



אוניברסיטת בן-גוריון
המחלקה לסוציולוגיה ולאנתרופולוגיה

ריכוז שיעורי למידה עצמית בקורס:
תוכנות סטטיסטיות

מרצה: ד"ר אורי ביבי

תשע"ג

מעודכן לתוכנות:

Excel, Office 2007
SPSS 18

תוכן :

מספר עמוד

3	שיעור מספר 1 :
16	תרגיל הגשה מספר 1 :
18	שיעור מספר 2 :
26	תרגיל הגשה מספר 2 :
27	שיעור מספר 3 :
40	תרגיל הגשה מספר 3 :
41	שיעור מספר 4 :
54	תרגיל הגשה מספר 4 :
55	שיעור מספר 5 :
67	תרגיל הגשה מספר 5 :
68	שיעור מספר 6 :
80	תרגיל הגשה מספר 6 :
81	שיעור מספר 7 :
86	תרגיל הגשה מספר 7 :

שיעור מספר 1

שיעור מס' 1: Excel; הזנה וארגון נתונים, חישובים מתמטיים והצגה גראפית

בשיעור זה תלמדו כיצד לעבוד עם תוכנת Excel לצורך ניתוח של נתונים שהתקבלו במחקר. בדרך כלל חוקרים עובדים עם תוכנות סטטיסטיות (כגון: Statistica ו-SPSS) לצורך ניתוח נתונים ממחקריהם, מאחר ואלו תוכנות שמותאמות יותר לניתוחים סטטיסטיים ומאפשרות ניתוחים סטטיסטיים מורכבים. אולם, בחרנו להקדיש את השיעור הראשון לתוכנת Excel. זוהי תוכנה שנמצאת כמעט בכל בית, ומאפשרת מגוון של אפשרויות לניתוחים סטטיסטיים בסיסיים. כמו-כן, באמצעות Excel ניתן לבנות ולהציג תרשימים וגרפים בקלות.

Excel הוא גיליון אלקטרוני. גיליון אלקטרוני הוא למעשה טבלה, המורכבת ממספר גדול של שורות ושל עמודות (באופן מדויק יותר: ב-Excel ניתן לפתוח במקביל מספר טבלאות, שכל אחת יכולה להכיל עד 65,536 שורות ו-256 עמודות – גודל שיספיק כמעט לכל הצרכים שלכם בעתיד). כל משבצת בגיליון נקראת "תא". לכל תא יש "כתובת", המורכבת מאות ומספר. האות מציינת את העמודה, והמספר – את השורה. כך, התא A1 נמצא בעמודה הראשונה ובשורה הראשונה; התא C13 נמצא בעמודה השלישית ובשורה ה-13. לתוך התאים בגיליון ניתן להזין מספרים, מילים, תאריכים או כל סוג אחר של נתונים לפי רצונכם. בנוסף, וזהו היתרון החשוב של Excel, ניתן להזין לתוך תאים אלה גם פונקציות מתמטיות וסטטיסטיות, שיעשו שימוש בנתונים הנמצאים בתאים אחרים בגיליון. למשל, נניח שהזנתם לתוך התאים בעמודה B את הטמפרטורות שנמדדו במשך חודש אוגוסט בבאר-שבע: הטמפרטורה ב-1 באוגוסט תירשם בתא B1, הטמפרטורה ב-2 באוגוסט תירשם בתא B2, וכך הלאה עד הטמפרטורה ב-31 באוגוסט, שתירשם בתא B31. כעת, תוכלו להגדיר שתאים אחרים בגיליון יכילו חישובים שונים על נתונים אלה (למשל: סכום, ממוצע, חציון, מספר התצפיות, ועוד). למשל, תוכלו להגדיר שהתא D3 (לדוגמה) יכיל את ממוצע התאים A1 עד A31, שהתא E5 יכיל את הסכום שלהם, וכן הלאה.

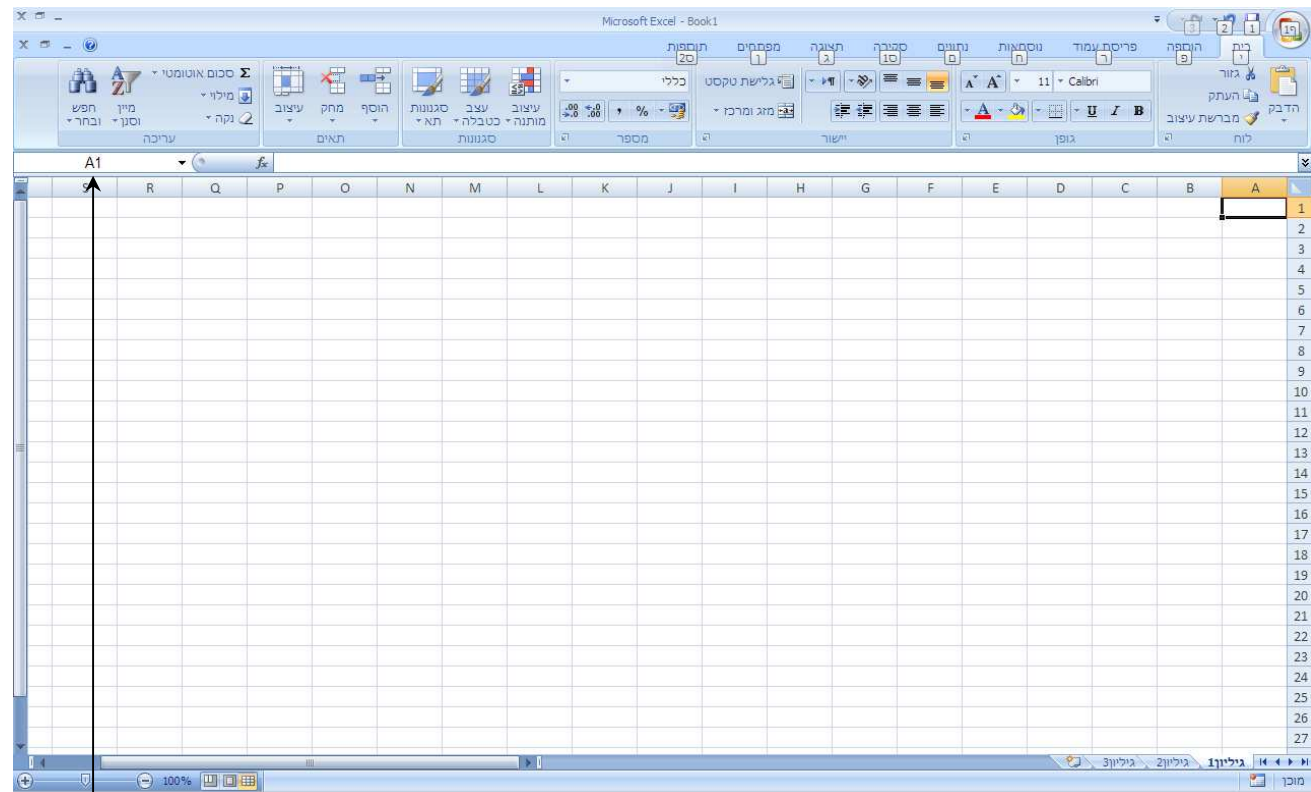
בשיעור זה תלמדו את השימוש הבסיסי בתוכנת Excel. במהלך השיעור תכירו את סביבת העבודה והתפריטים העיקריים בתוכנה, תלמדו להזין נתונים לגיליונות האלקטרוניים, ותכירו חלק מהפונקציות המתמטיות והסטטיסטיות השימושיות ביותר (כגון ממוצע, שונות, חציון, סכום, וכו'). כמו-כן, תלמדו ליצור תרשימים מהנתונים. זה המקום לציין כי Excel היא תוכנה בעלת מגוון עצום של אפשרויות. היא כוללת יותר מ-200 פונקציות מתמטיות, סטטיסטיות וכלכליות, ומאפשרת הוספה של פונקציות נוספות באמצעות תיכנות. בשיעור זה תכירו את תוכנת Excel על קצה המזלג. אולם, באמצעות הכלים שתקבלו, יחד עם תפריטי העזרה (Help) הנוחים של התוכנה, תוכלו להמשיך ולהרחיב את ידיעותיכם ב-Excel באופן עצמאי. נתחיל בהיכרות בסיסית עם סביבת העבודה של Excel.

את התוכנה ניתן לפתוח דרך התפריט הראשי של Windows : יש ללחוץ על "התחל" (או – "Start"), ואז על "תוכניות" (או – "Programs"). מתפריט התוכניות יש לבחור את התוכנית



Microsoft Excel. האייקון (צלמית) של התוכנה נראה כך :

לפניכם ייפתח המסך הבא :



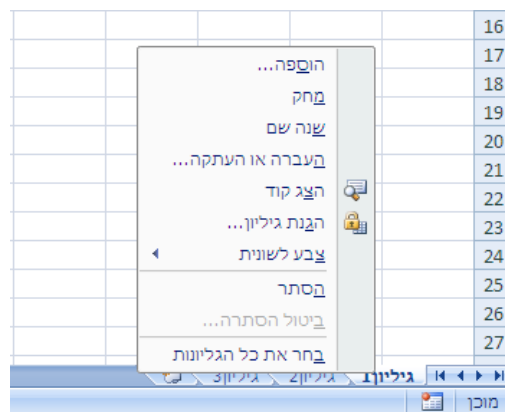
עיקר שטחו של המסך הוא טבלת הנתונים. זהו גיליון העבודה, שאליו תזינו את הנתונים ובו תבצעו חישובים שונים באמצעות נתונים אלה. שימו לב שאזור זה מחולק לטורים ולשורות.

אם תסמנו בעזרת הסמן את אחד התאים, תראו שהעמודה והשורה שבהן נמצא תא זה מודגשות.

כמו-כן, תראו את הכתובת של התא שסימנתם בחלונית שנמצאת בחלקו השמאלי של המסך.

בחלקו התחתון של המסך, תראו את התוויות "גיליון 1", "גיליון 2", ו-"גיליון 3".

למעשה, עם פתיחת התוכנה נפתחים שלושה גיליונות (כלומר, שלוש טבלאות שונות) במקביל. שם הגיליון שבו אתם עובדים כרגע מודגש בלבן. עם פתיחת התוכנה, זה יהיה תמיד "גיליון1". ניתן להוסיף עוד גיליונות, למחוק גיליונות או לשנות את שמותיהם. כדי לעשות זאת, יש לעמוד עם הסמן על אחת התוויות ולהקיש על המקש הימני של העכבר. ייפתח תפריט, שממנו תוכלו לבחור את האפשרות "הוספה..." (ואז – לבחור באפשרות "גיליון עבודה") כדי להוסיף גיליונות נוספים. כמו כן, תוכלו לבחור באפשרות "מחק" כדי למחוק את הגיליון, או באפשרות "שנה שם" כדי לשנות את שמו (הערה: כאשר עובדים עם מספר גיליונות, כדאי מאוד לשנות את שמותיהם לשמות בעלי משמעות, כגון "ציונים", "טמפרטורות", "כיתה ג" וכו'. שמות כאלה יקלו עליכם לדעת מה נמצא בכל אחד מהגיליונות).



מעל לגיליון העבודה תוכלו לראות צלמיות רבות, שהתפקיד של חלקן יוסבר בהמשך. מעליהן, יימצא תפריט התוכנה. תפריט זה מוכר לכם בודאי מתוכנות רבות אחרות, כגון Word ו-Internet Explorer. בראש חלון התוכנה מופיעים לחצנים בודדים שמאפשרים פעולות בסיסיות:



תת-התפריט הראשון בשורה (ובמידה רבה, החשוב ביותר), נקרא "בית".



הוא כולל פעולות כגון העתקה, הדבקה וגזירה של נתונים. פעולות אלה זהות לפעולות המקבילות בתוכנת Word.

עיצוב תוכן התאים (גופן)

תפריט זה כולל גם אפשרויות הוספה / מחיקת שורות / עמודות

וגם אפשרות סכימה של מספר תאים (התוצאה תופיע בתא אחר שייבחר על ידי המשתמש

לחיצה תת התפריטים הבאים: הוספה / פריסת עמוד / נוסחאות / נתונים / סקירה / תצוגה / מפתחים ותוספות.... תוביל לשינוי הצלמיות ולפתיחת אפשרויות רבות שאת חלקן נלמד במהלך השיעורים הקרובים.

את ההיכרות עם Excel נתחיל באמצעות הדוגמה הבאה

תרגיל 1

לפניכם נתונים שנאספו במסגרת מחקר על הרגלי האכילה של תושבי קיבוץ גלבוה. נדגמו 9 אנשים, שנשאלו למינם, גילם, מספר הארוחות השבועיות שלהם ומספר הסועדים עמם בארוחות לפי ימי השבוע.

נבדק	מין	גיל	יום א'	יום ב'	יום ג'	יום ד'	יום ה'	יום ו'	יום ש'	מס' ארוחות בשבוע
1	זכר	26	3	4	5	2	6	6	1	15
2	זכר	23	6	5	1	4	5	5	0	22
3	נקבה	45	1	0	3	3	7	8	8	19
4	זכר	65	7	3	5	5	4	2	6	30
5	נקבה	17	0	6	6	4	1	4	2	12
6	נקבה	29	3	7	4	3	2	1	3	10
7	זכר	42	2	4	3	3	1	6	2	42
8	זכר	16	7	9	1	6	3	5	5	12
9	נקבה	71	4	3	7	4	5	3	2	6

הוטלה עליכם המשימה לבנות קובץ נתונים, שיכיל את תוצאות המחקר, לשם ניתוח נתוניו בשלב מאוחר יותר.

1. פתחו קובץ נתונים חדש וקראו לו בשם "food".
 2. הוסיפו גיליון נתונים חדש, וקראו לו בשם "קיבוץ גלבוה".
 3. הזינו את הנתונים לגיליון החדש. הקפידו לשמור את הנתונים מדי פעם.
- *** שימו לב: את שמות המשתנים (כלומר, "נבדק", "מין", "גיל" וכן הלאה) יש לכתוב בתוך גיליון הנתונים, באותו האופן שבו מוזינים את הנתונים עצמם. ב-Excel אין אפשרות לשנות את שמות העמודות לשמות המשתנים. בתוכנות הסטטיסטיות שתלמדו בהמשך (SPSS ו-Statistica) אפשרות זו תהיה קיימת.

גיליון הנתונים שהזנתם יראה כך:

נבדק	מין	גיל	יום א'	יום ב'	יום ג'	יום ד'	יום ה'	יום ו'	יום ש'	מס' ארוחות בשבוע
1	זכר	26	3	4	5	2	6	6	1	15
2	זכר	23	6	5	1	4	5	5	0	22
3	נקבה	45	1	0	3	3	7	8	8	19
4	זכר	65	7	3	5	5	4	2	6	30
5	נקבה	17	0	6	6	4	1	4	2	12
6	נקבה	29	3	7	4	3	2	1	3	10
7	זכר	42	2	4	3	3	1	6	2	42
8	זכר	16	7	9	1	6	3	5	5	12
9	נקבה	71	4	3	7	4	5	3	2	6

כדי למחוק עמודה שלמה, יש להעמיד את הסמן על שם העמודה (האות). הסמן יהפוך לחץ, ואז יש ללחוץ על מקש העכבר הימני. בחירה באפשרות "מחק" תעלים את כל העמודה. בחירה באפשרות "נקה תוכן" תשאיר את העמודה, ורק התוכן שלה יימחק. נסו את כל אחת מהאפשרויות על גיליון הנתונים שלכם. מחקו את העמודה "יום ג'", ונקו את התוכן של העמודה "יום ד".

כדי למחוק שורה שלמה, יש להעמיד את הסמן על שם השורה (המספר). הסמן יהפוך לחץ, ואז יש ללחוץ על מקש העכבר הימני. בחירה באפשרות "מחק" תעלים את כל השורה. בחירה באפשרות "נקה תוכן" תשאיר את השורה, ורק התוכן שלה יימחק. נסו את כל אחת מהאפשרויות על גיליון הנתונים שלכם. מחקו את השורה שבה נמצאים נתוני נבדק מספר 5, ונקו את התוכן של השורה שבה נמצאים נתוני נבדק מספר 6.

חישוב ערכי משתנים חדשים:

נניח שברצוננו לחשב את מספר הארוחות הממוצע שאכל כל חבר קיבוץ ביום. מספר זה יהיה למעשה מספר הארוחות בשבוע, לחלק ל-7 ימים. כדי לעשות זאת, נכתוב את שם המשתנה "מס' ארוחות ממוצע ביום" בתא L1. לאחר מכן, נרצה לחשב את הערך של התא הרלוונטי עבור כל אחד מהנבדקים. לשם כך, נשתמש בערך המתאים בטור "מס' ארוחות בשבוע" ונחלק אותו ב-7.

סמנו באמצעות הסמן את התא L2. כעת כתבו את הסימן שווה (=). סימן זה מציין כי כעת אתם עומדים להזין נוסחה או פונקציה, ולא נתונים. לאחר מכן, סמנו את התא K2, ואז כתבו "/7" והקישו ENTER:

מס' ארוחות ממוצע ביום	מס' ארוחות בשבוע	יום ש'	יום ו'	יום ה'	יום ד'	יום ג'	יום ב'	יום א'	גיל	מין	נבדק
=K2/7	15	1	6	6	2	5	4	3	26	זכר	1
	22	0	5	5	4	1	5	6	23	זכר	2
	19	8	8	7	3	3	0	1	45	נקבה	3
	30	6	2	4	5	5	3	7	65	זכר	4
	12	2	4	1	4	6	6	0	17	נקבה	5
	10	3	1	2	3	4	7	3	29	נקבה	6
	42	2	6	1	3	3	4	2	42	זכר	7
	12	5	5	3	6	1	9	7	16	זכר	8
	6	2	3	5	4	7	3	4	71	נקבה	9

בתא K2 יירשם הערך של החישוב, כלומר 2.142857 בערך. כעת, נרצה לבצע פעולה דומה עבור כל הנבדקים. דרך אחת לעשות זאת, היא לחזור ולכתוב את הנוסחה שכתבתם בכל אחד מהתאים L3 עד L9. זוהי כמובן דרך מסורבלת (חשבו מה היה אם היה צריך לכתוב את הנוסחה 100 פעמים, או 1000 פעמים...). דרך אחרת לעשות זאת היא להשתמש בגרירה. גרירה היא העברה של הפונקציה לתאים חדשים. כדי לעשות זאת, סמנו באמצעות הסמן את תא L2. בצד השמאלי התחתון של התא מופיע ריבוע קטן. עמדו עליו עם העכבר, לחצו על המקש השמאלי של העכבר, וגררו בעזרת העכבר את תוכן התא כלפי מטה. כאשר תעובו את מקש העכבר, יופיעו שאר הממוצעים בכל התאים שאליהם גררתם את הנוסחה.

כעת, באופן דומה, צרו משתנה חדש (בעמודה M), שיכיל את מספר הארוחות שאכל כל נבדק בסוף השבוע (כלומר, סכום הארוחות שאכל ביום שישי ובשבת). נסו לבצע מטלה זו בעצמכם, לפני שתציצו ברמז שמתחת.

רמז :

כדי לעשות זאת, יש לעמוד עם הסמן על התא הרצוי (M2). לאחר מכן, יש לכתוב את הסימן שווה (=), כדי לציין שאתם עומדים לכתוב נוסחה. אז, סמנו את התא המכיל את מספר הארוחות ביום שישי (H2), כתבו את הסימן "+" , ואז סמנו את התא המכיל את מספר הארוחות בשבת (I2). לבסוף, הקישו ENTER. הערך שיתקבל בתא L2 הוא 7 (מפני שנבדק מספר 1 אכל 6 ארוחות ביום שישי וארוחה אחת בשבת). כעת, גררו את הנוסחה שבתא L2 כלפי מטה, כדי שתחושב עבור כל אחד מהנבדקים האחרים.

שימוש בפונקציות מתמטיות וסטטיסטיות :

בסעיף הקודם יצרתם, למעשה, פונקציות פשוטות. פונקציה היא פעולה המתבצעת על ערך של תא מסויים, או קבוצת תאים, כדי להפיק מהם ערך אחד אחר. דוגמאות לפונקציות כאלה הן הוצאת שורש, חישוב ממוצע, וחישוב סכום. כל אחת מהדוגמאות עושה שימוש בערכי תאים אחרים, כדי להפיק מהם ערך חדש, בהתאם לפונקציה הספציפית.

