

יוסי גודלניק

יישומי סטטיסטיקה

בגיליון אלקטרוני EXCEL

	A	B	C	D	E	F	G	H
1		ציוני המבחן						השכים
2		15	18	22	25	33		15
3		15	19	23	25	34		19
4		15	19	23	26	34		23
5		15	19	23	26	34		34
6		16	19	23	26	34		24
7		16	20	24	27	37		#N/A
8		16	20	24	27	35		#N/A
9		17	20	24	30	35		#N/A
10		18	21	24	33	35		#N/A
11		18	21	25	33	36		#N/A

15 \rightarrow =MODE.SNGL(B2:F11)

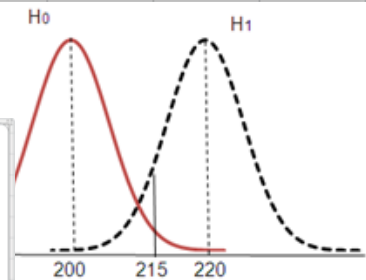
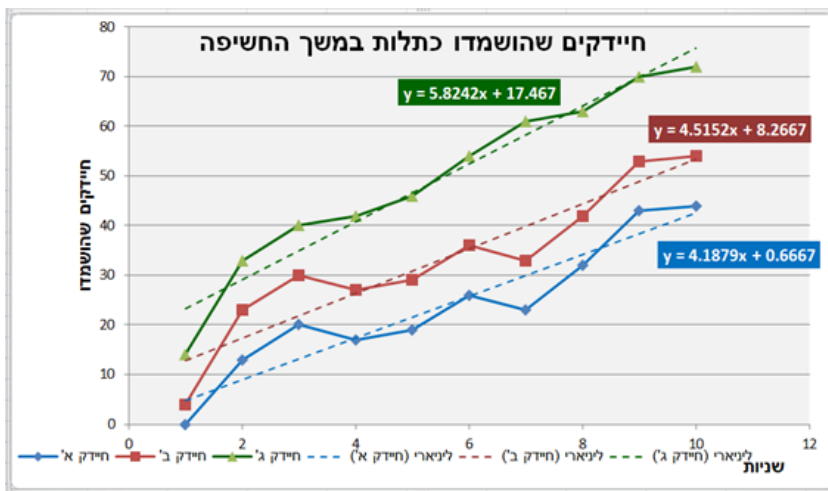
Anova: Single Factor

Input Range: \$C\$4:\$E\$8

Labels in first row:

Alpha: 0.05

Output Range: \$M\$4



© כל הזכויות שמורות למחבר יוסי גודלניק

המידע ניתן as is. נעשו מאמצים כדי שהחוברת תהיה אמינה ככל שניתן אך אין משתמעת מכך כל אחריות שהיא. המחבר והוצאת הוד-עמי אינם אחראים כלפי יחיד או ארגון עבור כל אובדן או נזק אשר ייגרם, אם ייגרם, מהמידע שבחוברת זו.

אין להעתיק, לשכפל, לצלם, להקליט, לתרגם, לאחסן במאגר מידע או להפיץ חוברת זו או קטעים ממנה, בשום צורה ובשום אמצעי אלקטרוני, אופטי או מכני, בין אם לשימוש פנימי ובין אם לשימוש מסחרי מבלי לקבל רשות מפורשת בכתב מבעל הזכויות.

לשם שטף הקריאה כתובה חוברת זו בלשון זכר בלבד. החוברת מיועדת לגברים ולנשים כאחד ואין בכוונתנו להפלות או לפגוע בציבור המשתמשים/ות.

