

תרגיל: שביל החלב

מה נלמד בתרגיל הזה?

מה זה Pivot table ואיך עובדים איתו.

לפני התרגיל

צפו [בסרטון הזה](#).

מה זה Pivot Table ובשביל מה זה טוב

פתחו את הקובץ calc.xlsx בגיליון results. דבר ראשון, בואו ניזכר [בהסבר הזה](#) לגבי הנתונים שמופיעים בקובץ ובגיליון הספציפי הזה.

עכשיו, בואו נסתכל קצת על תוצאות הניסוי. סביר להניח שנרצה לשאול את עצמנו שאלות מהסוג הזה:

☞ מה היה אחוז הפריטים עם תשובות נכונות?

☞ איך המיקום של חציית עשרת (ביחידות / בעשרות) השפיע על אחוז ההצלחה?

☞ כמה טעויות עשה כל משתתף, ואיך כמות הטעויות הושפעה מאסטרטגיית החישוב (L2R/R2L)?

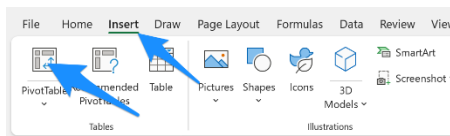
☞ מה סטיית התקן של זמני התגובה?

אפשר לענות על כל השאלות האלה בעזרת נוסחאות ופונקציות שלמדנו בתרגילים הקודמים. אבל היום נלמד דרך קלה יותר לענות עליהן – בעזרת Pivot Table. זה כלי של אקסל שעוזר לנו ביתוח בסיסי של הנתונים – לספור, לסכם, לחשוב ממוצעים וסטיות תקן, וגם להסתכל על הנתונים בכל מיני חתכים – כלומר לא על כל השורות ביחד אלא לחלק אותן לקבוצות לפי כל מיני פרמטרים.

בואו נראה איך זה עובד.

יצירת Pivot Table

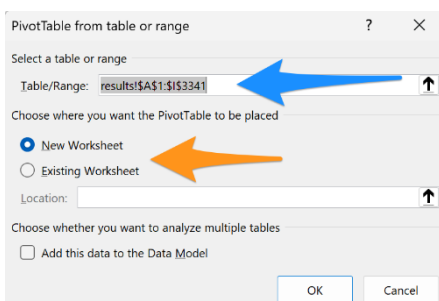
הדבר הראשון שאנחנו צריכים לעשות הוא ליצור את ה-Pivot Table. לשם כך, עמדו על מקום כלשהו בגיליון results (אבל לא על העמודות הריקות בצד ימין, אלא איפשהו באזור בו רשומים נתונים) ולחצו על כפתור Pivot Table, שנמצא בקצה של סרגל "הוספה":



בחלון שייפתח, נוכל להגדיר שני דברים:

א. איזה עמודות ושורות אנחנו הולכים לנתח (חץ כחול). ברירת המחדל היא כל הנתונים שרשומים בקובץ, אבל אם תרצו אפשר לבחור טווח נתונים מצומצם יותר.

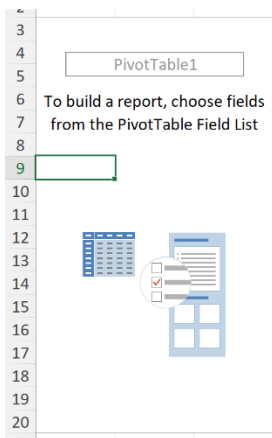
ב. איפה לשים את ה-Pivot Table החדש שניצור. אפשר לשים אותו בגיליון הנוכחי (כמו שריקדו הראה בסרטון), ואפשר ליצור אותו בגיליון חדש (זו ברירת המחדל). כרגע, הגיליון שלנו כבר די מלא אז נשים את הטבלה בגיליון חדש.