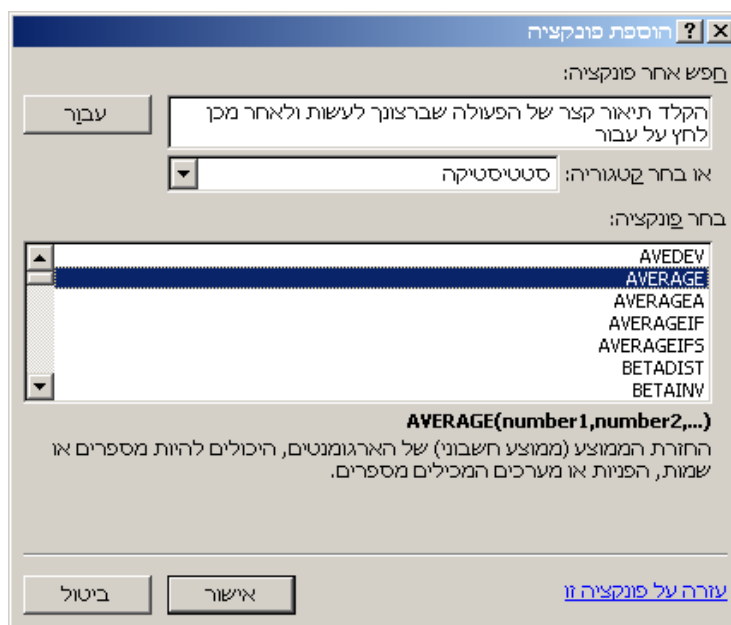
נמשיך את הדוגמה: נניח שרוצים לדעת מהו הגיל הממוצע של הנבדקים במדגם. עמדו עם הסמן על תא C12.

ביחרו בתת תפריט "נוסחאות" ← "הוספת פונקציה"

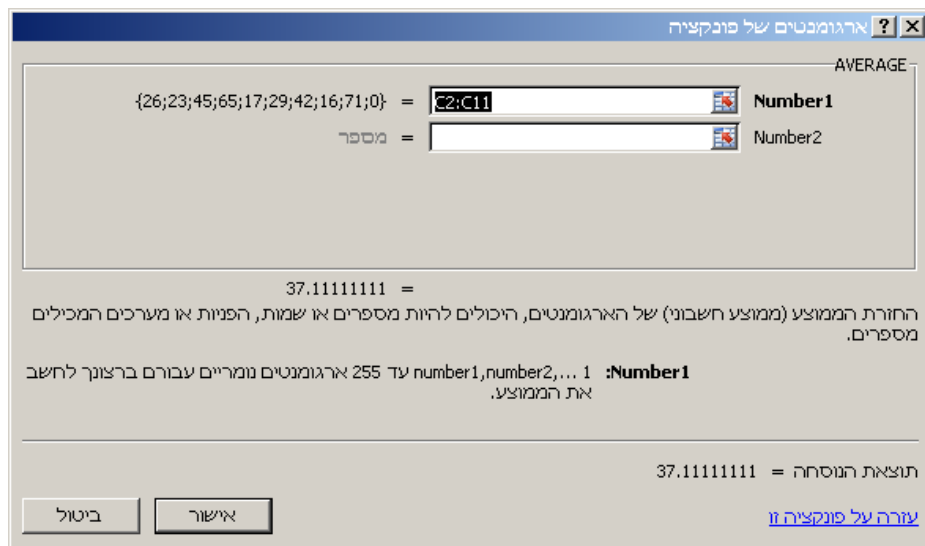


בלחיצה על צלמית זו ייפתח חלון שמכיל את כל הפונקציות הקיימות באופן מובנה ב-Excel, מחולקות לקטגוריות. בחרו בקטגוריה "סטטיסטיקה", ואז חפשו

את הפונקציה "AVERAGE" (ממוצע). סמנו אותה, ולחצו על "אישור".



ייפתח החלון הבא :



בשורה הראשונה, ליד הכותרת "Number1", יש להגדיר את טווח הערכים שעבורו רוצים לחשב את הממוצע. ניתן לעשות זאת בשתי דרכים: לכתוב ידנית את הטווח או לסמן אותו באמצעות העכבר. כדי לכתוב את הטווח באופן ידני, יש לכתוב את כתובת התא הראשון בטווח, ואז נקודותיים (:). ולאחר מכן כתובת התא האחרון (ר' בדוגמה שלמעלה). כדי לסמן את הטווח באמצעות העכבר, יש ללחוץ על השורה הלבנה שליד הכותרת "Number1", ולאחר מכן לסמן באמצעות העכבר את התאים הרצויים בגליון. לבסוף, יש להקיש על "אישור".

דרך נוספת להשתמש בפונקציות, שלא דרך תפריטי התוכנה, היא לכתוב אותן באופן ידני. כדי לעשות זאת, יש לעמוד עם הסמן על התא הרצוי, לכתוב את הסימן שווה (=) כדי לציין שעומדים להזין נוסחה, ואז את שם הפונקציה (למשל, Average) ולאחריו לפתוח סוגריים. לאחר הסוגריים, תוכלו להזין את כל הנתונים הנחוצים לשם חישוב הפונקציה שבחרתם. שימו לב, שהמחשב יכתוב לכם בתווית צהובה מתחת לפונקציה, אילו ערכים יש להכניס לתוכה. להלן רשימה חלקית של פונקציות סטטיסטיות ומתמטיות שימושיות נוספות, שבהן אפשר להשתמש באותו האופן שבו השתמשתם בפונקציית הממוצע Average. שימו לב כי לכל הפונקציות יש מסכי עזרה, וכדאי להשתמש בהם במקרה הצורך.

שם הפונקציה	תיאור
Sqrt	הוצאת שורש
Count	ספירת האיברים בטווח מסויים
Countif	ספירה מותנית (ספירת האיברים שמקיימים תנאי מסויים)
Sum	חישוב סכום של איברים
Sumif	חישוב סכום מותנה (סכום האיברים שמקיימים תנאי מסויים).

חישוב ממוצע	Average
חישוב סטיית תקן	Stdevp
חישוב מתאם פירסון	Pearson/ Correl
חישוב שונות משותפת (covariance)	Covar
חישוב הטרנספורציה של פישר (למשל, בעת ביצוע מבחן סטטיסטי למתאמים)	Fisher
מציאת הערך המקסימלי בקבוצת איברים	Max
מציאת הערך המינימלי בקבוצת איברים	Min
מציאת החציון	Median
מציאת השכיח	Mode
מציאת הערך שבאחוזון ("מאון") ה-k של קבוצת נתונים	Percentile
חישוב השונות של קבוצת איברים	Varp
ביצוע מבחן t על סדרת נתונים (או 2 סדרות, תלויות או ב"ת)	Ttest

מיון נתונים:

- ניתן למיין את קובץ הנתונים בסדר עולה או יורד, לפי ערכי אחד המשתנים (או יותר). מיינו את הקובץ על פי גיל הנבדק בסדר יורד:
- א. סמנו את כל הנתונים באמצעות העכבר.
- ב. בחרו 'נתונים' < 'מיון'.
- ג. יפתח חלון ובו אפשרויות למיין את הנתונים:

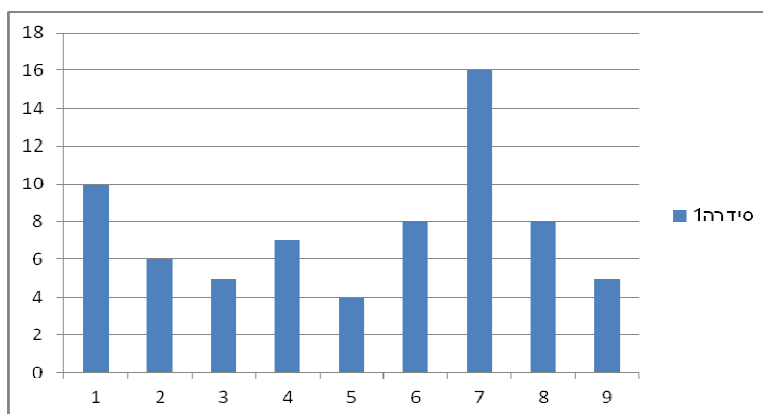
- ד. בחרו באפשרות 'יש שורת כותרת'.
- ה. בחרו במשתנה 'גיל' ובאפשרות 'סדר יורד', ולבסוף לחץ 'אישור'.

חשוב לשים לב שאכן כל הנתונים נבחרו (נצבעו) לפני פקודת המיון, במידה ולא, הדבר יוביל למיון של הנתונים הצבועים בלבד, שאר הנתונים לא יתמיינו. במצב זה לא תישמר השייכות של המשתנה הממוין לשאר הנתונים.

תרשימים וגרפים:

ב-Excel יש אפשרות להציג מגוון רחב של גרפים, וכן לערוך אותם בקלות מבחינה ויזואלית. כדי להציג נתונים בגרף, יש לסמן אותם תחילה, ואז להיכנס לתפריט "הוספה" ולבחור בסוג התרשים המבוקש.

צרו דיאגרמת עמודות שתתאר את מספר הארוחות שאכל כל אחד מהנבדקים בסוף השבוע. ציר ה-X יהיה מספר הנבדק, וציר ה-Y יהיה מספר הארוחות בסוף השבוע. כדי לעשות זאת, יש לסמן (להשחיר) את התאים M2 עד M10, המכילים את נתוני מספר הארוחות בסוף השבוע. לאחר מכן, יש להיכנס ל"הוספה" ולבחור ב"טורים" (למעשה, השם כאן מטעה: זהו תרשים עמודות), ובסוג המשנה הראשון.



עמידה על שטח הגרף ולחיצה על מקש העכבר הימני תפתח מספר אפשרויות לשינוי והתאמת הגרף.

תרגיל:

צרו כעת דיאגרמת עוגה (pie), שתתאר את אחוז הגברים ואחוז הנשים במדגם. להזכירכם, דיאגרמת עוגה משתמשת כדי לתאר שכיחות יחסית של קטגוריות שונות מתוך המדגם. שימו לב: כדי ליצור את התרשים המבוקש, יש לספור תחילה את שכיחות המקרים מכל סוג. עמדו על התא B20. בחרו בפונקציה 'count if', בחלון שיפתח יש לבחור את הטווח של הערכים (range) – נסמן את הטור כולו. ואת הקריטריון לבחירה (criteria) – נסמן תחילה את הערך של הגברים, "זכר" (הערה: יש לשמור על הסימן גרשיים " " כאשר הערך הרצוי אינו מספרי אלא מילה, כמו בדוגמה הנוכחית). כעת נחזור על אותן פעולות ונספור את מספר הנשים. ניתן לרשום מימין לכל ערך למה הוא מתייחס. כעת, השחירו את הערכים ושמותיהם ולחצו על צלמית התרשים. בחרו באפשרות 'עוגה', ובתוכה את אחד מתת-הסוגים לפי רצונכם. המשיכו לעצב את התרשים כרצונכם, ומקמו אותו כאובייקט בגיליון הנתונים הנוכחי. דוגמה אפשרית לעיצוב כזה נמצאת בקובץ food המצורף.

תרגיל הגשה מספר 1

לפניכם קטע מתוך שאלון העוסק בתופעת האלימות בבית הספר. להלן יופיעו הפריטים (השאלות), שם המשתנה שבו יופיע כל פריט בקובץ ה-Excel, ושם המשתנה שמודד הפריט. הנתונים מופיעים בקובץ violence המצורף.

מבנה השאלון:

א. מידע כללי -

שם בגיליון	הפריט	
CLASS	כיתה	1
SEX	מין התלמיד	2

ב. אוירה בבית הספר -

סולם המדידה בכל אחד מהפריטים:

1 - כלל לא, 2 - מעט, 3 - מדי פעם, 4 - די הרבה, 5 - הרבה מאוד.

משתנה תיאורטי	שם בגיליון	הפריט	
קללות	B1	בבית הספר יש בעיה של תלמידים המקללים מורים	1
ונדליזם	B2	תלמידים שוברים דברים בבית הספר (ונדליזם)	2
שתיה וסמים	B3	תלמידים שותים משקאות חריפים	3
שתיה וסמים	B4	תלמידים משתמשים בסמים	4
תחושת בטחון	B5	בדרך כלל אני מרגיש בטוח ומוגן בבית הספר	5
טיפול המורים	B6	המורים מצליחים לטפל טוב בתלמידים אלימים שעושים צרות	6
אכפתיות המורים	B7	למורים אכפת שלא יהיו התנהגויות אלימות	7
טיפול המורים	B8	המורים עושים דברים כדי להפחית התנהגויות אלימות	8
טיפול המורים	B9	המורים מצליחים להפחית התנהגויות אלימות	9
ידיעת חוקי בית הספר	B10	חוקי בית הספר מוכרים לי ואני יודע אותם	10
ידיעת חוקי בית הספר	B11	המורים דואגים לשוחח איתנו על חוקי בית הספר	11
ידיעת חוקי בית הספר	B12	חברי בכיתה יודעים את חוקי בית הספר	12

א. חשבו ממוצע וסטיית תקן לכל אחת מהשאלות, עבור שתי הכיתות יחד.

ב. חשבו ממוצע לכל אחת מהשאלות, עבור כל כיתה בנפרד. רמז: לשם כך יש למיין תחילה את הנתונים לפי כיתות.

ג. פריטי שאלון האווירה בבית הספר מאורגנים לפי משתנים תיאורטיים, כך שכל משתנה תיאורטי נמדד על-ידי מספר פריטים בשאלון (למשל, המשתנה "טיפול המורים" נמדד באמצעות הפריטים B6, B8 ו-B9). צרו עבור כל תלמיד משתנים חדשים עבור המשתנים "שתיה וסמים", "טיפול המורים" ו-"ידיעת חוקי בית הספר", שיכילו את ממוצע

- הפריטים הרלוונטיים עבור תלמיד זה. חשבו ממוצע וסטיית תקן לכל אחד ממשתנים אלה, עבור שתי הכיתות יחד.
- ד. הציגו גרף עמודות, שישווה את ממוצע כל אחד מהמשתנים, בשתי הכיתות. עצבו את הגרף כרצונכם, ומקמו אותו בגיליון חדש.
- ה. הציגו גרף עוגה, שיתאר את התפלגות התשובות לפריט 3 בשאלון.
- ו. שמרו את הקובץ.
- ז. אנא הגישו את תדפיס קובץ הנתונים ותדפיסי התרשימים למתרגלת שאליו/ה אתם רשומים. רשמו שם + ת"ז על גבי התדפיסים. נא לא להגיש את התרגיל בתיקיות או בשקיות ניילון ("שמרדף"), אין בכך צורך וחבל על הפגיעה באיכות הסביבה:

שיעור מספר 2

שיעור מס' 2: הכרות עם תוכנת SPSS – בניית קובץ נתונים וארגונו

תוכנת EXCEL היא גיליון אלקטרוני יעיל מאד לביצוע פעולות על בסיס נתונים גדול, כמו חישובים של מדדים והצגת נתונים בטבלה או בתצוגה גרפית. ב-EXCEL ניתן לבצע פונקציות סטטיסטיות בסיסיות. אולם, לצרכים סטטיסטיים ולצורך ניהול וניתוח נתוני מחקר קיימות תוכנות סטטיסטיות ייעודיות, בעלות אפשרויות סטטיסטיות וגראפיות רבות אשר אינן קיימות בתוכנות לשימוש כללי כגון EXCEL. בקורס זה תלמדו להשתמש בתוכנת SPSS (ראשי תיבות של Statistical Package for the Social Sciences). תוכנה זו נפוצה מאוד, ונועדה במיוחד לשימושם של חוקרים בתחומי מדעי החברה.

בשונה מ-EXCEL, שמורכב מטבלת נתונים אחת גדולה, ב-SPSS קיימים שלושה סוגי חלונות, המיועדים לצרכים שונים:

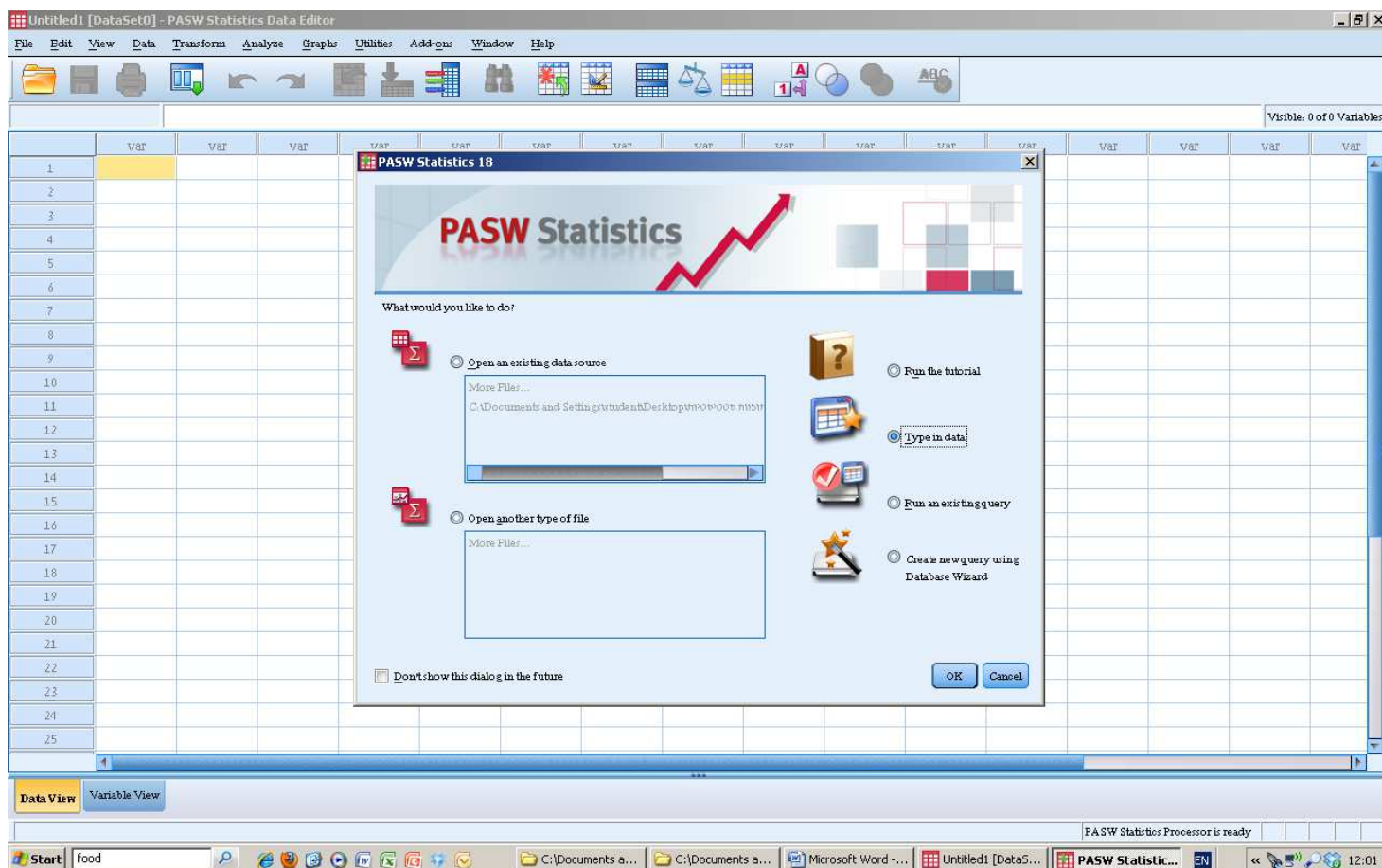
1. חלון נתונים – Data Editor: בחלון זה מקלידים את הנתונים, מגדירים את המשתנים ואת תוויות המשתנים והערכים. חלון זה דומה מעט לגיליון הנתונים של EXCEL, והוא מכיל טבלה בעלת מספר רב של שורות ושל עמודות. ניתן לשמור את טבלת הנתונים. קבצי נתונים ב-SPSS הינם בעלי הסיומת sav (לדוגמה, קובץ הנתונים שניצור בשיעור זה ייקרא inequality.sav).
2. חלון פקודות – Syntax: בחלון זה מקלידים את פקודות עיבוד הנתונים שאותן רוצים לבצע. ב-SPSS ניתן לבצע תהליכי עיבוד נתונים בשתי דרכים. דרך אחת הינה לעבוד באמצעות התפריטים של התוכנה, ומהם לבחור את סוג הניתוח המבוקש, המשתנים שעליהם ייערך הניתוח, והגדרות רלוונטיות נוספות. דרך אחרת היא לכתוב בחלון ה-Syntax את הפקודות, באופן ישיר בשפת SPSS. היתרון של כתיבת הפקודות בחלון ה-Syntax הוא היכולת לשמור קובץ פקודות זה, על מנת לבצע את אותו הניתוח גם בעתיד (על קבצי נתונים אחרים, למשל). קבצי Syntax ב-SPSS הינם בעלי הסיומת sps (לדוגמה, קובץ ה-Syntax שניצור בשיעור זה ייקרא inequality.sps).
3. חלון תוצאות – Output: בחלון זה יתקבלו התוצאות של הניתוחים הסטטיסטיים שבוצעו. למשל, אם בקובץ ה-Syntax נכתבה פקודה המחשבת ממוצע של משתנה, ערכו של הממוצע יוצג בחלון התוצאות. ניתן לשמור את חלון התוצאות בקובץ. הסיומת של קבצי תוצאות ב-SPSS היא spo (לדוגמה, קובץ התוצאות שניצור בשיעור זה ייקרא inequality.spo).