הפצה:

הוצאת הוד-עמי

09-9564716

info@hod-ami.co.il www.hod-ami.co.il

מסת"ב 978-965-361-415-4 ISBN

מוקדש באהבה לנכדיי

שירה, נועה, גיא ועומר

תוכן עניינים

5	פרק 1 – פתח דבר
6	פרק 2 – הכרת אקסל
6	העתקה [סוגים]
10	יצירת סדרה חשבונית
10	שינוי תוכן של תא
11	סימני פעולה
11	סדר הפעולות
12	ביצוע חישוב
12	כתיבה והעתקת נוסחה
13	הפניה [יחסית, מוחלטת, מעורבת]
16	פרק 3 – פונקציות מובנות
18	פרק 4 – פונקציות סטטיסטיות מובנות
20	פרק 5 – סטטיסטיקה תיאורית
21	הפונקציה MODE
22	הפונקציה PERCENTILE
24	הפונקציה FREQUENCY
25	HISTOGRAM
28	DESCRIPTIVE STATISTICS
30	פרק 6 – התפלגויות
30	התפלגות נורמלית
34	התפלגות t
38	התפלגות בינומית
40	התפלגות היפרגאומטרית
42	התפלגות פואסונית
44	התפלגות מעריכית
46	פרק 7 – סינון נתונים
48	סינון של "גם וגם"
48	סינון של "או/ו או"
50	פרק 8 – פונקציית התניה – IF
51	פונקציית התניה מורכבת
52	פרק 9 – פונקציות סטטיסטיות משולבות התניה
52	הפונקציה AVERAGEIF
52	הפונקציה SUMPRODUCT
56	פרק 10 – טבלאות ציר
58	סינון דוחות
59	פרק 11 – מבחני השערות ורווח סמך
59	מבחן על תוחלת כאשר השונות ידועה
61	מבחן על תוחלת כאשר השונות לא ידועה
63	מבחן על תוחלות של שתי אוכלוסיות, שונויות לא ידועות
66	מבחן על הפרש תוחלות של מדגמים מזווגים
69	פרק 12 – רגרסיה ליניארית
78	פרק 13 – מבחן חי בריבוע
81	פרק 14 – ניתוח שונות חד כיווני

אקסל הוא גיליון אלקטרוני המאפשר בניית מסד נתונים, ייבוא מסד נתונים, ביצוע חישובים הנוגעים לנתונים, לרבות חישובים סטטיסטיים, הצגת הנתונים בחתכים שונים, הצגה גרפית של הנתונים וכפי שכתוב באתר MICROSOFT OFFICE "ניתוח, שיתוף וניהול של מידע לקבלת החלטות מושכלות יותר".

ישנן תוכנות סטטיסטיות, חלקן חינמיות כמו PSPP [התוכנה מהווה תחליף חינמי לתוכנת SPSS ודומה לה מאוד] ו-IDAMS [התוכנה פותחה ע"י UNESCO] וחלקן בתשלום כמו SAS ו-SPSS המשמשות גופים רבים לרבות אוניברסיטאות. תוכנות אלו "חזקות" הרבה יותר מאשר האקסל מציע בנושא הסטטיסטיקה, אולם האקסל זמין יותר לכול וגם נותן מענה הולם ללומדים בקורסי טכנאים, הנדסאים ואפילו לסטודנטים במדעי החברה.

החומר שיובא, במסגרת חוברת זו, מניח שלמשתמש ידע בסיסי ב-OFFICE [פתיחת קובץ, סגירת קובץ וכו'] וכן ידע בסיסי באקסל – למרות זאת מובאת חזרה קצרה לגבי פונקציות בסיסיות באקסל. ההתייחסות תהיה לגרסאות EXCEL 2007-2010.

חשוב להדגיש כי חוברת זו **אינה** מהווה ספר לימוד לנושא הסתברות וסטטיסטיקה ורצוי לפנות לספר לימוד מתאים, להשתתף בהרצאות ובמקביל לשפר את מיומנויות החישוב בעזרת חוברת זו.

החוברת פונה לסטודנטים הלומדים בקורסי טכנאות והנדסאות לסוגיהם אך גם מתאימה לכל מי שלומד אקסל וסטטיסטיקה.

סדר פרקי החוברת אינו שגרתי והוא משלב, לפי הצורך, את הפונקציות האקסליות והסטטיסטיות.

