

שאלה 1

בחינת ה-SAT היא המקבילה האמריקנית של הבחינה הפסיכומטרית בישראל. לבחינה יש שני חלקים, מילולי ומתמטי, ולכל חלק ניתן ציון נפרד.

בתיקיית הנתונים בדרייב של הקורס נמצא קובץ אקסל המכיל נתונים על 50 המדינות של ארצות הברית, וכן עבור מחוז קולומביה (Washington DC). כן תוכלו למצוא שם הסבר על מבנה הקובץ ומה מייצג כל משתנה.

בשאלה זו נבנה בעזרת אקסל מודל שיסביר את ציוני החלק המתמטי של המבחן (המשתנה satMath), וזאת על יד השכר הממוצע של המורים (המשתנה teacherPay) ואחוז הניגשים לבחינה (המשתנה percentTaking).

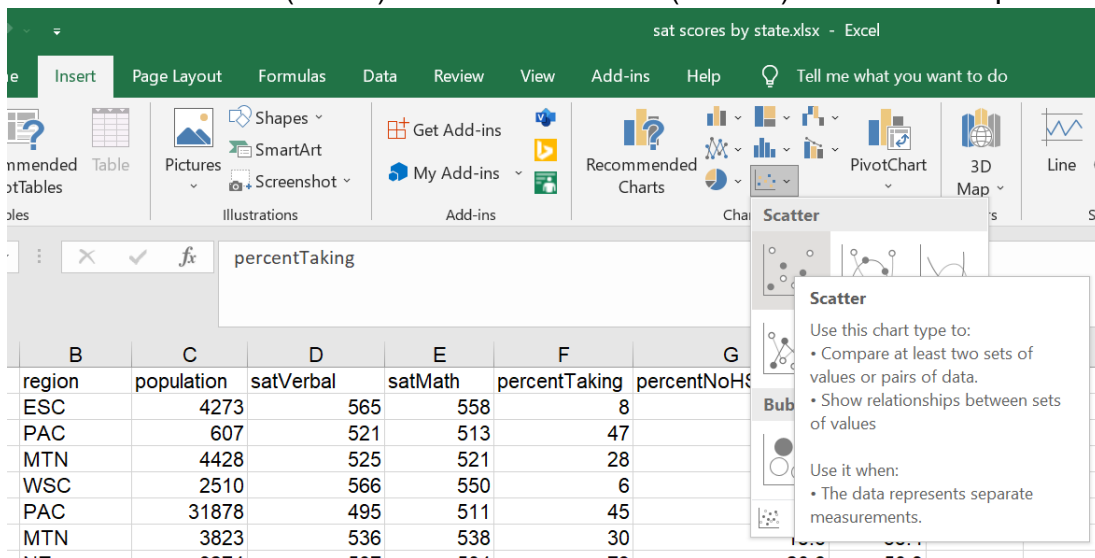
יש לענות על השאלות בעזרת אקסל. לצורך כך יש להתקין את התוסף Analysis ToolPak כפי שהוסבר בשיעור.

א. צרו שתי דיאגרמות פיזור: אחת מהן תציג את הציון במתמטיקה מול שכר המורים, והשנייה תציג את הציון במתמטיקה מול אחוז הניגשים לבחינה.

הדרכה/דוגמה: כדי ליצור דיאגרמת פיזור של גודל האוכלוסייה מול אחוז הניגשים לבחינה, יש לבחור בתחילה את שתי העמודות של המשתנים:

	A	B	C	D	E	F	G	H
1	state	region	population	satVerbal	satMath	percentTaking	percentNoHS	teacherPay
2	AL	ESC	4273	565	558	8	33.1	31.3
3	AK	PAC	607	521	513	47	13.4	49.6
4	AZ	MTN	4428	525	521	28	21.3	32.5
5	AR	WSC	2510	566	550	6	33.7	29.3
6	CA	PAC	31878	495	511	45	23.8	43.1
7	CO	MTN	3823	536	538	30	15.6	35.4
8	CT	NE	3274	507	504	79	20.8	50.3
9	DE	SA	725	508	495	66	22.5	40.5
10	DC	SA	543	489	473	50	26.9	43.7
11	FL	SA	14400	498	496	48	25.6	33.3
12	GA	SA	7353	484	477	63	29.1	34.1
13	HI	PAC	1184	485	510	54	19.9	35.8
14	ID	MTN	1189	543	536	15	20.3	30.9
15	IL	ESC	11847	564	575	14	23.8	40.9

לאחר מכן יש לבחור הוספה (או insert) ולבחור בדיאגרמת פיזור (scatter):