במהלך שיעור זה תכירו את השימוש בכל אחד מהחלונות הללו. שימו לב שכל חלון נראה אחרת, ושלכל סוג חלון יש סרגל כלים שונה. לכן, במידה וקיימים מספר חלונות פתוחים המוצגים על המסך בו זמנית, חשוב לבדוק מהו החלון הפעיל, כלומר החלון שעליו אנחנו עובדים כרגע. החלון הפעיל הינו החלון שאינו מכוסה על ידי שום חלון אחר במסך, ושסרגל הכותרת שלו מודגש (לעומת החלונות ברקע שבהם סרגל הכותרת הוא אפור). ניתן להפוך כל חלון לפעיל על ידי לחיצה באמצעות העכבר בכל מקום בשטח החלון.

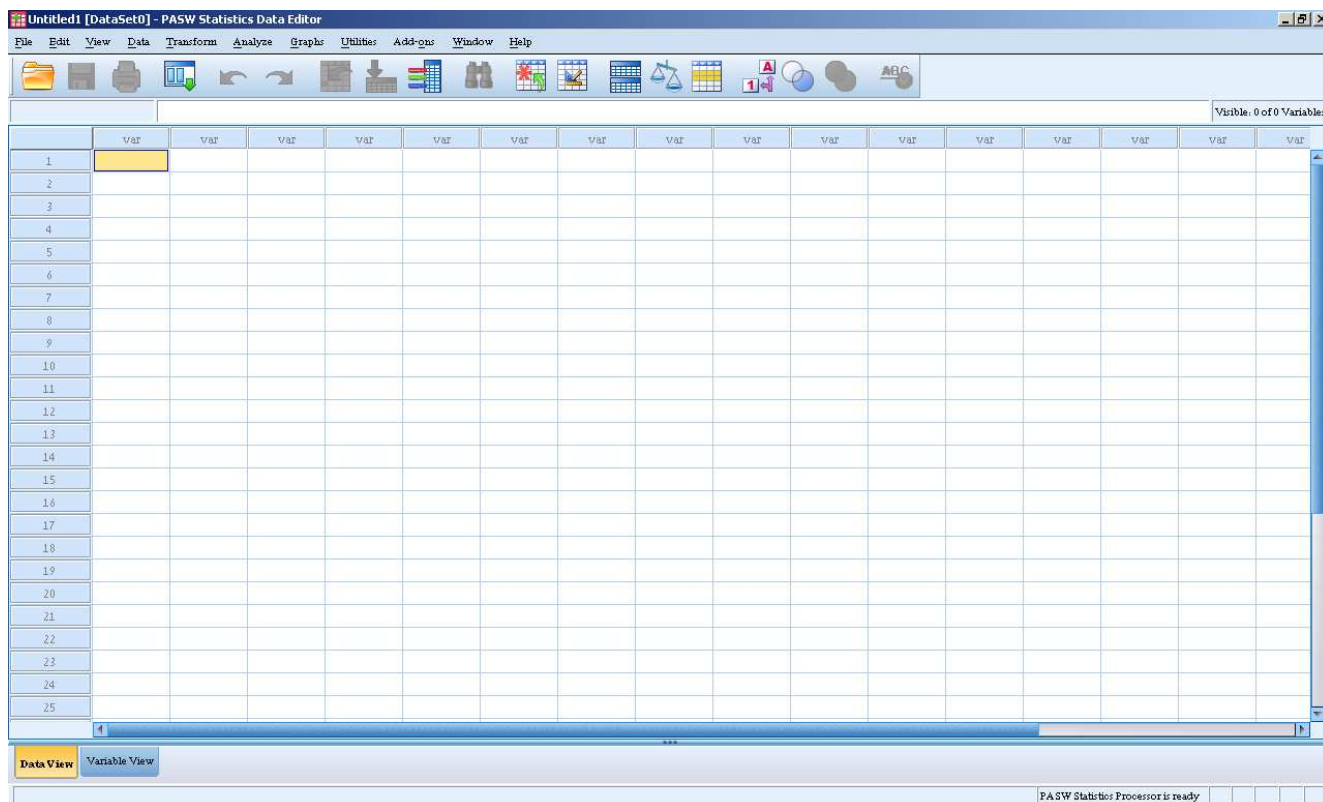
הדגמה ושימוש

את התוכנה ניתן לפתוח דרך התפריט הראשי של Windows: יש ללחוץ על "התחל" (או – "Start"), ואז על "תוכניות" (או – "Programs"). מתפריט התוכניות יש לבחור את התוכנית SPSS וזו את SPSS for Windows.

יפתח בפנינו החלון הבא:



נבחר באפשרות 'type in data' על מנת ליצור קובץ נתונים חדש, ואז נלחץ על OK. בהמשך, כאשר נרצה לפתוח קובץ נתונים קיים, נבחר באפשרות 'open an existing file' ולאחר מכן נאתר את הקובץ הרצוי לפי מיקומו במחשב. כעת ייפתח החלון הבא:



הזנת ערכים

זהו חלון הנתונים - Data Editor. כעת נרצה להזין נתונים לתוך הקובץ: לפניך נתונים שנאספו במסגרת מחקר בנושא אי-שוויון וריבוד חברתי. שאלת המחקר הכללית הינה האם החברה הישראלית היא חברה פתוחה ומובילית, או סגורה, בעלת מבנה ריבודי נוקשה ושיעורי ניעות חברתית נמוכים. הדרך האופרציונלית לבדוק זאת במחקר הנוכחי היא לבחון האם יש קשר בין מאפיינים שיוכיים (הקשורים להשתייכות האדם לקבוצות שונות בחברה, כגון מגדר וארץ מוצא) לבין משתנים השגיים (הקשורים בהישגיו). כדי לבחון את השאלה, נדגם מספר גדול של תושבי מדינת ישראל, ונאספו נתונים אודות המשתנים הבאים: משתנים שיוכיים:

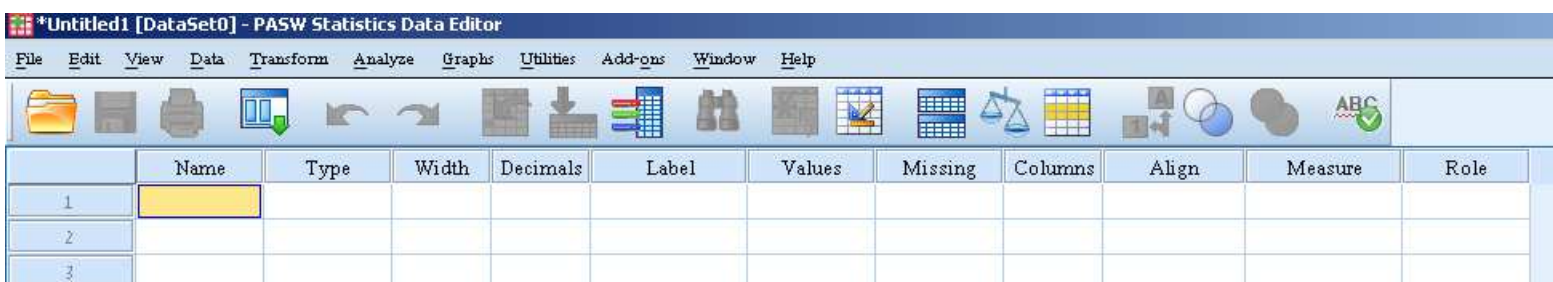
1. מגדר: זכר, נקבה. שם באנגלית - Gender
2. לאום: יהודי/ה, ערבי/ה, לא יהודי/ה - אחר. שם באנגלית - Nationality
3. מוצא (ליהודים בלבד): מזרחי/ה, אשכנזי/ה, אחר. שם באנגלית - Origin

משתנים הישגיים:

1. השכלה: מספר שנות לימוד. שם באנגלית - Education
 2. הכנסה: ממוצע שכר חודשי (ברוטו) בשנה האחרונה. שם באנגלית - Income
- להלן תוצאותיהם של 10 נבדקים מתוך מחקר זה:

מספר נבדק	מגדר	לאום	הגדרה עצמית	השכלה	הכנסה
1	זכר	יהודי	מזרחי	14	7500
2	זכר	יהודי	אשכנזי	10	6200
3	נקבה	יהודי	מזרחי	12	5400
4	זכר	אחר		9	3750
5	נקבה	יהודי	אשכנזי	8	12000
6	נקבה	מוסלמי		10	8750
7	נקבה	אחר		13	4990
8	זכר	מוסלמי		12	3550
9	נקבה	יהודי	אחר	10	8900
10	זכר	מוסלמי		11	10200

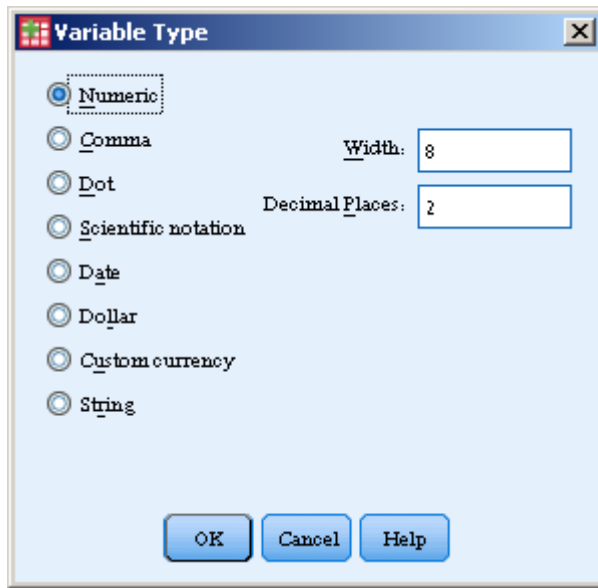
את הנתונים נקליד בחלון ה-Data Editor. בצד השמאלי התחתון של חלון זה, מופיעות שתי תגייות: Data view ו-Variable view. נתחיל בהגדרת המשתנים עצמם (בחלון ה-Variable view), ולאחר מכן נקליד את הנתונים בחלון ה-Data view. ניכנס לחלון ה-Variable view. כל שורה בטבלה זו מייצגת משתנה, ומתארת את המאפיינים שלו (כגון שם, סולם מדידה, ערכים אפשריים וכו'). להלן רשימת המאפיינים העיקריים:



❖ **Name** – שם המשתנה. SPSS, כמו תוכנות סטטיסטיות אחרות, לא תומכת בשמות משתנים בעברית. לכן, יש להגדיר את שמות המשתנים ואת ערכיהם באנגלית. שם המשתנה חייב להתחיל באות, יכול לכלול עד 8 תווים (אותיות, ספרות וסימנים, לא כולל סימנים חשבוניים ורווחים).

❖ **Type** – כאן מגדירים את סוג המשתנה. סוג המשתנה מתאר את סוג הנתונים (הערכים) שניתן יהיה להכניס לתוך משתנה זה. סוג המשתנה יכול להיות מספרי (Numeric), מילולי

(String), תאריך (Date) וכו'. סימון התא יציג כפתור עם 3 נקודות בחלקו הימני של התא. לחיצה עליו תכניס אותנו לתיבת דו-שיח להגדרת סוג המשתנה:



בתיבה זו ניתן לבחור את סוג המשתנה, ומאפיינים נוספים רלוונטיים (למשל, רוחב העמודה, ומספר הספרות שיופיעו לאחר הנקודה העשרונית).

בתיבת דו-שיח זו, כמו גם ברוב המסכים ב-SPSS, ניתן להשתמש באפשרויות העזרה של התוכנה כדי לקבל מידע אודות האפשרויות השונות. עברו בדרך זו על סוגי המשתנים השונים, ובדקו איזה סוג מידע יכול להכיל כל אחד מהם.

** עבור משתנים מספריים (Numeric), ניתן להגדיר את רוחב העמודה (Width) – מספר הספרות שיכילו המספרים שיוזנו בעמודה זו, ואת מספר הספרות שיש להציג לאחר הנקודה העשרונית (Decimals).

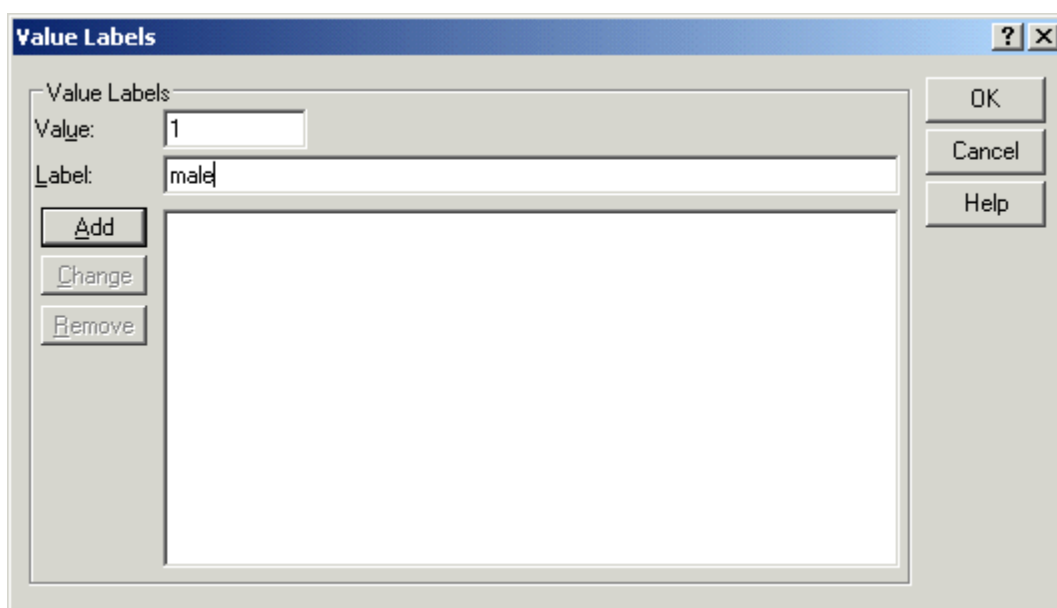
בשלב זה, הזינו את שמות המשתנים הרלוונטיים לדוגמה שלעיל, ואת סוג המידע שיכיל כל אחד מהם.

שימו לב: פעמים רבות, יהיה נוח יותר להזין נתונים מספריים גם עבור משתנים בסולם שמי. למשל, במקום לכתוב עבור כל אחת מהתצפיות בטבלה את הערך "Mizrachi", "Ashkenazi" או "Other", יהיה נוח יותר להשתמש במפתח מספרי לערכים אלה, למשל הספרות 1, 2 ו-3, בהתאמה (מזרחי=1, אשכנזי=2, אחר=3). במקרים אלה, ניצור משתנים מסוג מספרי (Numeric) גם עבור משתנים שערכיהם ה"אמיתיים" הינם בסולם שמי.

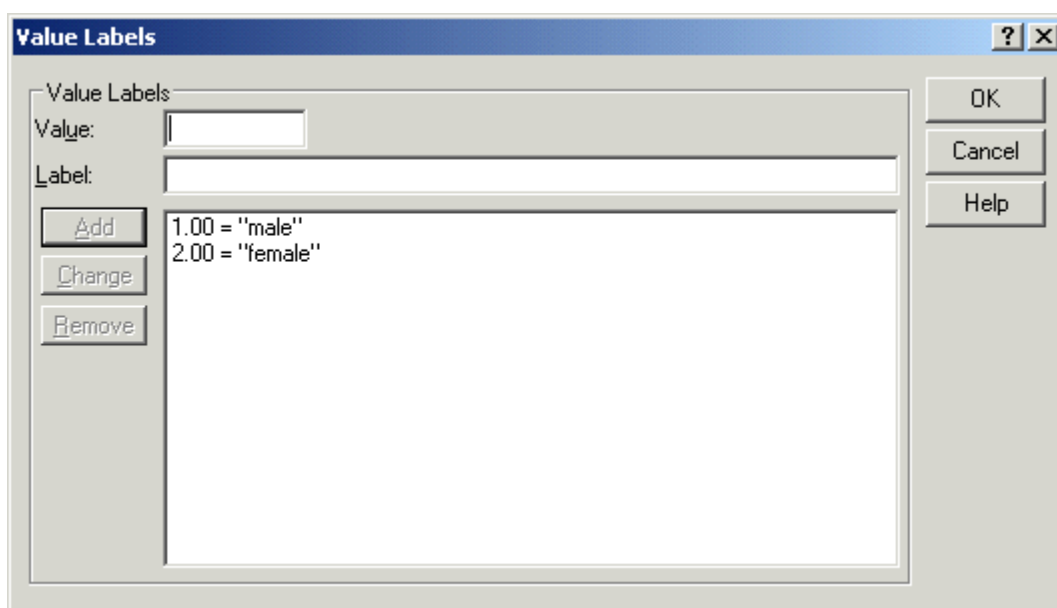
❖ **Label** – מתן תווית למשתנה. מאחר ושמות משתנים הינם קצרים, נוח פעמים רבות ליצור בתווית ארוכה יותר, שתכיל מידע רב יותר אודות המשתנה. התווית אינה שמו של המשתנה, והיא נועדה להקל על החוקר לזכור מה משמעותו של כל אחד מהמשתנים. צרו את התווית "Average monthly income in the previous year" עבור המשתנה Income.

❖ **Values** - כאן נותנים תוויות לערכי המשתנה. יעיל במיוחד כאשר משתמשים בערכים מספריים לתיאור רמות שמיות. למשל, במשתנה "Gender", נחליט שהערך '1' יינתן לתצפיות שבהן המין הוא זכר והערך '2' לתצפיות שבהן המין הוא נקבה. בכדי שלא נצטרך לזכור זאת, נכניס תוויות אלה להגדרת המשתנה עצמו. לחצו על התא בשורה הראשונה (של המשתנה

Gender), בעמודה Values. לחצו על שלושת הנקודות בחלקו הימני של התא. יפתח חלון שבו ניתן להגדיר את ערכי המשתנה. כדי לעשות זאת, יש לרשום את הערך בתא "Value", ואת התווית המתאימה לו בתא "Label", למשל:



לאחר מכן, יש ללחוץ על הכפתור ADD, ולהכניס את התווית הבאה (למשל, 2 ו- female):



עם סיום הכנסת כל התוויות הרצויות, יש ללחוץ על הכפתור OK.

❖ **Missing values** – ערכים חסרים. עמודה זו מאפשרת להגדיר אילו ערכים לא ייחשבו כערכים מספריים, ולא ייכנסו לחישוב. למשל, עבור נבדק שבחר שלא לענות על השאלה מהי הכנסתו החודשית, בחרו עורכי המחקר להקליד את הערך "1-" במשתנה Income. הערך 1- הינו סימן המציין כי הנבדק לא השיב, ואין משמעותו "ההכנסה היא -1 לך". לכן, לא נרצה לבצע חישובים סטטיסטיים על ערך זה (למשל, לא נרצה שהערך 1-

יכנס לחישוב הממוצע, השונות וסטיית התקן של ההכנסות הבודדות), ובעמודה זו נוכח להגדירו כערך חסר.

❖ **Columns** - כאן ניתן לקבוע את רוחב העמודה. ככל שהערך המוזן קטן יותר העמודה תהיה צרה יותר.

❖ **Align** - קובע את עימוד הטקסט בתא - יישור לימין שמאל, מרכז

❖ **Measure** - ניתן לקבוע את סולם המדידה של המשתנה כשמי-Nominal, סדר-Ordinal או רציף-Scale (מכיל רווח ומנה כאחד).

הגדירו את כל המשתנים הרלוונטיים שבטבלה בחלון ה-Variable view. הוסיפו תוויות לערכי המשתנים Gender, Nationality ו-Origin. בסיום, הטבלה תיראה כך:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Gender	Numeric	8	0		{1, male}...	None	8	Right	Nominal
2	Nationality	Numeric	8	0		{1, jewish}...	None	8	Right	Nominal
3	Origin	Numeric	8	0		{1, Mizrach...	None	8	Right	Nominal
4	Education	Numeric	8	2		None	None	8	Right	Scale
5	Income	Numeric	8	2		None	None	8	Right	Scale
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										

בשלב זה כדאי לשמור את קובץ הנתונים. בתפריט הראשי של SPSS (מתחת לשורת הכותרת הכחולה) יש לבחור בתפריט File, ואז באפשרות Save as. כעת ניתן לבחור את התיקייה במחשב שבה יישמר קובץ הנתונים, ואת שם הקובץ. שמרו את קובץ הנתונים בשם Inequality. הקובץ שיווצר ייקרא Inequality.sav.

לאחר הגדרת המשתנים, ניתן להזין את הנתונים עצמם. לשם כך, יש ללחוץ על התגית Data view שבחלקו השמאלי התחתון של המסך. את הנתונים יש להזין בתוך החלון שייפתח. על מנת להקליד ערך יש לסמן את התא המתאים, ללחוץ על הכפתור השמאלי בעכבר ולהזין את הערך הרצוי. על מנת לעבור בין תאים ניתן לנוע באמצעות החיצים.

לאחר הזנת הערכים על חלון ה- Data Editor להראות כך :

	Gender	Nationality	Origin	Education	Income	var	var	var	var	var	var
1	1	1	1	14.00	7500.00						
2	1	1	2	10.00	6200.00						
3	2	1	1	12.00	5400.00						
4	1	3	.	9.00	3750.00						
5	2	1	2	8.00	12000.00						
6	2	2	.	10.00	8750.00						
7	2	3	.	13.00	4990.00						
8	1	2	.	12.00	3550.00						
9	2	1	3	10.00	8900.00						
10	1	2	.	11.00	10200.00						

תרגיל הגשה מספר 2

צרו ב-SPSS קובץ נתונים, הזזה לקובץ הנתונים violence ששימש אתכם בתרגיל ההגשה הקודם ב-EXCEL.