בהמשך לתרגיל מספר 1, נרצה עכשיו לא רק להציג את הערכים ששכיחותם היא הגבוהה ביותר אלא את תדירות/שכיחות הופעתם של ערכים בתוך טווח מוגדר. זהו למעשה התחלת תהליך של **קיבוץ נתונים**. אפשר לבצע זאת [עם הפונקציות המובנות] בשתי דרכים. נתאר כאן את שתיהן:

I2											
fx: {=FREQUENCY(B2:F11,H2:H6)}											
	B	C	D	E	F	G	H	I	J		
1	ציוני המבחן						BINS - תאים		שכיחות		
2	26	35	15	20	34		18	11			
3	18	23	18	15	24		23	15			
4	19	33	23	33	20		28	11			
5	19	18	27	26	27		33	4			
6	19	22	34	25	34		38	9			
7	21	23	21	24	30						
8	24	23	25	16	34						
9	25	15	16	35	33						
10	26	16	21	35	15						
11	34	34	19	19	17						
12		15		35							
13		=MIN(B2:F11)		=MAX(B2:F11)		=FREQUENCY(B2:F11,H2:H6)					
14											

- שלב ראשון – לקבוע את הנתון שערכו הקטן ביותר והנתון שערכו הגדול ביותר.
- שלב שני – לקבוע את מספר התאים [BINS] וערך כל אחד מהם, אנו יודעים כי בטווח יש 50 נתונים, צריך לקבוע [אינטואיטיבית או לפי כללים שניתן למצוא בספרות] את מספר התאים.
- שלב שלישי – לרשום את ערכי התאים [בדוגמה – בטווח H2:H6]. ערך כל תא, בהתייחס לקיבוץ בתאים, הוא הגבול העליון המדומה של כל תא. נקבע את הגבול של התא הראשון H1 [18 באיור שלעיל] ואת הגבול של התא האחרון H6 [38 באיור שלעיל] הכלל הוא: $L1 \geq Min$ [אם נבחר $L1 < Min$ שכיחות התא תהיה 0] ו- $L2 \geq Max$ [אם נבחר $L2 < Max$ המערכת תיצור תא נוסף בשם MORE ובו ייצברו כל הנתונים שערכם גדול מאותו L2].
- שלב רביעי – להימצא בתא I2 ולסמן את טווח התאים I2:I6 ולהקליד את הנוסחה =FREQUENCY(B2:F1,H1:H6). כיוון שהנוסחה היא נוסחת **מערך**, בסיום ההקלדה יש ללחוץ בוזמנית על המקשים CONTROL+SHIFT+ENTER ומתקבלת התוצאות [השכיחות] כפי שמופיע בדוגמה.

לצד התא 23 השכיחות היא 15, המשמעות היא שמספר הנתונים בטווח המוגדר שערכם שווה או קטן ל- 23 אך גדול מ- 18 הוא 15 ובאופן דומה ביתר התאים.

התפלגות היפרגאומטרית

חישובי הסתברות, הנוגעים להתפלגות היפרגאומטרית, עשויים להיות מורכבים וכשהם נעשים באופן ידני או בעזרת מחשבוני הם מוגבלים [מעשית אך לא תיאורטית] לתחום מצומצם. נדגים את הדבר בעזרת התרגיל הבא:

תרגיל מספר 12

במשלוח של 25 צנצנות [N] המכילות [ע"פ הרשום על האריזה] כל אחת 200 טבליות של אומגה 3 [שמן דגים] 12 צנצנות [M] שבהן, בטעות, יש רק 175 טבליות. נלקח מדגם אקראי של 7 צנצנות [n]. מה ההסתברות שבמדגם לכל היותר 3 צנצנות [x] שבהן רק 175 טבליות?