- על סמך התבוננות בדיאגרמות, האם לדעתם ניתן להניח כי קיים קשר לינארי בין שכר המורים והציון במתמטיקה, וכן בין אחוז הניגשים לבחינה ובין הציון במתמטיקה?
- ב. צרו מודל רגרסיה בו המשתנה המוסבר הוא הציון במתמטיקה, והמשתנה המסביר הוא שכר המורים.
- 1) האם ההנחה כי התפלגות השאריות היא התפלגות נורמלית היא הנחה סבירה?
 - 2) מהי משוואת הרגרסיה?
 - 3) מהו אחוז השונות המוסברת?
 - 4) הסבירו כיצד על פי המודל שכר המורים משפיע על הציון במתמטיקה.
 - 5) אם במדינה מסויימת השכר הממוצע של המורים הוא 60 אלף דולר בשנה, מה צפוי להיות הציון במתמטיקה על פי המודל?
 - 6) חוו את דעתכם על טיב המודל
- ג. צרו מודל רגרסיה בו המשתנה המוסבר הוא הציון במתמטיקה, והמשתנה המסביר הוא אחוז הניגשים לבחינה
- 1) מהי משוואת הרגרסיה?
 - 2) מהו אחוז השונות המוסברת?
 - 3) מהו ערך השארית עבור מדינת אילינוי (קוד המדינה הוא IL)
 - 4) האם דוחים את ההשערות כי מקדמי הרגרסיה (a ו-b) שונים מאפס?
 - 5) הסבירו כיצד על פי המודל אחוז הניגשים לבחינה משפיע על הציון במתמטיקה.
 - 6) חוו את דעתכם על טיב המודל
- ד. צרו מודל רגרסיה בו המשתנה המוסבר הוא הציון במתמטיקה, הפעם עם שני משתנים מסבירים: שכר המורים ואחוז הניגשים לבחינה.
- 1) מהי משוואת הרגרסיה?
 - 2) מהו אחוז השונות המוסברת?
 - 3) מהו האמדן לסטיית התקן של השאריות?
 - 4) האם דוחים את ההשערות כי מקדמי הרגרסיה שונים מאפס?
 - 5) חוו את דעתכם על טיב המודל.
- ה. מהו לדעתכם הסבר אפשרי לשוני בין מקדם הרגרסיה של שכר המורים במודל הראשון ובין המקדם המקביל במודל השלישי?

שאלה 2

כדי לבדוק את הקשר בין המשתנה $time$: הזמן שסטודנט מתכוון למבחן (בשעות) ובין המשתנה $score$: הציון במבחן, חוקר אסף נתונים ובנה מודל רגרסיה. קו הרגרסיה הוא

$$score = 31.2 + 6.7 \cdot time$$

- א. מהו ניבוי קו הרגרסיה עבור סטודנט שהתכוון למבחן במשך 7 שעות?
- ב. סטודנט התכוון למבחן במשך 5 שעות ולבסוף קיבל ציון 70. מהי השגיאה/שארית של תצפית זו?
- ג. סטודנט תכנן ללמוד מספר מסויים של שעות, אז בסופו של דבר למד שעתיים פחות ממה שתכנן. מה יהיה השינוי בציון הצפוי לו על פי המודל?
- ד. לאחר סיום הניתוח הסטטיסטי, החוקר השיג נתון על סטודנט נוסף שהתכוון למבחן במשך 8 שעות. מהו הציון שקיבל סטודנט זה במבחן?
- ה. מה יהיה קו הרגרסיה אם נמדוד את הזמן בדקות במקום בשעות?

שאלה 3

כדי לבדוק את הקשר בין איכות טעמו של יין ובין מרכיביו הכימיים נלקח מדגם מייצג של בקבוקי יין אדום. מכל בקבוק נלקחה דגימה לבדיקת מעבדה. כמו כן, צוות של עשרה מומחים טעם את כל היינות ודירג את איכותם. איכות היין היא הממוצע של הדירוג של עשרת המומחים.

לצורך תרגיל זה נתמקד בקשר בין איכות היין ובין רמת האלכוהול שבו, באחוזים. כמו כן, נתעלם מהשאלה האם ההנחות הדרושות מתקיימות.
 כדי לבדוק את הקשר נבנה מודל רגרסיה. מצורף פלט תוצאת הרגרסיה מניתוח באקסל. למרבה הצער, המרצה של הקורס מחק ממנו כמה נתונים... עם זאת, נתון כי ממוצע דירוגי היינות הוא 5.61 וכי ממוצע רמת האלכוהול של היינות שבמדגם הוא 10.16.

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R						
R Square						
Adjusted R Square	0.189					
Standard Error	0.764					
Observations						
ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1		81.84			
Residual	596	347.79	0.58			
Total	597	429.63				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	1.916	0.31			1.30	2.53
alcohol		0.03			0.30	0.42

- א. מהו המשתנה המסביר ומהו המשתנה המוסבר?
- ב. מה היה גודל המדגם?
- ג. מהו האמדן למקדם הרגרסיה b?
- ד. האם תדחו את ההשערות כי מקדמי הרגרסיה a ו-b שונים מאפס?
- ה. מהו אחוז השונות המוסברת על ידי הרגרסיה?
- ו. מהו מקדם המתאם בין איכות היין ורמת האלכוהול?
- ז. הסבירו את ההשפעה של שינוי במשתנה המוסבר על הניבוי של ערך המשתנה המסביר.