לשם כך, הגדירו את המשתנים המתאימים, צרו תוויות עבור המשתנים ועבור ערכיהם, והקלידו את נתוני הקובץ violence ל-SPSS.

שמרו את קובץ ה-SPSS שנוצר בשם sav. מספר ת.ז.

(לדוגמא : 2221111.sav), והגישו אותו למייל : stat.bgu@gmail.com

נקודה למחשבה: האם יש דרך קלה יותר להעביר את הנתונים מ-EXCEL ל-SPSS, במקום להקלידם ידנית? נסו לבדוק זאת באמצעות בחירה בתפריט הראשי באפשרות File ואז Open. אם זה לא מצליח, כמובן שתוכלו להקליד את כל הערכים בעצמכם.

שיעור מספר 3

שיעור מס' 3: סטטיסטיקה תיאורית בסיסית ב-SPSS

בשיעור הקודם למדתם כיצד להזין קובץ נתונים לתוכנת ה-SPSS, וכיצד להגדיר את המשתנים בקובץ. נקודה חשובה היא, שב-SPSS (כמו גם בתוכנות סטטיסטיות אחרות כגון STATISTICA, ו-SAS), כל שורה בגיליון מתארת תצפית במדגם (למשל: נבדק), והעמודות מציינות את המשתנים.

בשיעור זה תלמדו לבצע ניתוחים סטטיסטיים בסיסיים על קבצי נתונים. בפרט, תלמדו כיצד להפיק מדדים סטטיסטיים תיאוריים, וכיצד להציג נתונים באופן גרפי.

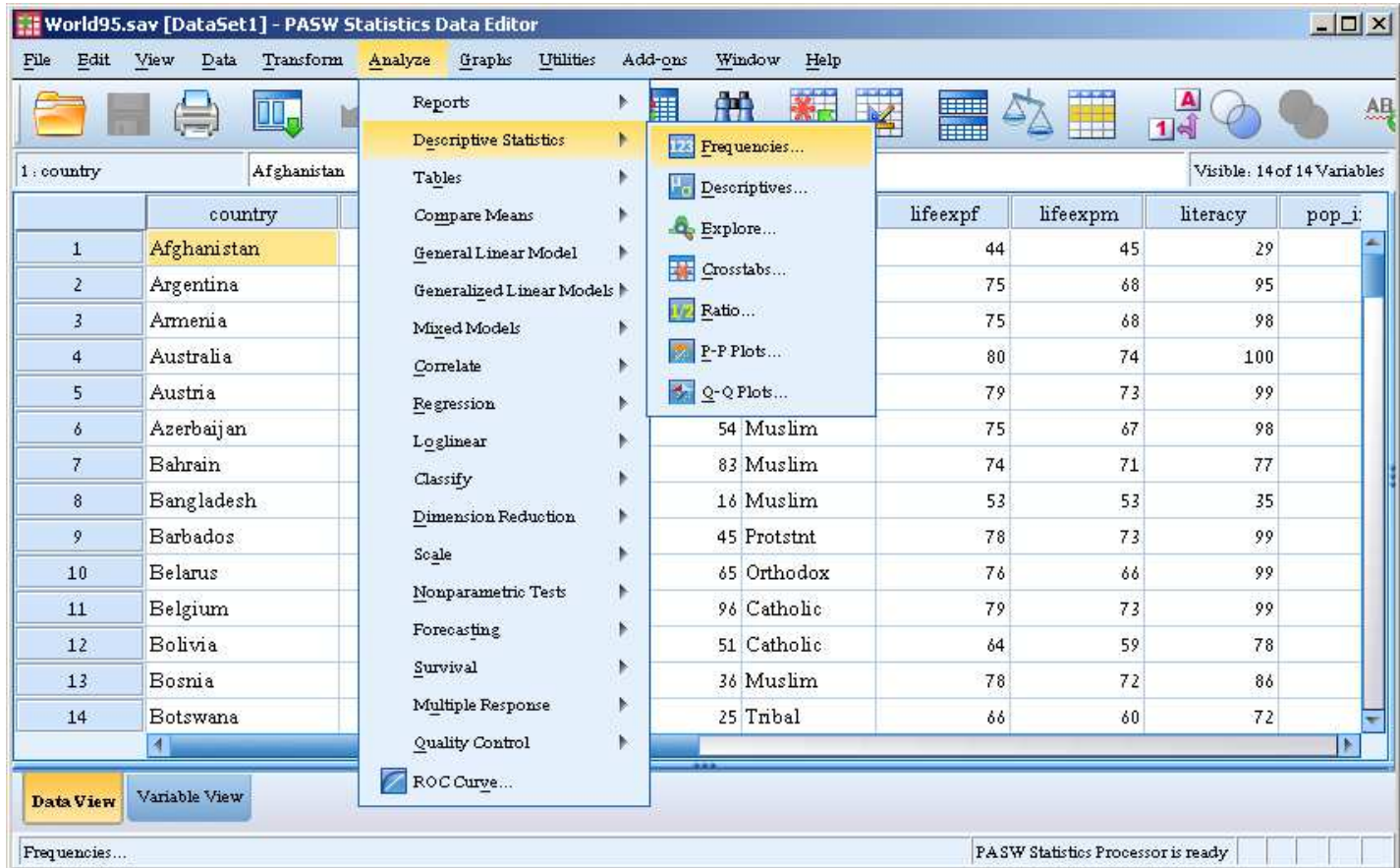
השיעור יתבסס על קובץ הנתונים World95.sav, המכיל מידע אודות כל מדינות העולם, הנכון לשנת 1995. להלן שמות המשתנים בקובץ:

country - שם המדינה
populatn – אוכלוסיה (באלפים)
density – צפיפות אוכלוסין (מספר תושבים לקילומטר רבוע)
urban – אחוז עיור (אחוז התושבים המתגוררים בערים)
lifeexpf – תוחלת החיים הממוצעת לנשים
lifeexpm – תוחלת החיים הממוצעת לגברים
literacy – אחוז התושבים היודעים קרוא וכתוב
pop_incr – קצב גידול האוכלוסין (באחוזים לשנה)
calories – צריכת קלוריות יומית ממוצעת בקרב התושבים
aids – מספר חולי איידס
fertility – מספר ילדים ממוצע במשפחה
lit_male – אחוז הגברים היודעים קרוא וכתוב
lit_female – אחוז הנשים היודעות קרוא וכתוב

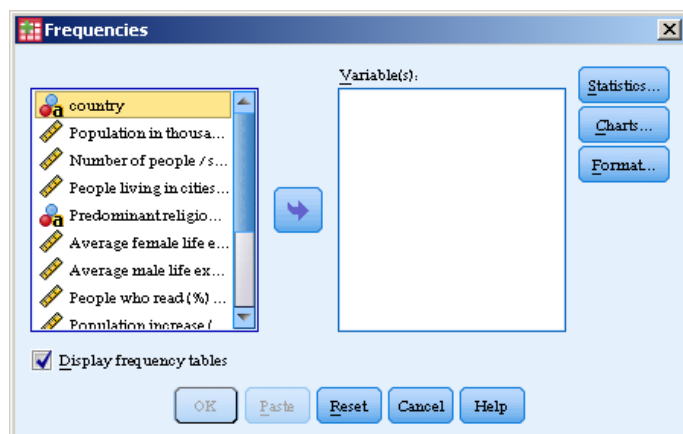
פתחו את קובץ הנתונים. שימו לב, שבקובץ הנתונים קיימים ערכים חסרים. למשל, במשתנה calories כתובה נקודה (.). עבור המדינות אפגניסטן, ארמניה, אזרביג'אן וכו', במשתנה aids כתובה נקודה עבור אזרביג'אן, בוסניה, וכן הלאה. משמעותה של נקודה היא "ערך חסר". מצב זה מציין כי הנתון הספציפי אינו ידוע, או אינו קיים בקובץ. לעיתים יופיעו ערכים חסרים שבהם לא מסומנת נקודה בתוך התא הרלוונטי, כמו למשל במשתנה religion, עבור המדינה דרום-אפריקה (south Africa). באופן מעשי, אין הבדל בין המצב שבו כתובה נקודה למצב שבו היא אינה כתובה.

יצירת טבלת שכיחויות

מהי התפלגות הדתות בין המדינות השונות? מהו אחוז המדינות שבהן הדת השלטת הינה איסלאם? או בודהיזם? כדי לענות על שאלות אלה, ושאלות אחרות הקשורות לאופן שבו מתפלגים משתנים בתוך המדגם, נרצה לראות את התפלגות השכיחויות של המשתנה הרלוונטי. בתפריט הראשי של SPSS, בחלון ה- Data View, נבחר בתפריט Analyze. תפריט זה ימשך עבור כל הניתוחים הסטטיסטיים שנרצה לבצע (כעת ובעתיד). בהמשך, נבחר בתת-תפריט Descriptive Statistics, ואז נבחר באפשרות Frequencies (שכיחויות):

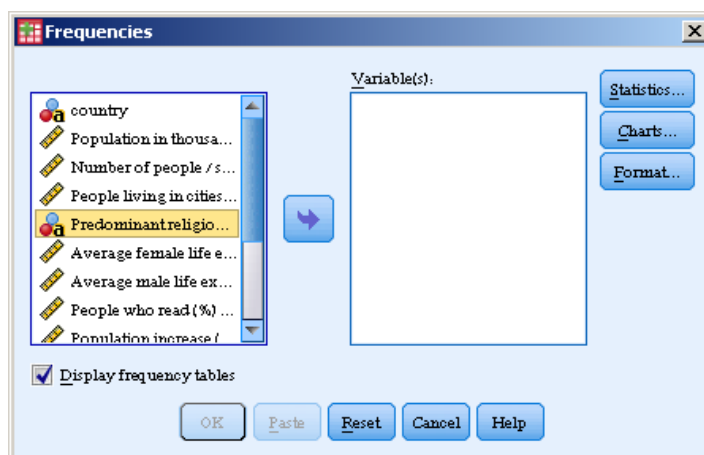


לאחר הבחירה באפשרות Frequencies, ייפתח החלון הבא:

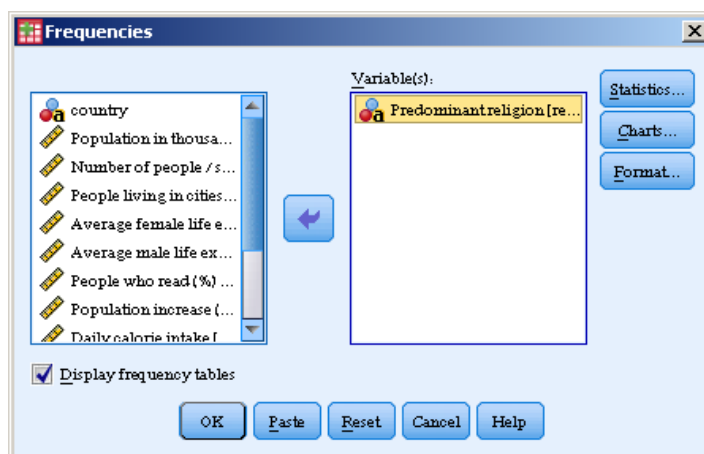


בחלקו השמאלי של החלון, תופיע רשימת כל המשתנים בקובץ הנתונים. הרשימה לא מכילה את שמות המשתנים עצמם, אלא את התוויות שלהם (תוכלו לעבור ולהסתכל בחלון ה- Variable View) כדי לראות את התוויות של המשתנים המוגדרים בקובץ. שימו לב שמשתנים שערכיהם מוגדרים באופן מילולי (כמו country ו-religion, המסומנים בשני עיגולים ובאות a) יסומנו באופן שונה ממשתנים שערכיהם הינם מספריים (כמו population ו-density, המסומנים בסרגל).

בחלקו הימני של החלון, תופיע רשימת המשתנים שעליהם ברצונכם לבצע את הניתוח הסטטיסטי. כרגע, לא מופיע בה אף משתנה. כדי לבחור את המשתנים הרלוונטיים, יש לסמן אותם בחלונות השמאלית (המכילה את רשימת המשתנים), ואז ללחוץ על החץ שבין שני החלונות:



המשתנה שבחרתם (בדוגמה זו, religion, שעליו נבצע את המשך הניתוח) ייכנס לחלון המשתנים שנבחרו:



באופן עקרוני, ניתן להגדיר מספר משתנים (ולא רק אחד) שעליהם יתבצע הניתוח. במקרה זה, הניתוח יתבצע עבור כל אחד מהם בנפרד. לדוגמה, תוצג טבלת השכיחויות עבור המשתנה הראשון, ולאחר מכן טבלת השכיחויות עבור המשתנה השני, וכן הלאה.

לאחר שהוגדרו המשתנים לניתוח, יהפכו לפעילים שני הלחצנים בחלון הימני העליון של החלון: Paste ו-OK. חשוב להבין היטב את ההבדל בין לחצנים אלה. לחיצה על הכפתור OK תגרום לביצוע הניתוח הסטטיסטי שהוגדר (במקרה זה, יצירת טבלת שכיחויות), ולהצגת התוצאות בחלון ה-output. זוהי הדרך הקצרה והמהירה ביותר לבצע את הניתוח. לעומת זאת, לחיצה על הכפתור Paste לא תבצע את הניתוח ממש, אלא תפתח חלון פקודות (syntax), ותכתוב בתוכו את פקודות ה-SPSS המתאימות לביצוע הניתוח שהוגדר. באופן עקרוני, קיימות שתי דרכים לביצוע ניתוחים סטטיסטיים ב-SPSS. דרך אחת היא באמצעות השימוש בתפריטי התוכנה (כמו שעשינו כדי לפתוח את החלונת Frequencies). תפריטים אלה מאפשרים להגדיר את המשתנים הרלוונטיים, ואת סוג הניתוח הנדרש. הדרך השנייה היא לכתוב את ההוראות לביצוע הניתוח בחלון ה-Syntax, באופן הדומה לתיכנות בשפות מחשב אחרות. לאחר כתיבת הפקודות, ניתן להריץ אותן כדי לבצע את הניתוח בפועל. למרות השימוש בתפריטים הוא קל ונוח יותר (אין צורך לכתוב פקודות ולזכור כיצד להגדיר את הניתוחים הרצויים), היתרון הגדול של כתיבת הפקודות בחלון ה-Syntax הוא היכולת לשמור אותן בקובץ. באופן זה, ניתן יהיה לבצע את הניתוח פעמים רבות, מבלי להגדירו בכל פעם מחדש. באמצעות הכפתור Paste, הנמצא בכל אחד מתפריטי הניתוחים הסטטיסטיים, ניתן להדביק בחלון ה-Syntax את הפקודות (בשפת SPSS) המתאימות לניתוח שהוגדר בתפריט. בהמשך השיעור נשתמש באפשרות זו.

כעת, ניגש לביצוע הניתוח הסטטיסטי שהגדרנו. למעשה, ביקשנו להציג טבלת שכיחויות של המשתנה religion. נלחץ על הכפתור OK לביצוע הפעולה.

תוצאות הניתוח הסטטיסטי יוצגו תמיד בחלון נפרד, הנקרא חלון ה-output (פלט). חלון זה מורכב משני חלקים עיקריים. החלק הימני והעיקרי של החלון מכיל את התוצאות שהתקבלו מהניתוח. החלק השמאלי של החלון מכיל את רשימת הפקודות שמכיל חלון ה-output, ואת אוסף הטבלאות שנוצר על-ידי כל פקודה. כרגע, הרצנו את הפקודה Frequencies בלבד. אולם, ניתוח סטטיסטי מלא כולל לרוב הרבה מאוד פקודות שמריצים אחת לאחר השנייה, ויוצר הרבה מאוד טבלאות נתונים שונות בחלון ה-output. וחלקו השמאלי של חלון ה-output יעזור מאוד במקרה זה להתמצא בקובץ הפלט הארוך שיתקבל.

*Output1 [Document1] - PASW Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Output

- Log
- Precedences
 - Title
 - Notes
 - Active Dataset
 - Statistics
 - Predominant religion

→ **Frequencies**

[DataSet1] C:\Documents and Settings\student\Desktop\נתונים - עם שינוי נתונים\World95.sav

Statistics

Predominant religion

N	Valid	108
	Missing	1

Predominant religion

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Animist	4	3.7	3.7	3.7
	Buddhist	7	6.4	6.5	10.2
	Catholic	41	37.6	38.0	48.1
	Hindu	1	.9	.9	49.1
	Jewish	1	.9	.9	50.0
	Muslim	27	24.8	25.0	75.0
	Orthodox	8	7.3	7.4	82.4
	Protestant	16	14.7	14.8	97.2
	Taoist	2	1.8	1.9	99.1
	Tribal	1	.9	.9	100.0
	Total	108	99.1	100.0	
Missing	missing	1	.9		
	Total	109	100.0		

PASW Statistics Processor is ready

Start food C:\Document... C:\Document... untitled - Paint Microsoft Wo... *Output1 [World95.sav ... 13:01

כעת, נתבונן בפלט שהתקבל. נוצרו שני טבלאות. הטבלה הראשונה נראית כך:

Statistics

Predominant religion

N	Valid	108
	Missing	1

מעל לטבלה כתובה התווית (label) של המשתנה שעליו התבצע הניתוח: Predominant religion. בתוך הטבלה כתובות מספר התצפיות שכולל המשתנה (N). את התצפיות ניתן לחלק לשני סוגים: תצפיות "תקפות" הכוללות ערכים במשתנה (valid), ותצפיות שיש עבורן ערכים חסרים במשתנה זה (missing). בדוגמה שלעיל, ניתן לראות כי קיימות 108 תצפיות תקפות, ותצפית אחת שעבורה קיים ערך חסר במשתנה זה. לאחר מכן, תופיע טבלת השכיחויות שאותה ביקשנו:

Predominant religion

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Animist	4	3.7	3.7	3.7
	Buddhist	7	6.4	6.5	10.2
	Catholic	41	37.6	38.0	48.1
	Hindu	1	.9	.9	49.1
	Jewish	1	.9	.9	50.0
	Muslim	27	24.8	25.0	75.0
	Orthodox	8	7.3	7.4	82.4
	Protstnt	16	14.7	14.8	97.2
	Taoist	2	1.8	1.9	99.1
	Tribal	1	.9	.9	100.0
	Total	108	99.1	100.0	
Missing	missing	1	.9		
Total		109	100.0		

מעל לטבלה תופיע, כמו בטבלה הקודמת, התווית של המשתנה שעליו נערך הניתוח:

.Predominant religion

העמודה השמאלית ביותר כוללת את כל הערכים הקיימים במשתנה.