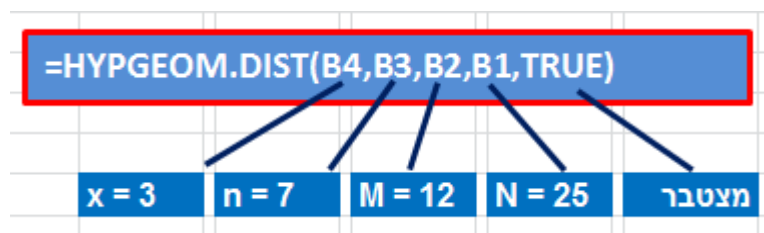
פתרון

הפתרון ה"רגיל" נראה כך:

$$P(x \leq 3) = \left[\frac{\binom{12}{0} \cdot \binom{25-12}{7-0}}{\binom{25}{7}} + \frac{\binom{12}{1} \cdot \binom{25-12}{7-1}}{\binom{25}{7}} + \frac{\binom{12}{2} \cdot \binom{25-12}{7-2}}{\binom{25}{7}} + \frac{\binom{12}{3} \cdot \binom{25-12}{7-3}}{\binom{25}{7}} \right]$$

באקסל ייראה הפתרון כך:

	A	B	C	D	E	F	G	H
1	N =	25						
2	M =	12						
3	n =	7						
4	x =	3	0.5503					
5								



תרגיל מספר 13

במשלוח של 250 צנצנות [N], המכילות [ע"פ הרשום על האריזה] כל אחת 200 טבליות של אומגה 3 [שמן דגים], 120 צנצנות [M] שבהן, בטעות, יש רק 175 טבליות. נלקח מדגם אקראי של 70 צנצנות [n]. מה ההסתברות שבמדגם לכל היותר 30 צנצנות [x] שבהן רק 175 טבליות?

פתרון

למעשה תרגיל זה דומה לתרגיל מספר 12 אולם כל אחד מהפרמטרים הוכפל פי 10. הפתרון ה"רגיל" כולל 31 אברים $[x+1]$ אותם יש לחשב [וזה לא פשוט כלל] והוא נראה כך:

$$P(x \leq 30) = \left[\frac{\binom{120}{0} \cdot \binom{250-120}{70-0}}{\binom{250}{70}} + \frac{\binom{120}{1} \cdot \binom{250-120}{70-1}}{\binom{250}{70}} + \dots + \frac{\binom{120}{30} \cdot \binom{250-120}{70-30}}{\binom{250}{70}} \right]$$

לאחר שהקלדנו, בגיליון ה- EXCEL, את הנתונים ואת הנוסחה של התרגיל הקודם די אם "נתלבש" על הפתרון ונחליף את ה- 25 ב- 250, את ה- 12 ב- 120, את ה- 7 ב- 70 ואת ה- 3 ב- 30 ונקבל את הפתרון לתרגיל:

	A	B	C	D	E	F	G	H
1	N =	250						
2	M =	120						
3	n =	70						
4	x =	30	0.1911					
5								

התפלגות פואסונית

התפלגות פואסונית, הינה התפלגות של משתנה בדיד המאפיינת הסתברות של מופעים [מקרים] למרווח זמן או ליחידה אחרת, למשל מספר המכוניות המגיעות במשך שעה למכון לשטיפת מכוניות, מספר דליפות מים ל- 10 ק"מ של צינור מים ראשי ועוד.

לצרכים מעשיים ובעיקר לבקרת איכות סטטיסטית, משמשת ההתפלגות הפואסונית כקירוב להתפלגות הבינומית [לחישוב ההסתברות למציאת x חלקים פגומים במדגם.

ההסתברות הפואסונית מחושב על פי:

$$P(X \geq x) = 1 - \sum_{x=0}^{x-1} \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad P(X \leq x) = \sum_{x=0}^x \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

חישובי הסתברות, הנוגעים להתפלגות פואסונית, עשויים להיות מורכבים וכשהם נעשים באופן ידני או בעזרת מחשבוניים הם מוגבלים [מעשית אך לא תיאורטית] לתחום מצומצם. נדגים את הדבר בעזרת הדוגמה הבאה.