סיכום נתונים בסיסי

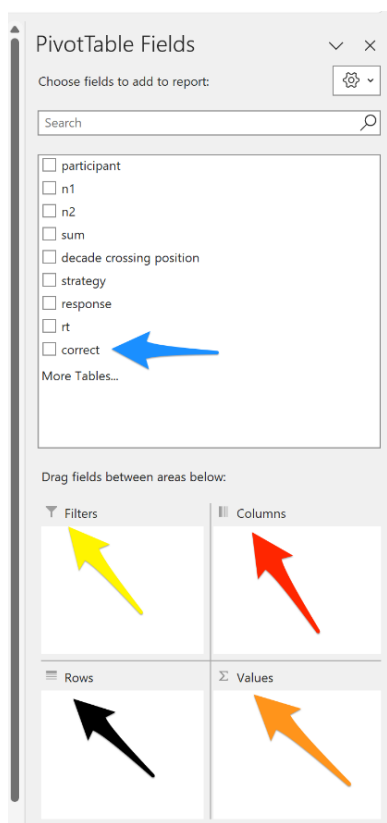
בתור התחלה, ננסה לענות על שאלה פשוטה: בכמה צעדים התשובה היתה נכונה?

✓ בטרמינולוגיה של ניסויים, "צעד" הוא פריט אחד בניסוי, במקרה זה – תרגיל אחד שהשתתף ענה עליו. בקובץ שלנו, כל שורה היא צעד.



1. העמידו את הסמן במקום כלשהו על איזור ה-Pivot Table –

כדי לדעת כמה תשובות נכונות היו, אנחנו יכולים פשוט לסכם את הערכים של כל התאים בעמודת correct. לשם כך, לכו אל חלון הגדרות ה-Pivot Table (תרשים 1), שנמצא בצד ימין של מסך האקסל. בחרו את עמודת correct (חץ כחול) וגררו אותה לאיזור של values (חץ כתום) – לכאן נגרור את העמודות שאנחנו רוצים להציג את ערכיהן (הממוצע שלהן, סכומן וכו').

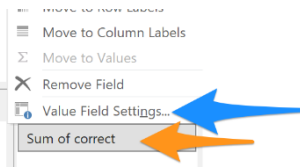


תרשים 1: חלון הגדרות ה-Pivot Table:

ברירת המחדל של Pivot Table היא לסכם את כל הערכים בעמודה שבחרנו, וזה מה שתקבלו:

2	
3	Sum of correct
4	2817

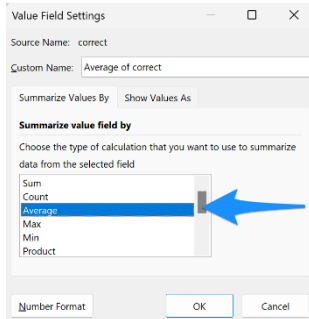
תוכלו לבדוק בקלות שהסיכום אכן נכון (למשל, חזרו לגיליון results, סמנו את עמודת correct כולה, ובדקו את הסכום שכתוב בתחתית המסך).



2. מה אחוז הצעדים בהם התשובה היתה נכונה?

בשביל לגלות את זה, נרצה לחשב לא את הסכום של העמודה אלא את הממוצע שלה. כדי לעשות זאת, נקליק על שם העמודה באיזור values (חץ כתום), ונבחר ב"גדרות" (חץ כחול).

שימו לב: לא מגדירים על העמודה ברשימת העמודות (חץ כחול בתרשים 1) אלא באיזור values (חץ כתום בתרשים 1).



בחלון שייפתח ("חלון הגדרות ערך"), נוכל להגדיר באיזו פונקציה להשתמש כדי "לסכם" את העמודה. ברירת המחדל כאמור היתה סיכום (הפונקציה SUM), עכשיו נוכל לבחור בממוצע:

מה יצא?

3. כרגע הממוצע מוצג בתור מספר בין 0 ל-1. זוכרותים איך לשנות להציג אותו בתור אחוזים?

4. שוב בעזרת "חלון הגדרות ערך", בדקו מה זמן התגובה המינימלי, המכסימלי, הממוצע ומה סטיית התקן של זמני התגובה.