העמודה השנייה משמאל (Frequency) מכילה את השכיחות של כל אחד מהערכים בקובץ הנתונים. כך, למשל, ניתן לראות כי ב-27 מדינות הדת השלטת היא האיסלאם, וב-7 מדינות הדת השלטת היא בודהיזם. כמו כן ניתן לראות כי קיים ערך חסר אחד (Missing), וכי בסך הכל היו

במדגם 109 תצפיות (Total)

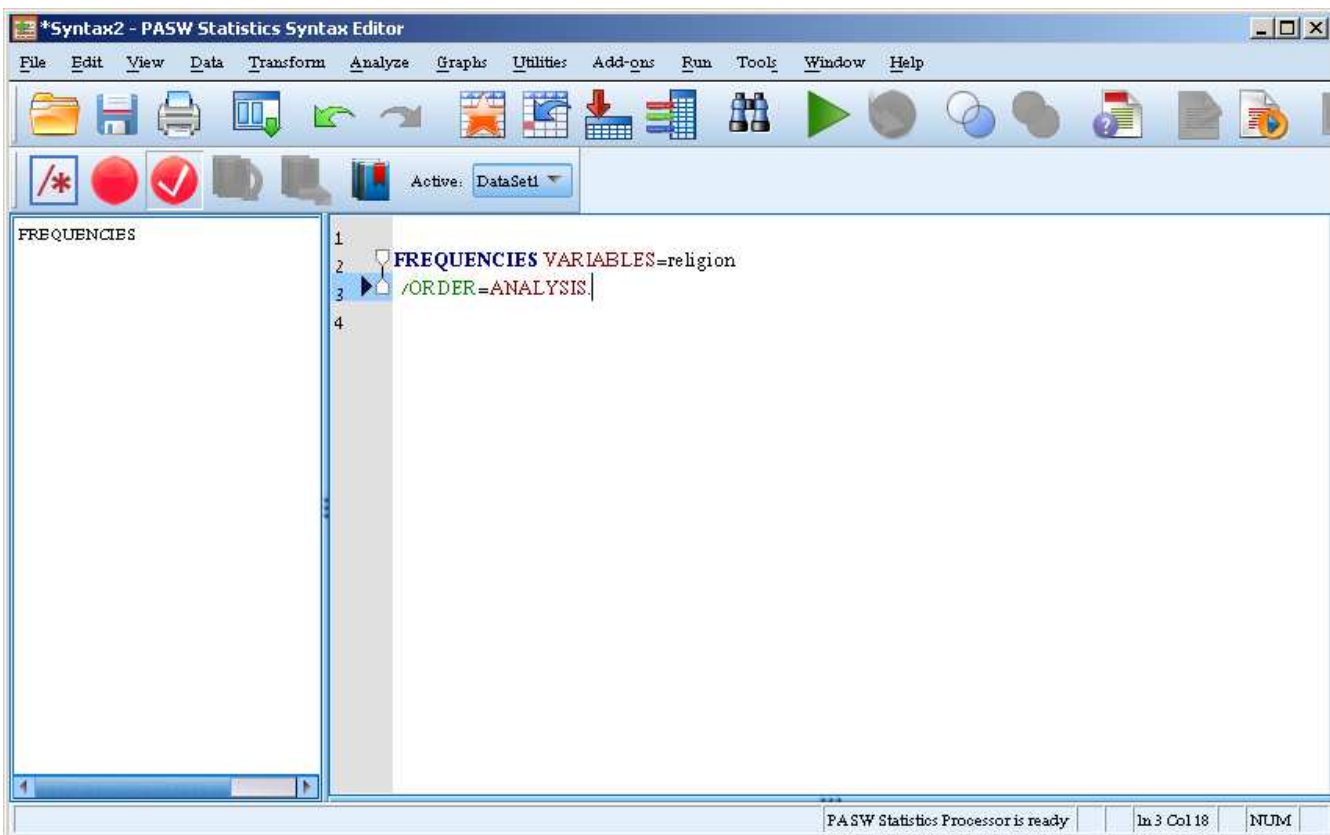
העמודה השלישית משמאל (Percent) מציגה את השכיחות היחסית במדגם. השכיחות היחסית היא היחס בין השכיחות של כל ערך, לבין מספר התצפיות הכללי (במילים אחרות, $f(x)/N$). כך, למשל, ניתן לראות כי ב-24.8% מהמדינות בקובץ הדת השלטת היא האיסלאם (החישוב שהתבצע הוא $27/109$), וב-6.4% מהמדינות בקובץ הדת השלטת היא בודהיזם.

העמודה השנייה מימין (Valid percent) מציגה את השכיחות היחסית, אולם לא ביחס לכל התצפיות במדגם, אלא ביחס למספר הערכים התקפים בלבד. בדוגמה זו, קיימות 109 תצפיות בסך-הכל. אולם, רק 108 מהן הן תצפיות "תקפות" שאינן מכילות ערכים חסרים. עמודה זו מחשבת את השכיחות היחסית ביחס למספר הערכים התקפים בלבד. כך למשל, השכיחות היחסית מתוך הערכים התקפים של דת האיסלאם היא 6.5% (החישוב הוא $27/108$, ולא $27/109$ כמו בעמודה הקודמת).

לבסוף, העמודה הימנית ביותר (Cumulative percent) מציינת את השכיחות היחסית המצטברת. שכיחות יחסית מצטברת מוגדרת כיחס בין השכיחות המצטברת $F(x)$, לבין סך כל המקרים התקפים במדגם. שימו לב שלשכיחות מצטברת יש משמעות רק כאשר סולם המדידה הוא סדר לפחות, מפני שרק החל מסולם סדר יש משמעות לסידור הקטגוריות. לכן, עמודה זו הינה חסרת משמעות עבור המשתנה religion, שנמדד בסולם שמי.

*** נסו להפיק טבלת שכיוויות, כפי שעשיתם קודם, עבור המשתנה populatn. מה המשמעות של השכיחות המצטברת היחסית במשתנה זה?

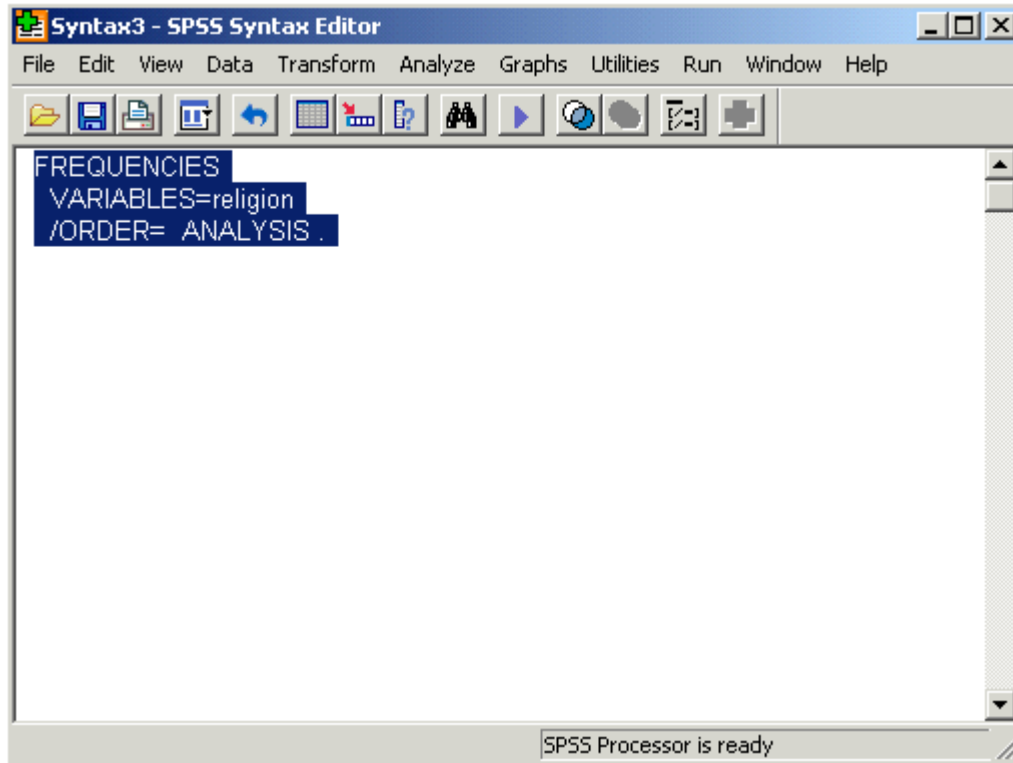
כעת, נחזור ונבחן מה עושה הכפתור Paste. סגרו את חלון ה-output באמצעות הלחצן X שבחלקו הימני העליון. המחשב ישאל אם ברצונכם לשמור את קובץ התוצאות. באופן עקרוני, ניתן לשמור קבצי תוצאות על המחשב, ולפתוח אותם בעתיד. כרגע, אין צורך לעשות זאת. חזרו את חלון ה-Data view, והיכנסו שוב לחלונית Frequencies (באמצעות בחירה בתפריטים Analyze -> Descriptive statistics -> Frequencies). בחרו שוב במשתנה religion לניתוח (אם שינייתם זאת מאז הניתוח הקודם). עתה, אל תלחצו על כפתור ה-OK אלא על הכפתור Paste. ייפתח חלון syntax שייראה כך:



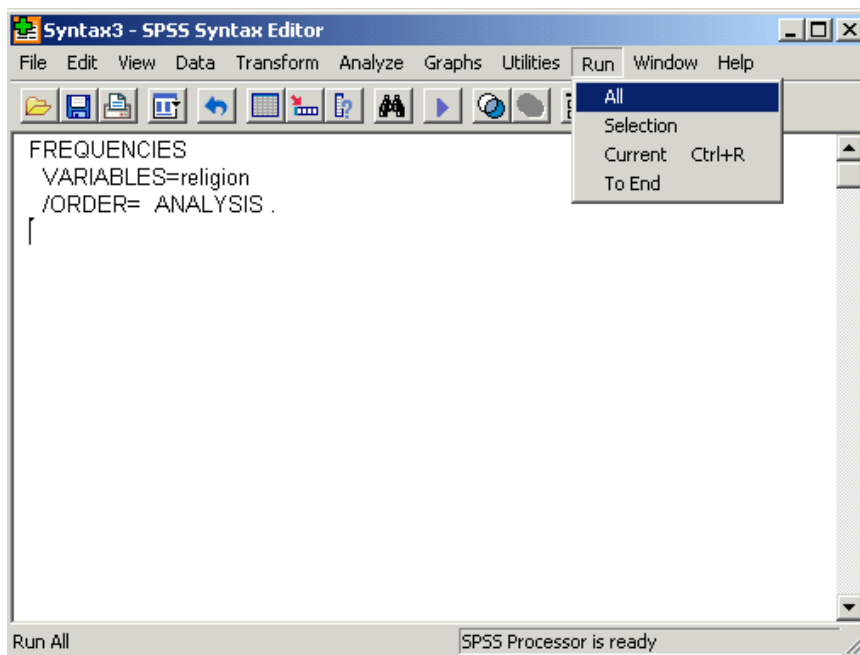
חלון ה-syntax ("חלון הפקודות") מאפשר לכתוב בתוכו פקודות המגדירות ניתוחים שונים, או להדביק אליו פקודות שנוצרות באופן אוטומטי, באמצעות ההגדרות שנבחרו בתפריטים של התוכנה.

הפקודה שכתובה כרגע בחלון מתאימה לאותו הניתוח שהגדרנו קודם באמצעות התפריט Frequencies. היתרון הגדול של חלון ה-syntax הוא שניתן לשמור את הפקודות בקובץ, ולהריץ אותן לאחר מכן (עם אותו קובץ נתונים, או אפילו עם קובץ נתונים אחר). הדרך לשמור את הפקודות היא להיכנס לתפריט של חלון ה-syntax, לבחור ב-File ואז Save As..., ולשמור את הקובץ באופן הרגיל. כדי להריץ פקודות מתוך חלון ה-syntax, ניתן לסמן את הפקודה הרצויה

(כרגע קיימת רק פקודה אחת, אבל בקובץ של ניתוח מלא יהיה מספר רב של פקודות), ולחוץ על מקש החץ הכחול.



לחלופין, אפשר לבחור בתפריט Run, ומשם לבחור האם ברצונכם להריץ את כל הפקודות (All) או רק את הפקודות המסומנות (Selection):



באופן כללי, בקורס זה תידרשו לעבוד באמצעות תפריטי התוכנה בלבד, ולא באמצעות כתיבת פקודות. עם זאת, מומלץ להכיר את האפשרות לעבוד באמצעות חלון ה-syntax במידת הצורך.

מדדי מרכז ופיזור

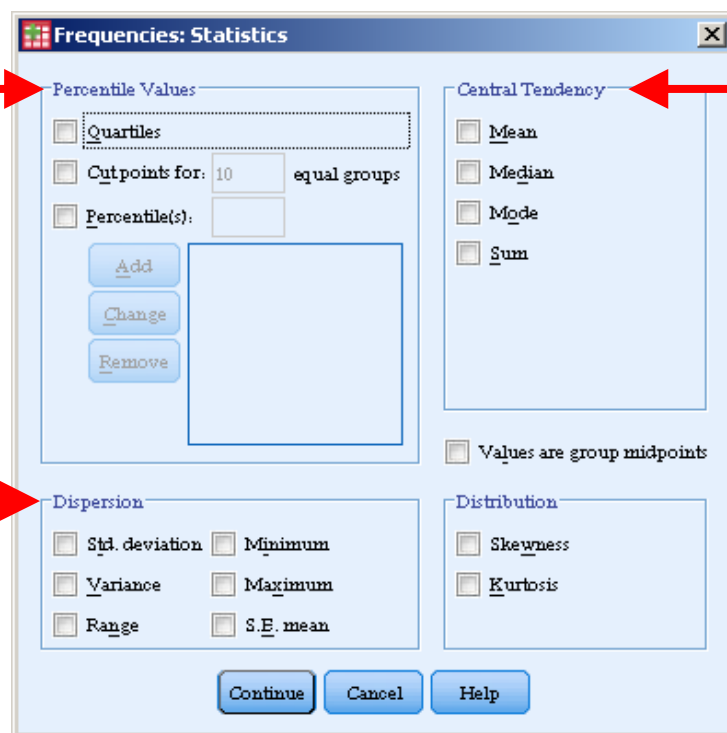
מהו ממוצע התושבים במדינה? מהו החציון? וסטיית התקן? למדדים אלה יש אין משמעות עבור המשתנה religion, ולכן נדגים את הניתוח על המשתנה populatn (גודל האוכלוסיה באלפים).

בתפריט הראשי של SPSS, בחלון ה-Data View, נבחר בתפריט Analyze. תפריט זה ישמש עבור כל הניתוחים הסטטיסטיים שנרצה לבצע (כעת ובעתיד). בהמשך, נבחר בתת-התפריט Descriptive Statistics, ואז נבחר באפשרות Frequencies (שכיחויות). אם נלחץ על OK, נקבל טבלת שכיחויות בלבד. כדי לקבל את המדדים המבוקשים, נלחץ על הכפתור Statistics שנמצא בצד ימין של חלון ה-Frequencies. יפתח חלון, שבו ניתן לסמן את המדדים הסטטיסטיים שברצונכם לחשב:

מדדים למיקום יחסי

מדדי מרכז

מדדי פיזור



כדי לבצע חישוב של מדד מסויים, יש לסמן V בתוך הריבוע הלבן שליד שם המדד. המדדים העיקריים (והרלוונטיים בשלב זה) שניתן להגדיר בחלון זה הם:

מדדים למיקום מרכזי:

ממוצע Mean, חציון Median, שכיח Mode, וסכום התצפיות Sum.

מדדי פיזור:

סטיית תקן Std. deviation, שונות Variance, טווח Range, הערך הקטן ביותר במדגם Minimum, הערך הגדול ביותר במדגם Maximum, טעות תקן S.E. mean (מדד זה יילמד בהרחבה בסמסטר ב').

מאונים:

רבעונים Quartiles, ומאונים ספציפיים אחרים. כדי להגדיר מאונים ספציפיים (למשל, מהם הערכים הנמצאים במאון ה-17 ובמאון ה-83), יש לסמן V בריבוע הלבן שליד המילה Percentiles. ייפתח ריבוע לבן קטן שבתוכו ניתן לרשום את המאון המבוקש (למשל, 17). לאחר מכן, יש ללחוץ על הכפתור Add, ורשום מאונים נוספים שרוצים לחשב (אם יש כאלה):

לאחר שהגדרנו את כל המדדים הסטטיסטיים שברצוננו לחשב, יש ללחוץ על הכפתור Continue. לחיצה על כפתור זה תחזיר אותנו לחלון ה-Frequencies, ושם יש ללחוץ OK כדי לקבל את תוצאות הניתוח.

כעת, נחשב עבור המשתנה population את המדדים הסטטיסטיים הבאים: ממוצע, חציון, שכיח, שונות, סטיית תקן, טווח, מינימום, מקסימום, רבעונים, מאון 10, ומאון 90. כדי לעשות זאת, סמנו V בתוך הריבוע הלבן שליד המדדים הרלוונטיים, ולחצו על הכפתור Continue. מאחר ואין ברצוננו להציג טבלת שכיחויות עבור המשתנה, בטלו את הסימן V ליד הכיתוב "Display frequency tables" שמתחת לרשימת המשתנים, בחלון ה-Frequency.

בחלון ה-output יתקבל הפלט הבא:

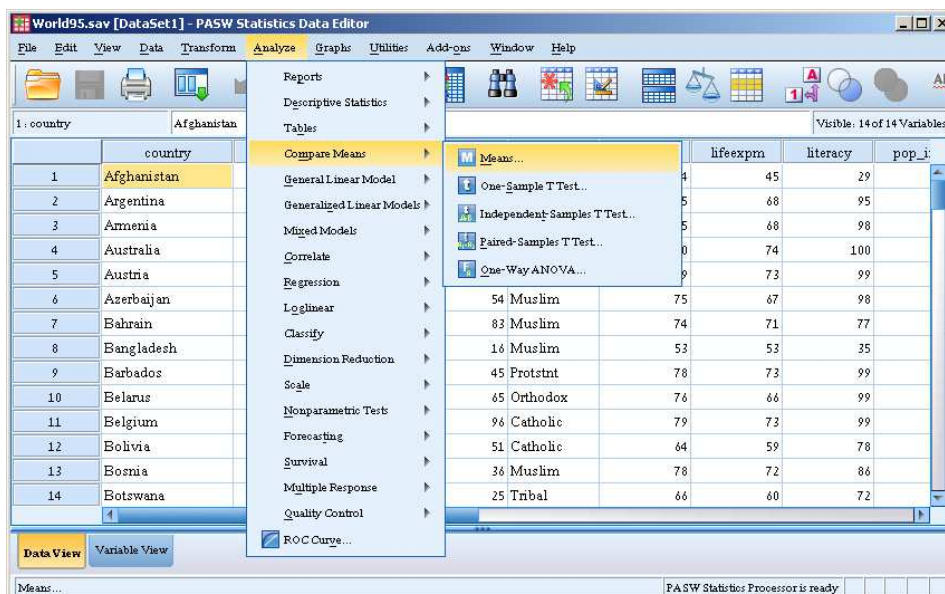
Population in thousands		
N	Valid	109
	Missing	0
Mean		47723.88
Median		10400.00
Mode		2900 ^a
Std. Deviation		146726.364
Variance		21528625814.
Range		143
Minimum		256
Maximum		1205200
Percentiles	10	2700.00
	25	5000.00
	50	10400.00
	75	37100.00
	90	91800.00

a. Multiple modes exist. The smallest value is shown

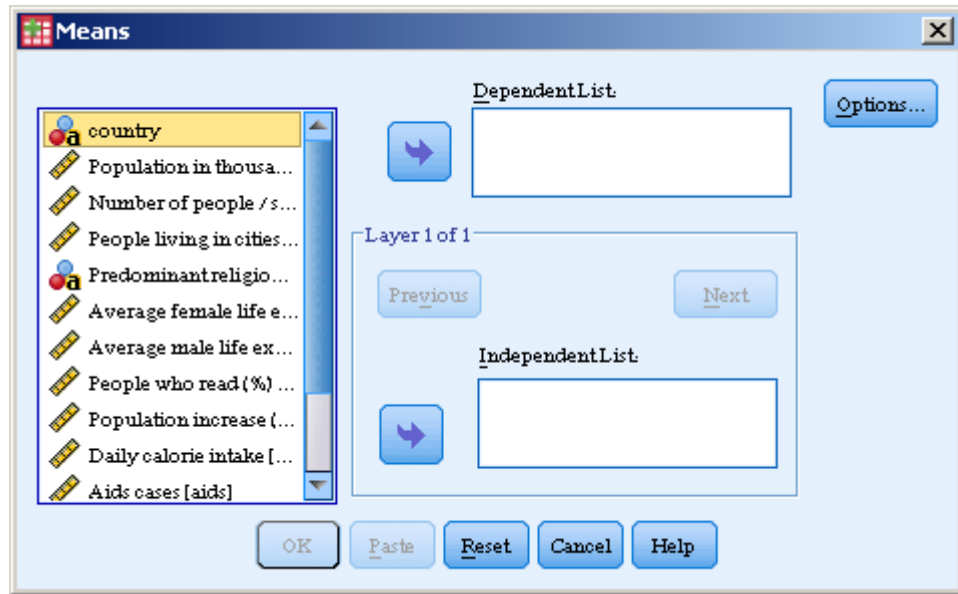
הטבלה מכילה את כל המדדים התיאוריים שביקשנו, בזה אחר זה. שם המדד כתוב בעמודה השמאלית, והערך כתוב בעמודה הימנית. בנוסף, בטבלה מופיעים באופן אוטומטי נתונים אודות מספר התצפיות התקפות במדגם (N Valid), ומספר הערכים החסרים (N Missing). מאחר ובמשתנה זה (populatn) אין ערכים חסרים, מספר הערכים החסרים המופיעים בטבלה זו הוא 0. שימו לב לשני דברים נוספים בטבלה. ראשית, לגבי השכיח: ליד הערך 2900 כתובה האות a, המפנה אתכם אל הערה בתחתית הטבלה. בהערה מצויין כי קיימים בהתפלגות מספר שכיחים, והשכיח המוצג הינו הקטן ביותר. באופן עקרוני, אין אפשרות להציג בטבלת המדדים הסטטיסטיים את כל השכיחים. הדרך לדעת מהם כל השכיחים היא לחזור ולהציג את טבלת השכיחויות המלאה, ולבדוק זאת ידנית. שנית, לגבי השונות: לעיתים, המספר שמופיע הוא $2E+010$. צורת הצגה זו של מספרים מוכרת לחלקכם מהמחשבון האישי. דרך כתיבה זו נועדה לקצר את הכתיבה של מספרים ארוכים מאוד. במקרה זה, $2E+010$ הוא קיצור של הביטוי $2 \cdot 10^{10}$, או 2 ואחריו 10 אפסים. באופן דומה, הביטוי $3.5E-012$ (למשל) הוא קיצור של הביטוי $3.5/10^{12}$, או 3.5 ולפניו (משמאלו) 11 אפסים.

