב את מקדם המתאם.

ה. 0.35