לחלק את הנתונים לקבוצות

עד עכשיו, כל הדוגמאות שראינו הרגישו קצת פשוטות מדי – פשוט עשינו כל מיני סיכומים על עמודה שלמה, ואת זה אפשר לעשות גם בדרכים קלות יותר, בלי Pivot Table. אז בואו נראה בשביל מה הוא באמת מועיל לנו.

5. נחשב את אחוז הדיוק הממוצע בנפרד עבור צעדים עם חציית עשרת ביחידות או בעשרות.

ראשית, גררו את עמודת correct לאיזור values ובחרו להציג את הממוצע שלה, בדיוק כמו שעשינו בשאלה 2.

עכשיו נפצל את הנתונים האלה – כלומר נציג בנפרד את הממוצע של השורות בהן החציה היתה ביחידות (בעמודת decade crossing position כתוב U) או בעשרות (כתוב D).

בעצם אנחנו רוצים לקבל טבלת-סיכום קטנה עם 2 שורות – שורה של ממוצע היחידות ושורה של ממוצע העשרות. אז נגרור את עמודת decade crossing position לאיזור rows (חץ שחור בתרשים 1).

3	Row Labels	Average of correct
4	D	86.5%
5	U	82.1%
6	Grand Total	84.3%

זו טבלת הסיכום שנקבל:

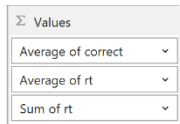
הטבלה מציגה את אחוז ההצלחה (הממוצע של עמודת correct) עבור חציה ביחידות, חציה בעשרות, וגם בסך הכל. כמו שסיפרתי באחד התרגילים הקודמים, תוכלו לראות ששיעור ההצלחה היה גבוה יותר כשחציית העשרת היתה בספרת העשרות, ונמוך יותר כשהיא היתה בספרת היחידות.

6. אם אנחנו רוצים לראות גם את זמן התגובה הממוצע – לא חייבים למחוק את אחוז הטעויות. אפשר להוסיף גם אותו לאיזור values, בלי להוציא משם את עמודת correct.

Row Labels	Average of correct	Average of rt
D	86.5%	11.36
U	82.1%	12.46
Grand Total	84.3%	11.90

כמו שאפשר לראות, כאן גם שיניתי את הפורמט של התאים בהם מוצג rt – לא באחוזים, כמו שיצא כברירת המחדל, אלא בתור מספר עם 2 מקומות אחרי הנקודה.

אנחנו יכולים לראות שכאשר חציית העשרת היתה בספרת העשרות, המשתתפים לא רק הצליחו יותר, אלא גם עבדו מהר יותר.



7. אפשר גם להציג את אותה עמודה פעמיים, עם שני שני סיכומים שונים. לדוגמה, גם את סטיית התקן של זמני התגובה. גררו את עמודת rt שוב לאיזור values, כך שהיא תופיע שם פעמיים:

ואז פתחו את ההגדרות של העמודה שהוספנו, ושנו אותן להציג סטיית תקן במקום ממוצע:

Row Labels	Average of correct	Average of rt	StdDev of rt
D	86.5%	11.36	4.47
U	82.1%	12.46	5.13
Grand Total	84.3%	11.90	4.84

8. כרגע "פצלנו" את הסיכום שלנו ל-2 שורות – סיכום נפרד עבור הצעדים עם חציה ביחידות ובעשרות. איך נפצל לפי אסטרטגיית החישוב?

9. בואו נפצל את הנתונים גם לפי מיקום חציית העשרת וגם לפי אסטרטגיית החישוב. הוסיפו בחזרה את עמודת decade crossing position לאיזור rows. שימו אותה מתחת ל-strategy, כך:



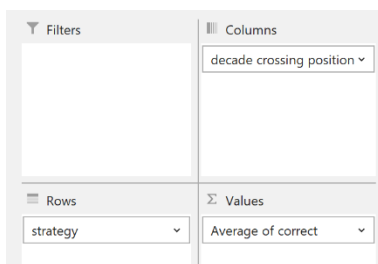
טבלת הסיכום תציג את הנתונים מפוצלים ב-2 רמות: לפי אסטרטגיה, ובתוך כל אסטרטגיה – לפי מיקום חציית העשרת. בשורה הראשונה של כל אסטרטגיה אנחנו גם רואים את סיכום הנתונים של אותה אסטרטגיה (עבור 2 מיקומי חציית העשרת):