ממוצעים מותנים

עד עתה, למדתם כיצד להציג נתונים סטטיסטיים תיאוריים אודות משתנה אחד בלבד (לדוגמה, religion או populatn). בסעיף זה תלמדו כיצד להציג נתונים אודות השילוב של שני משתנים יחד. למשל: עכשיו ברצוננו לדעת כיצד מתפלג המשתנה populatn (גודל האוכלוסיה), בתוך הרמות השונות של המשתנה religion (דת שלטת). כלומר, מהו גודל האוכלוסיה הממוצעת במדינות המוסלמיות, מהו גודל האוכלוסיה הממוצעת במדינות הבודהיסטיות וכן הלאה. מובן, שאין הכרח לחשב ממוצע בלבד, אלא ניתן לחשב באופן זה כל מדד סטטיסטי אחר (לדוגמה, מהי סטיית התקן של גודל האוכלוסיה לפי הדת השלטת, או החציון, וכן הלאה). כדי לעשות זאת, יש לגשת לחלון הנתונים, לבחור מהתפריט הראשי את האפשרות Analyze, ואחר-כך Compare Means (השווה ממוצעים) ואז Means...

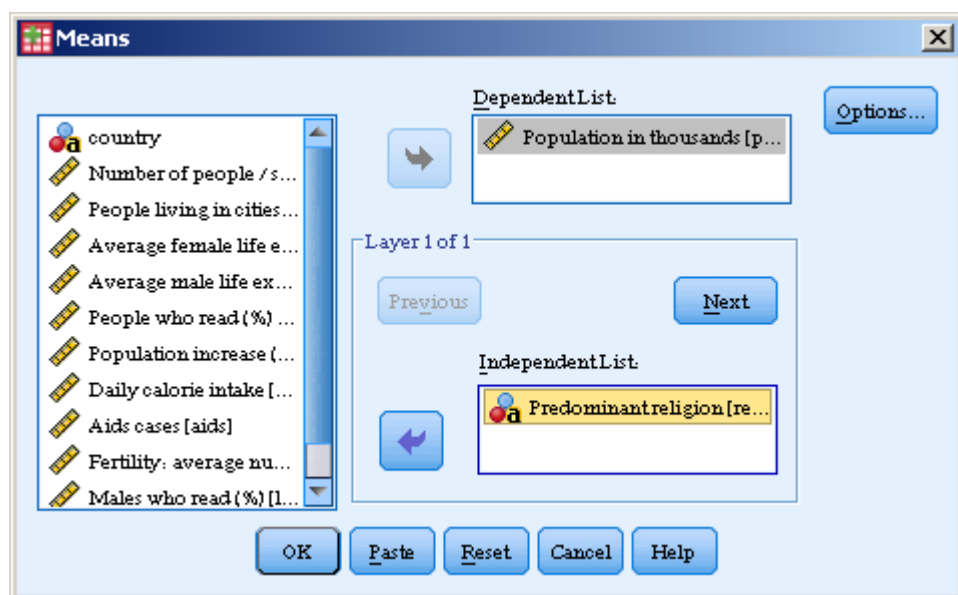


ייפתח החלון הבא :



בצידו השמאלי של החלון תופיע רשימת התוויות של המשתנים בקובץ הנתונים. מימין, יופיעו שני חלונות, Dependent List (רשימת משתנים תלויים) ו- Independent List (רשימת משתנים בלתי-תלויים). בהקשר זה, משתנים תלויים הם המשתנים שאותם רוצים להשוות, ואותם רוצים לחשב. בדוגמה שלעיל, המשתנה התלוי הוא גודל האוכלוסיה, מפני שרצינו להשוות משתנה בין דתות שונות, ולחשב עבורו ממוצעים. לכן, יש להעביר משתנה זה לחלון ה- Dependent List (באמצעות החץ השחור הקטן, כפי שעשיתם בחלון ה-Frequencies). לעומת זאת, משתנים בלתי-תלויים הם משתנים שעל-פיהם מאורגנת ההשוואה. בהקשר זה, רצינו להשוות בין הדתות השונות, ולכן המשתנה הבלתי-תלוי הוא religion. לכן, יש להעביר משתנה זה לחלון ה-

: Independent List



אם תלחצו OK, תקבלו טבלה המציגה את הממוצעים וסטיות התקן של המשתנה populatn, בכל אחת מהרמות של המשתנה religion. כדי לבחור במדדים סטטיסטיים אחרים, ניתן ללחוץ על כפתור ה- Options... ושם לבחור במדדים הרצויים.

הפלט שיתקבל מורכב משתי טבלאות. הטבלה הראשונה נראית כך :

Case Processing Summary						
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Population in thousands *	108	99.1%	1	.9%	109	100.0%
Predominant religion						

בטבלה זו מוצגים מספר התצפיות הכללי (Total), מספר התצפיות התקפות (Included) ומספר הערכים החסרים (Excluded), עבור הניתוח שהתבצע. במקרה זה, ניתן לראות כי תצפית אחת הוצאה מהניתוח בשל ערכים חסרים. לאחר מכן, תוצג הטבלה הבאה :

Report

Population in thousands

Predominant religion	Mean	N	Std. Deviation
Animist	13950.00	4	11396.637
Buddhist	45406.29	7	43389.718
Catholic	21974.39	41	29583.816
Hindu	911600.00	1	.
Jewish	5400.00	1	.
Muslim	36533.59	27	49483.047
Orthodox	32900.00	8	49525.693
Protstnt	31708.94	16	65594.845
Taoist	604050.00	2	850154.483
Tribal	1359.00	1	.
Total	47759.29	108	147409.939

טבלה זו מציגה ליד כל קטגוריה של המשתנה religion, את הממוצע, מספר התצפיות, וסטיות התקן המתאימים לה במשתנה populatn. לדוגמה, ממוצע האוכלוסיה במדינות המוסלמיות הוא 36,533.59 (באלפים, כלומר 36 מליון בערך). המשמעות של מספר זה היא שקיימות 27 מדינות מוסלמיות. לכל אחת מהן יש ערך אחר במשתנה populatn, והממוצע של כל הערכים הללו הוא 36,533.59. סטיית התקן של הערכים הללו (המבטאת את מידת הפיזור בין המדינות המוסלמיות השונות) היא 49,483.047.

תרגיל הגשה מספר 3

1. השתמשו בקובץ World95.sav
2. באמצעות הניתוחים הסטטיסטיים שלמדתם, צרו קובץ תוצאות (output) שיכיל טבלאות, שבהן נתונים העונים על השאלות הבאות. הדפיסו את קובץ התוצאות (מתוך חלון ה-output ניתן לבחור בתפריט File ואז Print). לבסוף, צרפו דף נוסף, ובו התשובות המספריות בלבד לשאלות שלהלן (לדוגמה: א – 65, ב – 537, וכן הלאה).
 - א. מהי תוחלת החיים הממוצעת של גברים ושל נשים?
 - ב. מהי סטיית התקן של תוחלות החיים של גברים ושל נשים?
 - ג. מה היה מספר חולי האיידס בעולם בשנת 1995?
 - ד. מהו הטווח הבין רבעוני של צריכת הקלוריות היומית?
 - ה. מהו מספר הילדים הממוצע במשפחה בודהיסטית? ובמשפחה קתולית?
 - ו. כמה ערכים חסרים ישנם במשתנה aids?
 - ז. הציגו טבלה, המשווה את אחוז הגברים היודעים קרוא וכתוב לאחוז הנשים היודעות קרוא וכתוב, לפי הדת השלטת במדינה. מטרת הטבלה היא לזהות אי-שוויון בין המינים, לפי דת (למשל: מה אחוז הגברים לעומת הנשים היודעות קרוא וכתוב במדינות בודהיסטיות, לעומת מוסלמיות, קתוליות, וכו'). שימו לב שבשיעור לא למדתם באופן ישיר כיצד לעשות זאת, אבל ניתן להציג את ההשוואה המבוקשת באמצעות "משחק" באחד משני סוגי הניתוחים שנלמדו.
 - ח. המשיכו לשחק עם הקובץ כרצונכם. הוסיפו שלוש שאלות מעניינות משלכם ובדקו אותן. כתבו את השאלות והתשובות להן.
 - ט. שימרו את הקובץ ושילחו אות למייל: stat.bgu@gmail.com בצירוף קובץ וורד (word) ובו התוצאות לשאלות בהן התשובה היא תוצאה מספרית. אנא ציינו את שימכם ומספר התרגיל בנושא המייל.

שיעור מספר 4

שיעור מס' 4: מתאמים ורגרסיה ב-SPSS

בשיעור זה תלמדו לחשב מתאם קרמר, פירסון וספירמן, וכן לבצע רגרסיה פשוטה ומרובה ב-SPSS. השיעור יתבסס על הקובץ Employee.

להלן המשתנים בקובץ זה:

- id – מספר סידורי של העובד
- gender – מגדר (f=נקבה, m=זכר)
- bdate – תאריך לידה
- educ – מספר שנות השכלה
- jobcat – קטגוריה תעסוקתית (1=פקיד/ה, 2=מתכנת/ת, 3=מנהל/ת)
- salary – שכר נוכחי בדולרים
- salbegin – שכר התחלתי בדולרים
- jobtime – משך הזמן שבו העובד מועסק בחברה (בחודשים)
- prevexp – ניסיון קודם (בחודשים)
- minority – השתייכות לקבוצת מיעוטים באוכלוסיה (0=לא, 1=כן, 9=לא ידוע)

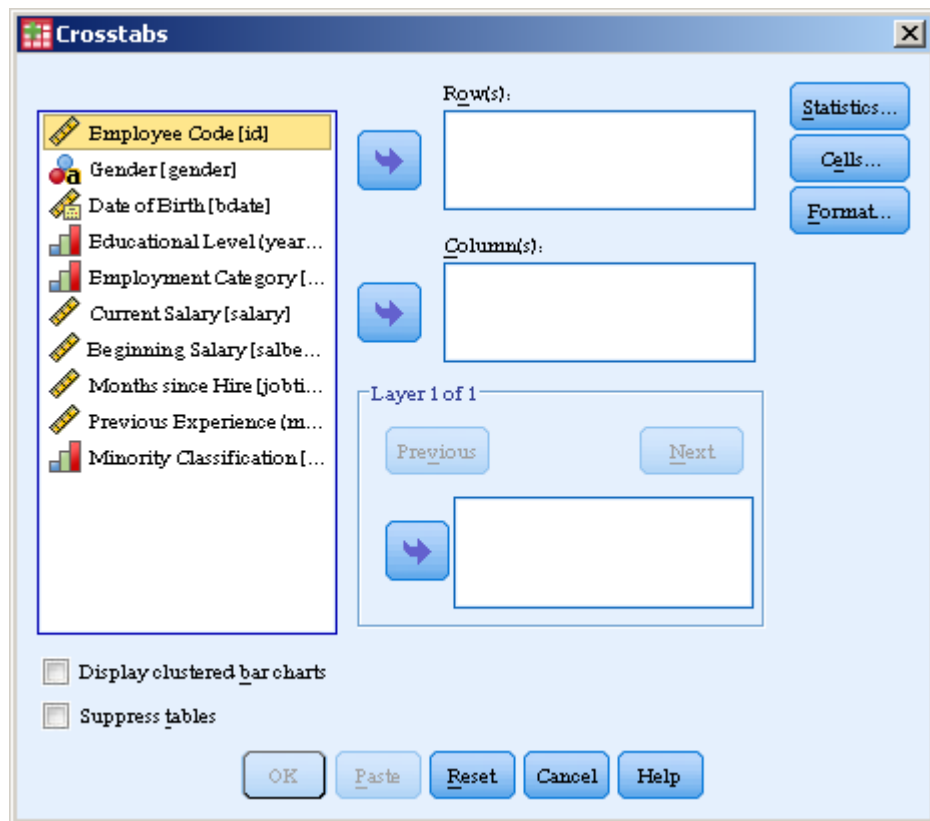
מתאם קרמר

מתאם קרמר הינו מדד לקשר בין שני משתנים, שהנמוך ביניהם נמדד בסולם שמי. למשל, בקובץ זה, ברצוננו לבחון את עוצמת הקשר בין מגדר לבין קטגוריה תעסוקתית. כדי לעשות זאת, יש לבחור בתפריט הראשי של SPSS באפשרות Analyze, ואז Descriptive statistics ו-Crosstabs.

The screenshot shows the SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the 'Crosstabs...' option is selected. The data table below shows the first 15 rows of the dataset.

id	gender	salary	salbegin	jobtime	prevexp
1	m	\$57,000	\$27,000	98	
2	m	\$40,200	\$18,750	98	
3	f	\$21,450	\$12,000	98	
4	f	\$21,900	\$13,200	98	
5	m	\$45,000	\$21,000	77	
6	m	\$32,100	\$13,500	98	
7	f	\$36,000	\$18,750	98	
8	f	\$21,900	\$9,750	98	
9	f	\$27,900	\$12,750	98	
10	f	\$24,000	\$12,000	88	
11	f	\$21,000	\$16,500	98	
12	m	\$28,350	\$12,000	98	
13	m	\$27,750	\$14,250	98	
14	f	\$35,100	\$16,800	98	
15	1				

ייפתח החלון הבא :



מתאם קרמר מבוסס על טבלת שכיחות משותפת של שני משתנים. טבלה זו נראית למשל כך :

	נקבה	זכר
פקיד/ה		
מתכנת/ת		
מנהל/ת		

בטבלה שלעיל המשתנה "מין" (Gender) מגדיר את העמודות, והמשתנה "קטגוריה תעסוקתית" (Jobcat) מגדיר את השורות. כדי ליצור טבלה כזו ב-SPSS, יש להכניס את המשתנה Gender לחלון הלבן שמעליו כתוב Rows (שורות), ואת המשתנה jobcat לחלון שמעליו כתוב Columns (עמודות). ניתן להכניס את המשתנים לחלוניות הלבנות באמצעות החץ, כפי שעשיתם בשיעורים הקודמים.

כעת, אם תלחצו על OK (בחלקו התחתון של החלון), יתקבל בחלון ה-output שתי טבלאות. הטבלה הראשונה מציגה את מספר התצפיות שנכנסו לניתוח, ואת מספר התצפיות החסרות.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Employment Category	474	100.0%	0	.0%	474	100.0%

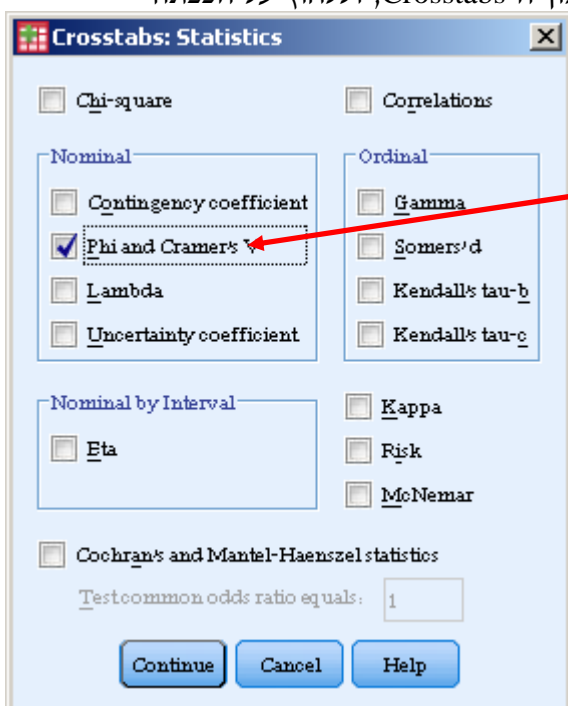
במקרה זה, אין ערכים חסרים, וכל הערכים (474 תצפיות, 100%) נכנסו לניתוח הסטטיסטי. הטבלה השנייה הינה טבלת שכיחויות משותפת:

Gender * Employment Category Crosstabulation

Count

		Employment Category			Total
		Clerical	Custodial	Manager	
Gender	Female	208	0	10	218
	Male	155	27	74	256
	Total	363	27	84	474

למעשה, זו טבלת ה-observed שעל-פיה ניתן לחשב את מתאם קרמר. כדי לחשב את ערכו של מתאם קרמר עצמו, יש לחזור לחלון ה-Crosstabs, וללחוץ על הכפתור Statistics שבחלקו התחתון של החלון. יפתח החלון הבא:



בחלון זה, יש לסמן את הריבוע הזה, וללחוץ על Continue (בצד ימין למעלה) ואז שוב OK. יתקבל הפלט הבא:

Symmetric Measures

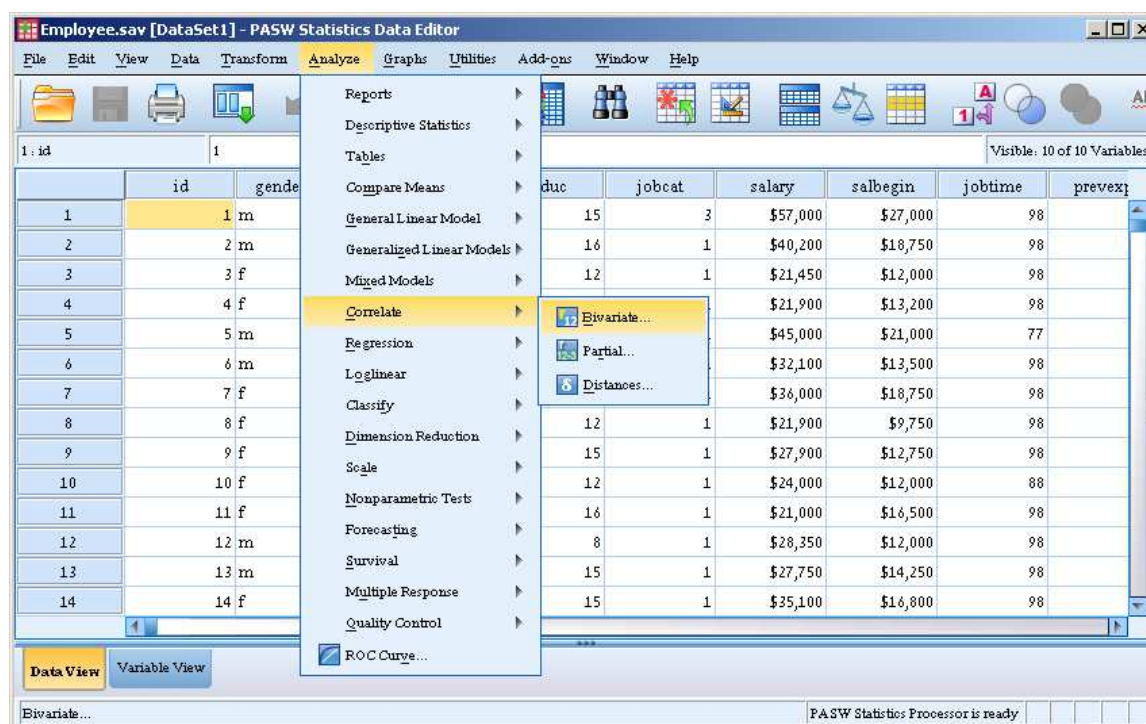
		Value	Approx. Sig.
Nominal by Nominal	Phi	.409	.000
	Cramer's V	.409	.000
N of Valid Cases		474	

בעמודה Value, נמצא הערך של מתאם קרמר ליד הכותרת Cramer's V. בדוגמה שלנו, ערכו של מתאם קרמר הוא 0.409. הטור הימני בטבלה (Approx. Sig.) מתייחס למובהקות המתאם. בסמסטר ב' תלמדו רבות על מובהקות, ואז גם תהיה לנתונים בטור זה משמעות.

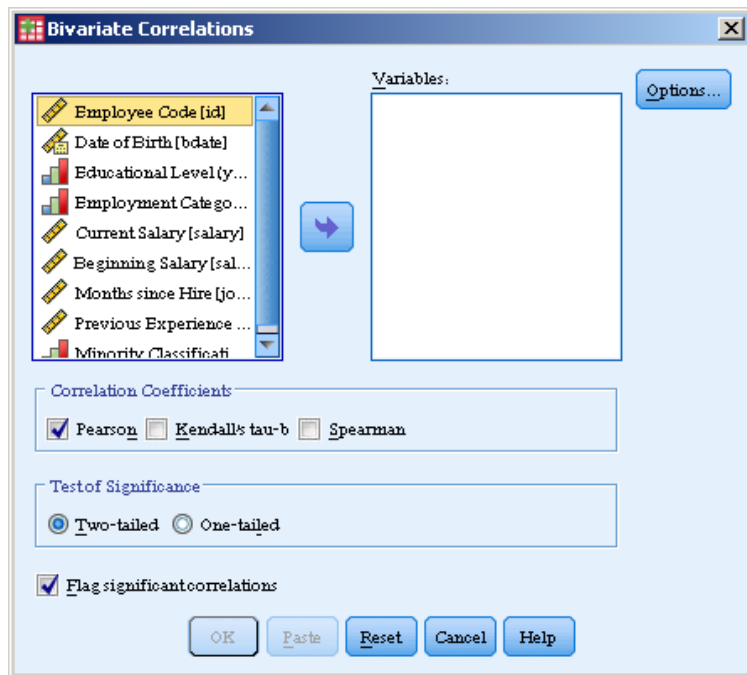
מתאם ספירמן ומתאם פירסון

מתאם ספירמן הוא מדד לקשר בין שני משתנים, שהנמוך מביניהם נמדד בסולם סדר. מתאם פירסון הוא מדד לקשר בין שני משתנים בסולם רווח או מנה. ב-SPSS, הדרך לחשב מתאמים אלה דומה מאוד, ונעשית דרך אותו חלון. למשל, בקובץ זה, ניח כי ברצוננו לחשב את המתאם בין השכר ההתחלתי לבין השכר הנוכחי. למרות ששני המשתנים הללו נמדדים בסולם מנה, נחשב תחילה את מתאם ספירמן (רק לצורך ההדגמה), ואז את מתאם פירסון.

בתפריט הראשי של התוכנה יש לבחור Analyze, ואז Correlate ו-Bivariate:

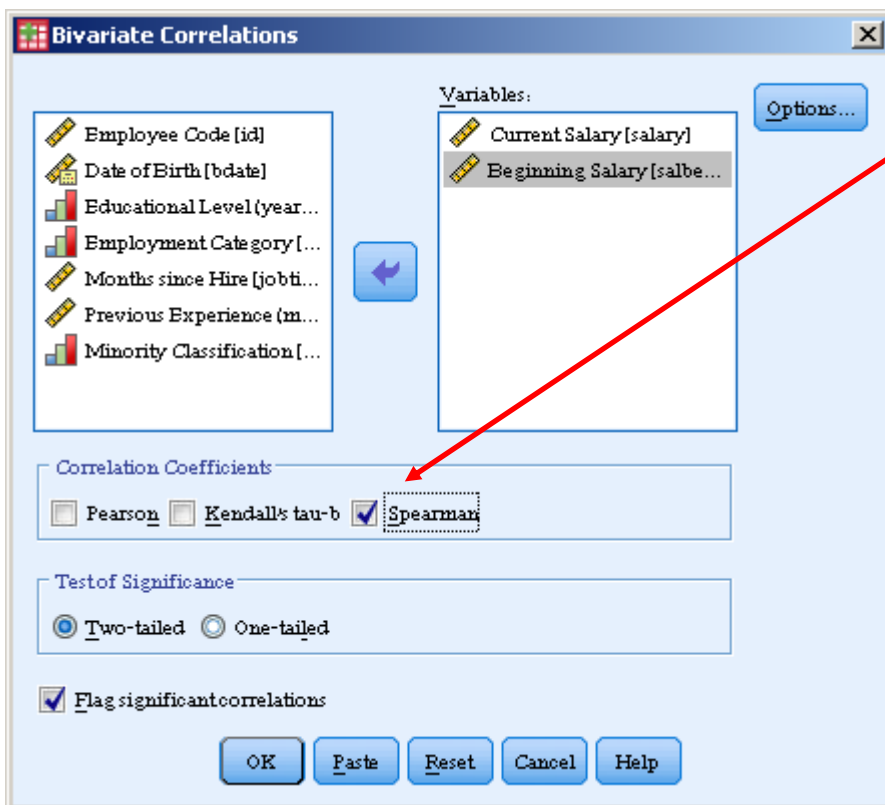


ייפתח החלון הבא:



אל חלקו הימני של החלון, תחת הכותרת Variables, ניתן להעביר את המשתנים שביניהם רוצים לחשב את המתאם. במקרה זה, נעביר את המשתנים salary ו-salbegin לחלון הימני באמצעות החץ.

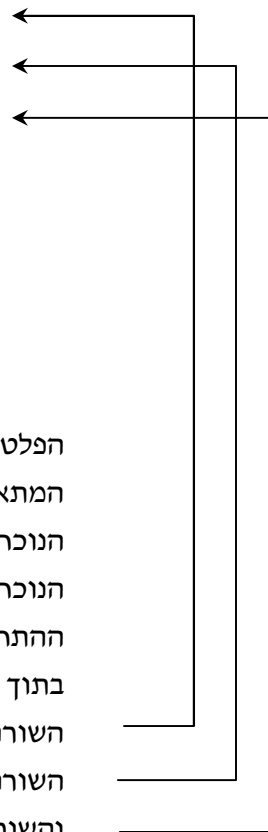
לאחר מכן, נבחר איזה סוג מתאם ברצוננו לחשב. ניתן לעשות זאת על-ידי סימון V ליד סוג המתאם הרצוי, בשורה שמתחת לשמות המשתנים. במקרה זה, נבטל את הסימון שליד מתאם פירסון, ונבחר במתאם ספיירמן.



כעת, יש לחוץ על OK. בחלון ה-Output, יתקבל הפלט הבא:

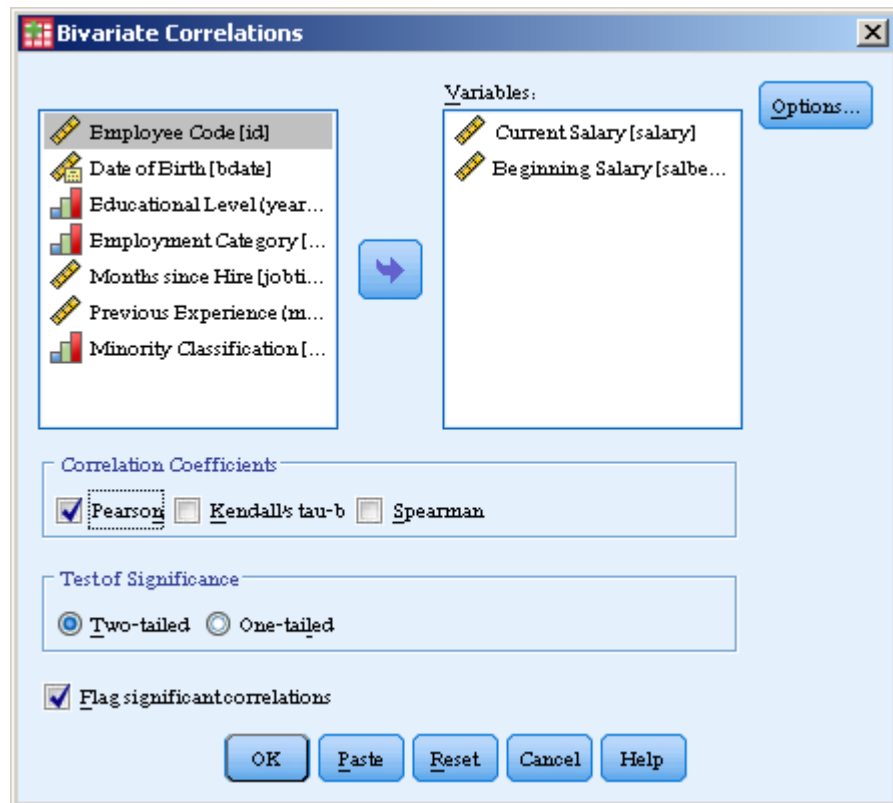
Correlations

			Current Salary	Beginning Salary
Spearman's rho	Current Salary	Correlation Coefficient	1.000	.824**
		Sig. (2-tailed)	.	.000
		N	474	474
	Beginning Salary	Correlation Coefficient	.824**	1.000
		Sig. (2-tailed)	.000	.
		N	474	474



** . Correlation is significant at the 0.01 level (2-tailed).

הפלט הינו טבלת מתאמים. שמות המשתנים מופיעים הן בטורים והן בשורות. באופן זה, המתאם בין שני משתנים נמצא בתא שבו מצטלבים שני המשתנים. למשל, את המתאם בין השכר הנוכחי לשכר ההתחלתי נמצא בתא שבו השורה מציינת את השכר ההתחלתי, והטור את השכר הנוכחי. מתאם זה נמצא גם בתא שבו השורה מציינת את השכר הנוכחי, והטור את השכר ההתחלתי. במובן זה, טבלת המתאמים היא סימטרית. בתוך כל תא מופיעות 3 שורות. השורה הראשונה מציינת את ערכו של המתאם, השורה השנייה מציינת את מובהקות המתאם (גם כאן, זהו מושג שנדבר עליו בסמסטר ב' בלבד), והשורה השלישית מציינת את מספר התצפיות שעליהן מבוסס המתאם. לכן, בדוגמה שלנו, ערכו של מתאם ספירמן בין השכר ההתחלתי לבין השכר הנוכחי הוא 0.824 (הכוכביות קשורות גם הן למושג המובהקות, וניתן להתעלם מהן בשלב זה). באופן זה לגמרי, ניתן לחשב את מתאם פירסון. במקום לסמן V ליד האפשרות Spearman, יש לסמן V ליד האפשרות Pearson.



יש לחץ על OK. יתקבל הפלט הבא:

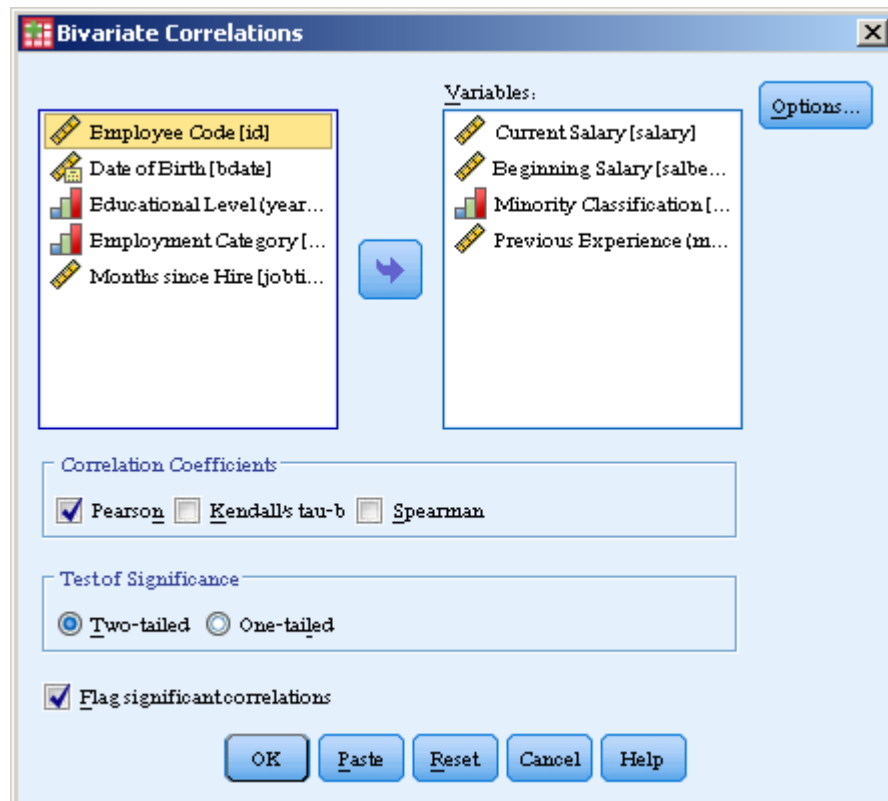
Correlations

		Current Salary	Beginning Salary
Current Salary	Pearson Correlation	1	.784**
	Sig. (2-tailed)		.000
	N	474	474
Beginning Salary	Pearson Correlation	.784**	1
	Sig. (2-tailed)	.000	
	N	474	474

** . Correlation is significant at the 0.01 level (2-tailed).

המבנה של פלט זה זהה לחלוטין למבנה הפלט של מתאם ספירמן, אלא שהמתאמים בתוכו הינם מתאמי פירסון.

יש לציין, כי טבלת מתאמים אינה חייבת לכלול שני משתנים בלבד. על-ידי הוספה של מספר משתנים גדול יותר לחלון ה-Variables, ניתן ליצור טבלת מתאמים בין יותר משני משתנים, לדוגמה:



במקרה זה, יתקבל הפלט הבא:

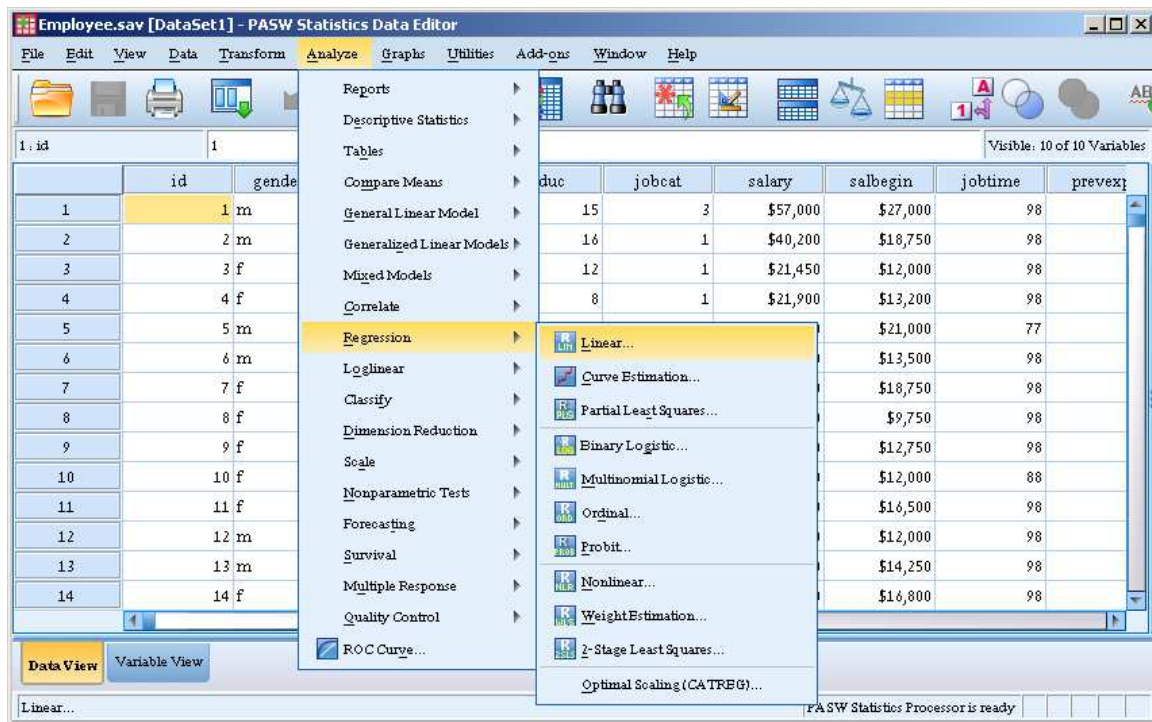
		Current Salary	Beginning Salary	Months since Hire	Previous Experience (months)
Current Salary	Pearson Correlation	1	.880**	.084	-.097*
	Sig. (2-tailed)		.000	.067	.034
	N	474	474	474	474
Beginning Salary	Pearson Correlation	.880**	1	-.020	.045
	Sig. (2-tailed)	.000		.668	.327
	N	474	474	474	474
Months since Hire	Pearson Correlation	.084	-.020	1	.003
	Sig. (2-tailed)	.067	.668		.948
	N	474	474	474	474
Previous Experience (months)	Pearson Correlation	-.097*	.045	.003	1
	Sig. (2-tailed)	.034	.327	.948	
	N	474	474	474	474

פלט זה מכיל את כל המתאמים, בין כל זוג משתנים מבין אלה שנבחרו. למשל, המתאם בין ניסיון קודם לבין השכר ההתחלתי הוא -0.097 , והמתאם בין מספר חודשי העסקה לבין ניסיון קודם הוא 0.003 .

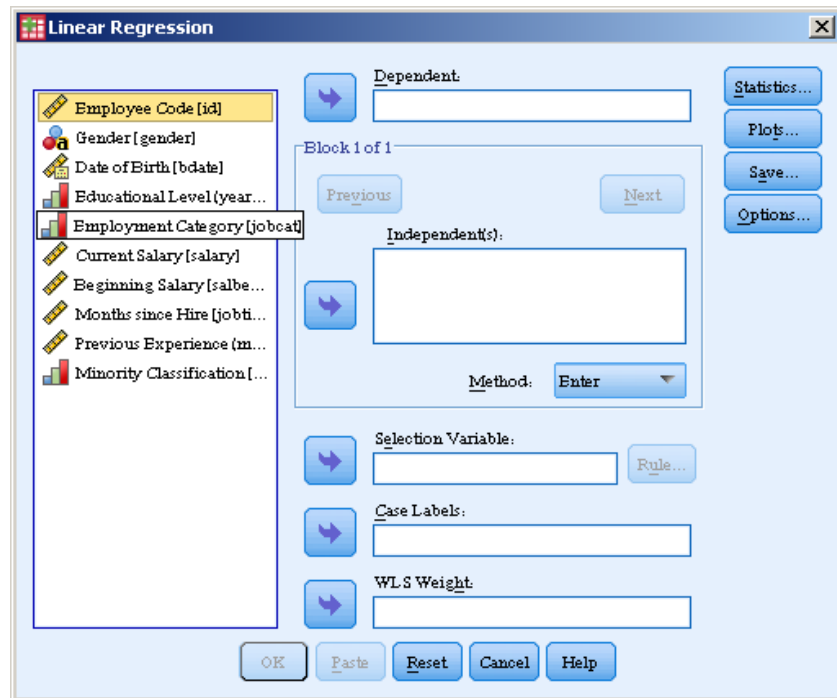
רגרסיה פשוטה

באמצעות רגרסיה פשוטה ניתן לנבא את ערכו של משתנה אחד (Y) באמצעות משתנה מנבא אחר (X). למשל, ברצוננו לדעת מהי נוסחת הרגרסיה לניבוי השכר הנוכחי (Y) מתוך השכר ההתחלתי (X).

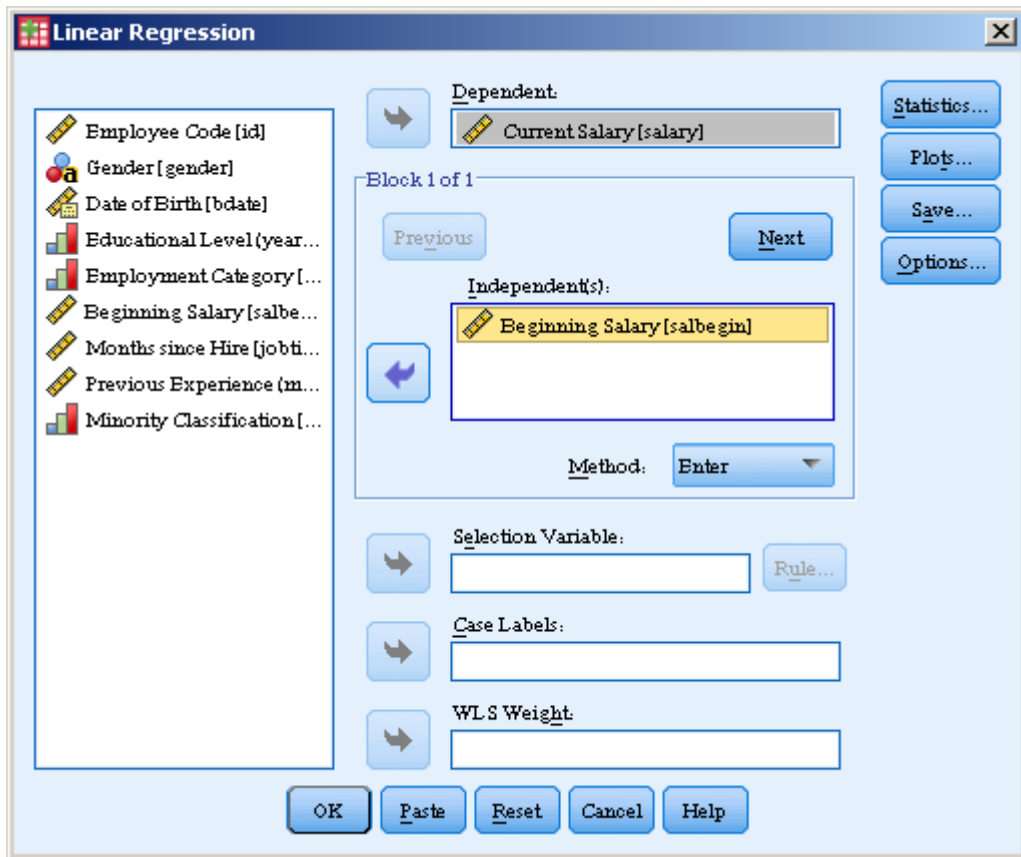
לשם כך, יש לבחור בתפריט הראשי Analyze, ואז Linear Regression.



יפתח החלון הבא:



בחלקו הימני עליון של החלון, מופיעה הכותרת Dependent. לחלונית הלבנה שמתחתיה יש להעביר את המשתנה שאותו רוצים לנבא (כלומר, Y, או במקרה שלנו – salary). מתחת, מופיעה הכותרת Independent(s). לחלונית הלבנה שמתחתיה יש להעביר את המשתנה שבאמצעותו אנו רוצים לנבא (כלומר, X, או במקרה שלנו – salbegin).