	Row Labels	Average of correct	Average of rt	StdDev of rt2
3				
4	L2R	86.9%	11.97	4.87
5	D	89.2%	11.26	4.01
6	U	84.7%	12.70	5.52
7	R2L	81.7%	11.83	4.80
8	D	83.9%	11.47	4.89
9	U	79.5%	12.22	4.69
10	Grand Total	84.3%	11.90	4.84

10. נניח שאנחנו רוצים את הסיכום הפוך – כלומר שאקסל יחלק את הנתונים קודם כל לפי מיקום החציה, ורק בתוך כל מיקום חציה יחלק לפי אסטרטגיה. יש לך ניחוש איך לעשות את זה?

11. עוד דרך להציג את החלוקה לפי 2 פרמטרים (אסטרטגיה, מיקום חציה) היא כשאחד מהם בשורות ואחד בעמודות, כך:

	Row Labels	Average of correct	Column Labels		
4			D	U	Grand Total
5	L2R	89.2%	84.7%	86.9%	
6	R2L	83.9%	79.5%	81.7%	
7	Grand Total	86.5%	82.1%	84.3%	

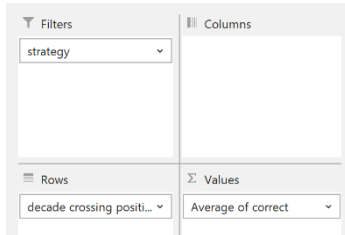


כמובן, נוח לראות את הנתונים בצורה הזאת אם שמנו באיזור values רק עמודה אחת (כמו בדוגמה הזאת). אז בואו נוציא משם את rt. ואז העבירו את עמודת decade crossing position מאיזור rows אל איזור columns (חץ אדום בתרשים 1). כך:

12. "סובבו" את הטבלה: מיקום חציית העשרת בשורות, אסטרטגיה בעמודות.

סינון נתונים

13. מחקו את strategy מאיזור columns. נשארו עם טבלת סיכום מחולקת רק לפי מיקום חציית העשרת: אחוז ההצלחה בצעדים עם חציית עשרת ביחידות / בעשרות.



עכשיו אנחנו רוצים לסנן את הנתונים. כלומר, להשאיר את טבלת הסיכום בדיוק במבנה הנוכחי שלה, אבל שיוצגו בה רק הנתונים של הצעדים בהם המשתתפים חישבו משמאל לימין. כדי לעשות זאת, קודם כל נגרור את עמודת strategy לאיזור Filters (חץ צהוב בתרשים 1). כך:

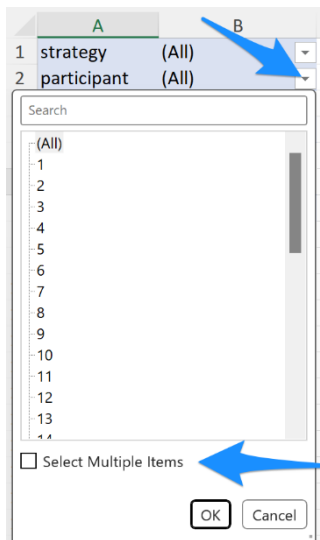
כתוצאה מכך, אנחנו רואים שנוסף סימון strategy מעל טבלת הסיכום:

A	B
strategy	(All)
2	
3	Row Labels - Average of correct
4	D 86.5%
5	U 82.1%
6	Grand Total 84.3%

בתא B1 נוכל לבחור להציג רק את הנתונים של אחת האסטרטגיות. נסו זאת.

14. הוסיפו גם סינון לפי משתתף, ובדקו נתונים של חלק מהמשתתפים.

15. אפשר לבחור, כמו שעשינו קודם, להציג רק את הנתונים של משתתף אחד. אבל אפשר גם לבקש להציג נתונים של כמה משתתפים:



4	Row Labels - Average of correct	
5	D	86.5%
6	U	82.1%
7	Grand Total	84.3%

16. גם אם לא גררנו עמודה לאיזור Filters אלא השתמשנו בה בשורות או בעמודות, עדיין אפשר לסנן לפיה. לחצו כאן:

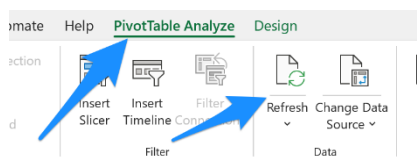
שינוי נתונים

נחזור לגיליון results. נוסיף שוב עמודה שכבר ראינו בתרגיל קודם ("אל תשאל למה") – עמודה שאומרת האם חציית העשרת היתה בשלב הראשון או השני של החישוב.

הוסיפו עמודה בין עמודה F לעמודה G, תנו לה כותרת, ורשמו בה את הנוסחה:

$$=IF(OR(AND(E2="D", F2="L2R"), AND(E2="U", F2="R2L")), 1, 2)$$

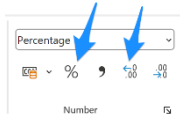
אם נחזור אל גיליון ה-pivot table, נראה שהעמודה החדשה לא נוספה שם. בשביל שהיא תתווסף, צריך להגיד לאקסל בפירוש שעשינו שינוי בנתונים. בסרגל "PivotTable analyze", נלחץ על כפתור רענון נתונים:



תשובות לשאלות בקובץ זה

2. הממוצע הוא 84.3%

3	Average of correct	84.3%
4		0.843413174



3. עומדים על התא של הממוצע ולוחצים על כפתור פורמט אחוז בסרגל "בית", ואם אנחנו רוצים – גם נגדיל קצת את רמת הדיוק:

התוצאה

3	Average of correct	84.3%
4		

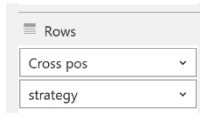
4. בשביל זה נצטרך לגרור את עמודת rt לאיזור values, ולהגדיר שאנחנו רוצים לסכם אותה בעזרת פונקציות AVERAGE, MAX, MIN או StDev. אם קודם שינינו את הפורמט של התא לאחוזים, כנראה נרצה להחזיר אותו עכשיו לפורמט של מספר רגיל.

- זמן תגובה מינימלי: 4 שניות
- מכסימלי: 61 שניות
- ממוצע: 11.9 שניות
- סטיית תקן: 4.83 שניות

8. נוציא את decade crossing position מאיזור rows, ובמקומו נשים את strategy. התוצאה:

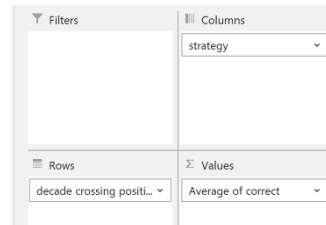
Row Labels	Average of correct	Average of rt	StdDev of rt
L2R	86.9%	11.97	4.87
R2L	81.7%	11.83	4.80
Grand Total	84.3%	11.90	4.84

כלומר המשתתפים הצליחו יותר בחישוב משמאל לימין מאשר בחישוב מימין לשמאל -- אבל הם גם היו טיפ טיפה איטיים יותר. כמובן שכדי לעמוד מאחורי אמירה כזאת, לא מספיק הנתונים בטבלה כאן. צריך גם לעשות ניתוח סטטיסטי כדי לראות אם ההבדל מובהק!

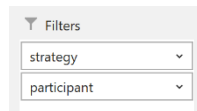


10. באיזור rows, נחליף את הסדר של 2 העמודות (גוררים את אחת מהן) שיהיה כך:

Average of correct	Column Labels		
Row Labels	L2R	R2L	Grand Total
D	89.2%	83.9%	86.5%
U	84.7%	79.5%	82.1%
Grand Total	86.9%	81.7%	84.3%



12. ככה:



14. גוררים גם את עמודת משתף לאיזור הפילטרים:

15.