לבסוף, יש ללחוץ על OK. יתקבל הפלט הבא :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.784 ^a	.614	.614	\$10,622.750

a. Predictors: (Constant), Beginning Salary

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	84837153648.3	1	84837153648.3	751.817	.000 ^a
		46		46		
	Residual	53261809509.5	472	112842816.757		
		13				
	Total	138098963157.	473			
		859				

a. Predictors: (Constant), Beginning Salary

b. Dependent Variable: Current Salary

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8653.668	1058.134		8.178	.000
	Beginning Salary	1.495	.055	.784	27.419	.000

a. Dependent Variable: Current Salary

בשלב זה, רק הטבלה התחתונה (coefficients) היא רלוונטית. טבלה זו מציינת את מקדמי הרגרסיה הליניארית, בציוני גלם ובציוני תקן.

בטבלה זו שני שורות, Constant ו-Beginning salary. השורה Constant (קבוע) מתייחסת לקבוע הרגרסיה a. השורה Beginning salary מתייחסת לקבוע הרגרסיה b.

[תזכורת: משוואת הרגרסיה הליניארית הפשוטה היא $\tilde{Y} = b \cdot X + a$, כאשר y הוא המשתנה המנובא ו-x הוא המשתנה המנבא. ערכו של a יוצג בשורה (constant), וערכו של b יוצג בשורה Beginning Salary, מאחר והוא קשור למשתנה זה].

ערכיהם של a ו-b מופיעים בעמודה השנייה משמאל (תחת הכותרת "B"). בדוגמה שלנו, ערכו של a הוא 8653.668 וערכו של b הוא 1.495. משוואת הרגרסיה, לכן, היא $\hat{y} = 1.495x + 8653.668$. העמודות האחרות בטבלה זו (Std. Error, t, Sig.) קשורות גם הן למובהקות, ויילמדו במהלך סמסטר ב'. על העמודה Beta תלמדו דרך הדוגמה הבאה.

רגרסיה מרובה

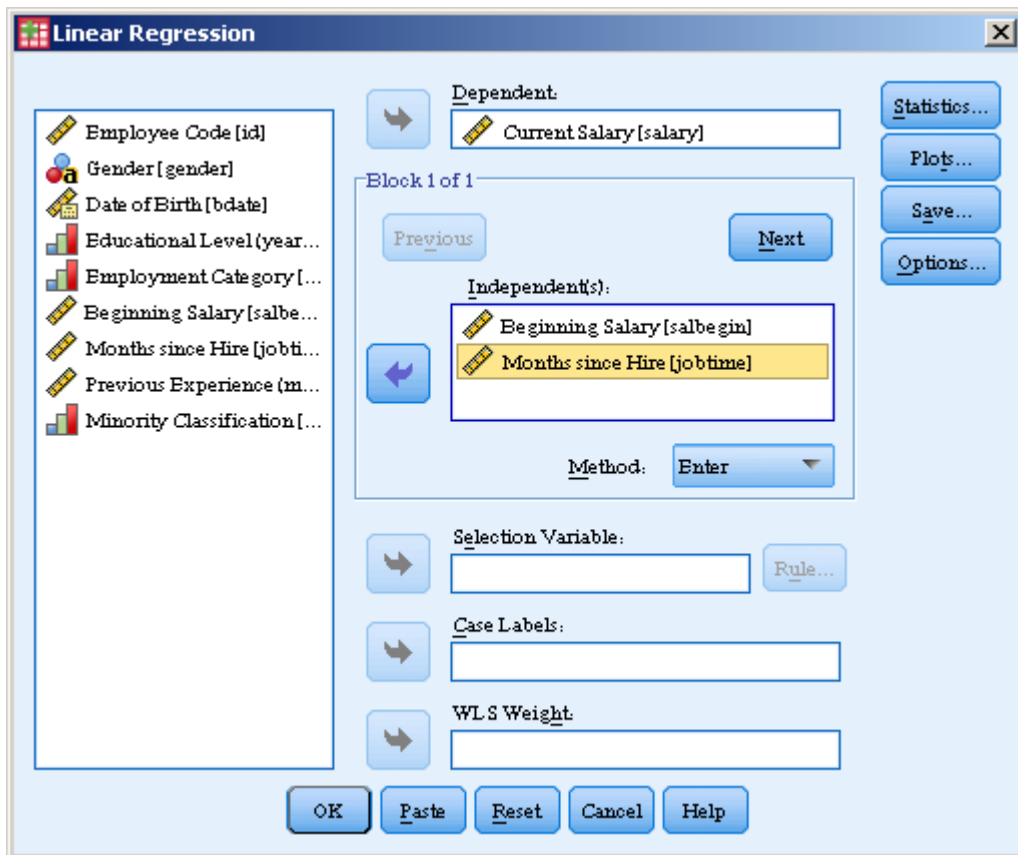
רגרסיה מרובה היא הרחבה של אותם העקרונות של הרגרסיה הפשוטה. גם כאן, ברצוננו לנבא את ערכו של משתנה מנובא אחד (Y). אולם, הפעם, נשתמש בכמה משתנים מנבאים, שיסומנו X1, X2, X3 וכן הלאה. היתרון ברגרסיה מרובה הוא דיוק: ככל שמשמשים במשתנים רבים יותר לניבוי, כך הניבוי יהיה מדויק יותר. למשל: ברצוננו לנבא את השכר הנוכחי של העובד (salary). אנו נדייק יותר בניבוי אם נדע גם את השכר ההתחלתי שלו (salbegin) וגם את מספר חודשי ההעסקה שלו (jobtime), מאשר אם נדע את השכר ההתחלתי שלו בלבד.

באופן עקרוני, אין מגבלה על מספר המשתנים המנבאים ברגרסיה מרובה. משוואת הרגרסיה הפשוטה, בציוני גלם, היא $\tilde{Y} = b \cdot X + a$. ברגרסיה מרובה, מאחר וקיימים כמה מנבאים

(כלומר, כמה X-ים), יהיה לכל אחד מהם b משלו. כך, הופכת המשוואה להיות

$$\tilde{Y} = b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3 + \dots + a$$
 במשוואה זו יהיה b נפרד עבור כל מנבא, וקבוע a אחד בלבד.

נמשיך את הדוגמה הקודמת. הפעם, נרצה לנבא את השכר הנוכחי באמצעות השכר ההתחלתי ובאמצעות מספר חודשי ההעסקה. ניכנס לניתוח הרגרסיה כמו מקודם. הפעם, מאחר וישנם שני משתנים מנבאים, נעביר את שניהם (jobtime ו-salbegin) לחלון ה-Independents.



לאחר לחיצה על OK, יתקבל הפלט הבא בחלון ה-Output:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.788 ^a	.621	.619	\$10,540.477

a. Predictors: (Constant), Months since Hire, Beginning Salary

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	85770086042.339	2	42885043021.169	385.998	.000 ^a
	Residual	52328877115.520	471	111101649.927		
	Total	138098963157.859	473			

a. Predictors: (Constant), Months since Hire, Beginning Salary

b. Dependent Variable: Current Salary

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2748.864	4072.592		-.675	.500
	Beginning Salary	1.496	.054	.784	27.640	.000
	Months since Hire	140.644	48.535	.082	2.898	.004

a. Dependent Variable: Current Salary

נסתכל תחילה על הטבלה הראשונה, Model Summary. העמודה השניה בטבלה, R square, מציגה את R^2 – פרופורציית השונות המוסברת. ברגרסיה מרובה (כמו ברגרסיה פשוטה), זהו החלק של השונות במשתנה המנובא, שמוסבר על-ידי שונות במשתנים המנבאים. בדוגמה שלפנינו, ניתן לראות כי 62.1% מהשונות במשתנה salary מוסברת על-ידי שונות במשתנים salbegin ו-jobtime. כמו ברגרסיה פשוטה, גם ברגרסיה מרובה R^2 מהווה מדד לטיב ההתאמה של משוואת הרגרסיה. זהו ערך שנע בין 0 ל-1, וככל שערכו גבוה יותר – כך הניבוי טוב יותר. הטבלה השניה, ANOVA, קשורה לבדיקת המובהקות של מודל הרגרסיה, ועליה לא נדבר בשלב זה.

הטבלה השניה, Coefficients, מכילה את ערכיהם של מקדמי הרגרסיה. מאחר ובדוגמה זו יש שני מנבאים, משוואת הרגרסיה כוללת 3 מקדמים: a (קבוע), b1 (המקדם של המשתנה המנבא הראשון, salbegin), ו-b2 (המקדם של המשתנה המנבא השני, jobtime). העמודה השניה משמאל בטבלה זו מכילה את ערכי המקדמים. לכן, משוואת הרגרסיה בציוני גלם היא:

$$\hat{Y} = 1.496 \cdot X_1 + 140.644 \cdot X_2 - 2748.864$$

(כאשר X_1 ו- X_2 הם המשתנים salbegin ו-jobtime בהתאמה).

כמו בדוגמה הקודמת, העמודות t, Std. Error, ו-Sig. מתייחסות לבדיקת מובהקות, ואנו נתעלם מהן בשלב זה.

לבסוף, העמודה Beta מכילה את ערכיהם של מקדמי הרגרסיה בציוני תקן. כמו ברגרסיה פשוטה, גם ברגרסיה מרובה ניתן לחשב משוואת רגרסיה בציוני גלם ובציוני תקן. בציוני תקן, המשוואה היא:

שימו לב שכאן כל המשתנים מופיעים כציוני תקן ולא כציוני גלם. בנוסף, המשוואה לא כוללת את הקבוע a. כמו-כן, במקום מקדמי הרגרסיה b1, b2, וכו', המקדמים כאן נקראים β_1 , β_2 וכו', בהתאמה. גם מקדמים אלה הינם מספרים, שניתן לחשב אותם. לכן, על פי הטבלה, משוואת הרגרסיה בציוני תקן היא:

$$\tilde{Z}_Y = 0.784 \cdot Z_{X_1} + 0.82 \cdot Z_{X_2}$$

תרגיל הגשה מספר 4

1. השתמשו בקובץ World95.sav, שבו השתמשתם בשיעור 3 בתוכנות סטטיסטיות.
2. באמצעות הניתוחים הסטטיסטיים שלמדתם, צרו קובץ תוצאות (output) שיכיל טבלאות, שבהן נתונים העונים על השאלות הבאות. שמרו את קובץ התוצאות (מתוך חלון ה-output ניתן לבחור בתפריט File ואז Save as).
- צרפו קובץ נוסף של WORD, ובו התשובות המספריות בלבד לשאלות שלהלן (לדוגמה: א – 65, ב – 537, וכן הלאה). שמרו את הקבצים שניהם תחת הת.ז שלכם ושלחו לאותו מייל של העבודות הקודמות (STAT.BGU@GMAIL.COM).
- א. מהו ערכו של מתאם פירסון בין תוחלת החיים הממוצעת לגברים לבין תוחלת החיים הממוצעת לנשים?
- ב. מהו ערכו של מתאם ספירמן בין אחוז העיור לבין צריכת הקלוריות היומית הממוצעת?
- ג. מהי משוואת הרגרסיה, בציוני גלם, לניבוי צריכת הקלוריות היומית על-ידי גודל האוכלוסיה?
- ד. מהי משוואת הרגרסיה, בציוני גלם ובציוני תקן, לניבוי אחוז התושבים היודעים קרוא וכתוב על-ידי גודל האוכלוסיה, צפיפות האוכלוסין ומספר הילדים הממוצע במשפחה?

שיעור מספר 5

שיעור מס' 5: יצירת משתנים חדשים ושינוי במשתנים קיימים

חישוב משתנים חדשים - Compute:

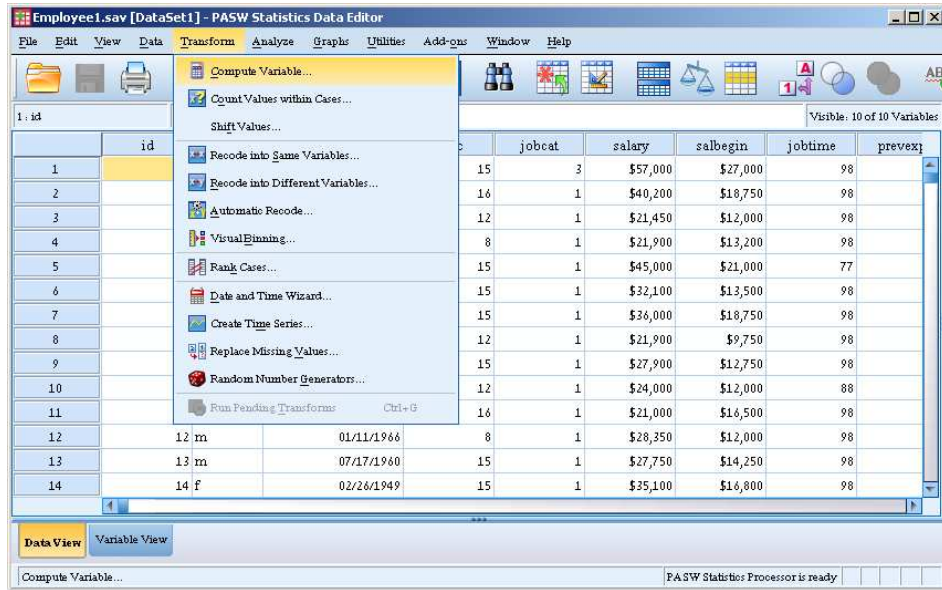
עד כה למדתם כיצד לנתח קובץ נתונים, ולחשב מדדים סטטיסטיים על הנתונים שהיו קיימים בקובץ. אולם, פעמים רבות תרצו ליצור משתנים נוספים בקובץ, שיכילו מידע שונה מהמידע שכבר קיים בו. למשל, נניח וקובץ הנתונים מכיל את ציוני הבוחן והמבחן של סטודנטים:

מספר סטודנט	ציון בוחן	ציון מבחן
1	80	100
2	90	99
3	52	67
4	43	70
5	67	85
6	89	85
7	98	100

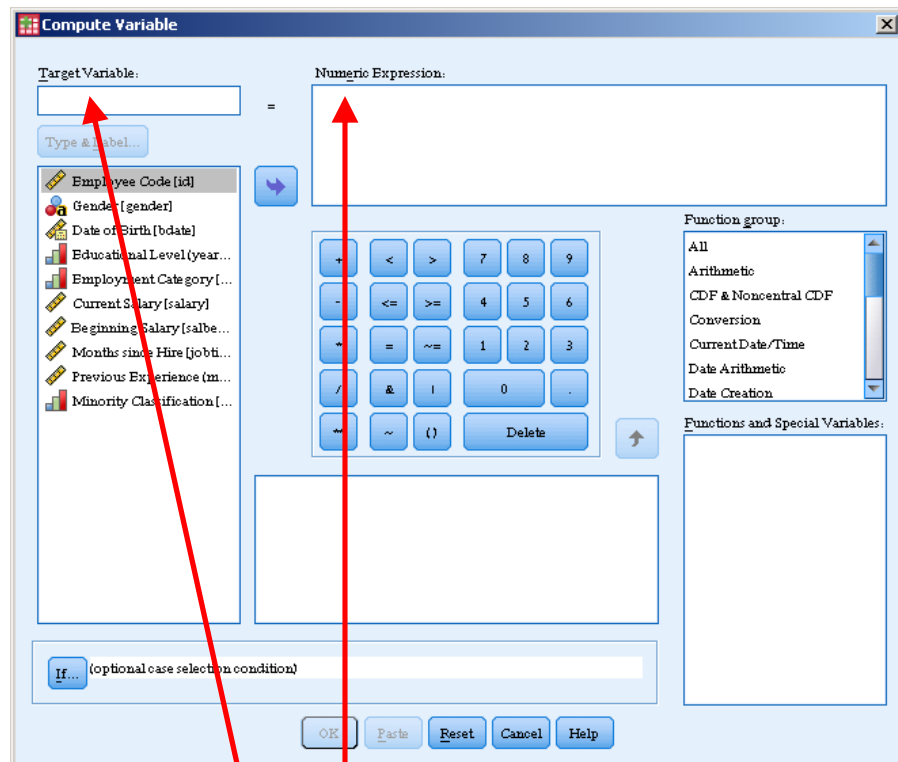
כעת, על מנת לחשב את הציון הסופי בקורס, תרצו ליצור עמודה חדשה, נוספת, בקובץ הנתונים. הערך שיהיה בעמודה זו יחושב על ידי שקלול הערכים (הציונים) בעמודות המכילות את ציוני הבוחן והמבחן. הדרך לעשות זאת ב-SPSS היא באמצעות חלון הפקודה Compute.

פתחו את קובץ הנתונים Employee, שבו השתמשתם בשיעור מספר 4. במהלך שיעור זה נשנה חלק מהמידע בקובץ. כדי שמידע הקיים לא יילך לאיבוד, אנא שמרו את הקובץ כפי שהוא בשם Employee1. המשתנה Salary מכיל את השכר החודשי הנוכחי של כל עובד, בדולרים. נניח עתה, כי ברצוננו ליצור משתנה נוסף בקובץ, המכיל את השכר השנתי של כל עובד.

לשם כך, יש להיכנס לתפריט הראשי Transform, ואז לבחור באפשרות Compute Variable:



כעת יפתח החלון הבא :

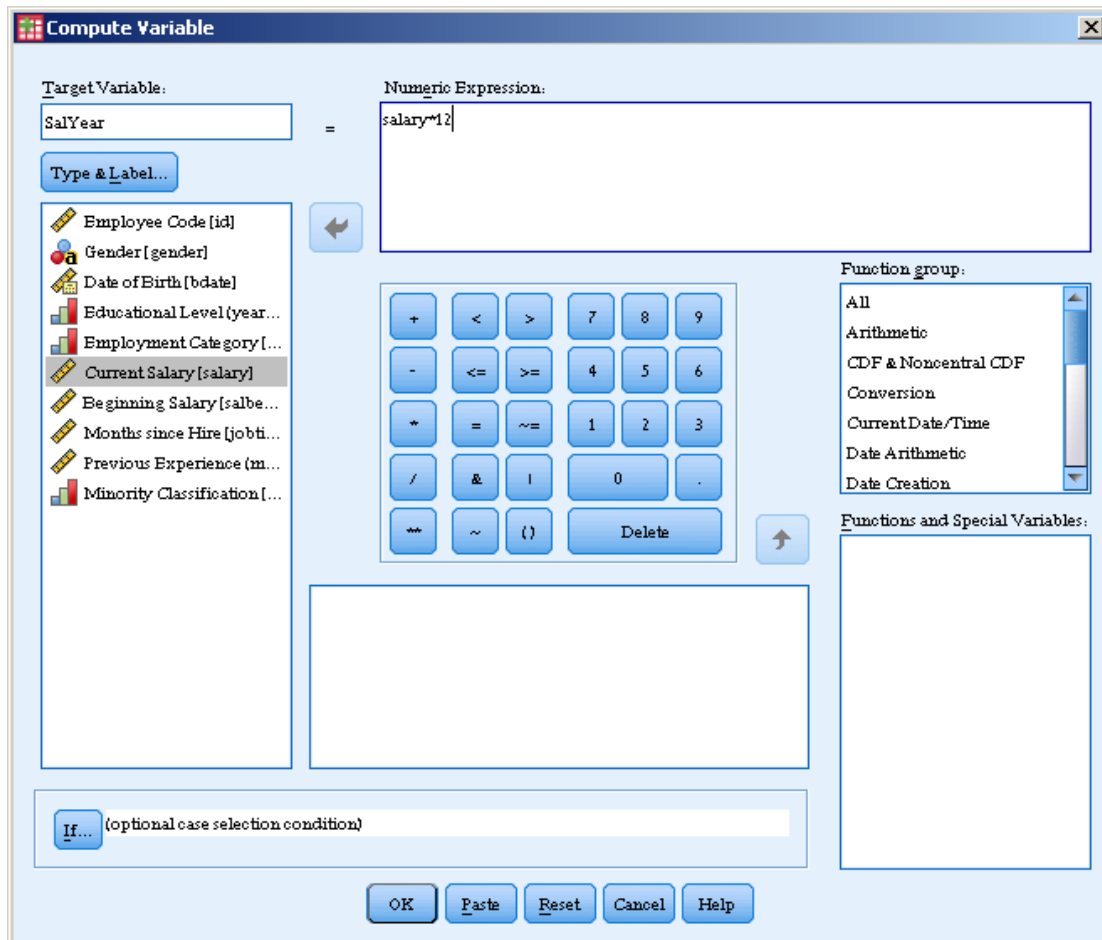


בחלונת השמאלית העליונה, תחת הכותרת Target Variable, יש לכתוב את שם המשתנה שברצונכם ליצור. במקרה זה, השם יכול להיות SalYear.

בחלונת הימנית העליונה, תחת הכותרת Numeric Expression, ניתן לכתוב את הנוסחה שבאמצעותה יחושבו הערכים במשתנה החדש. במקרה זה, הנוסחה תהיה המשכורת החודשית כפול 12, כלומר, הערך שכתוב עבור כל תצפית במשתנה Salary, כפול 12. כדי להגדיר זאת, יש לסמן את המשתנה Salary בחלון המשתנים, לגרור אותו לחלון ה-Numeric Expression באמצעות החץ השחור, ולכתוב "12*" (הערה: הסימן כוכבית * מציין כפל).

לאחר מכן, החלון ייראה

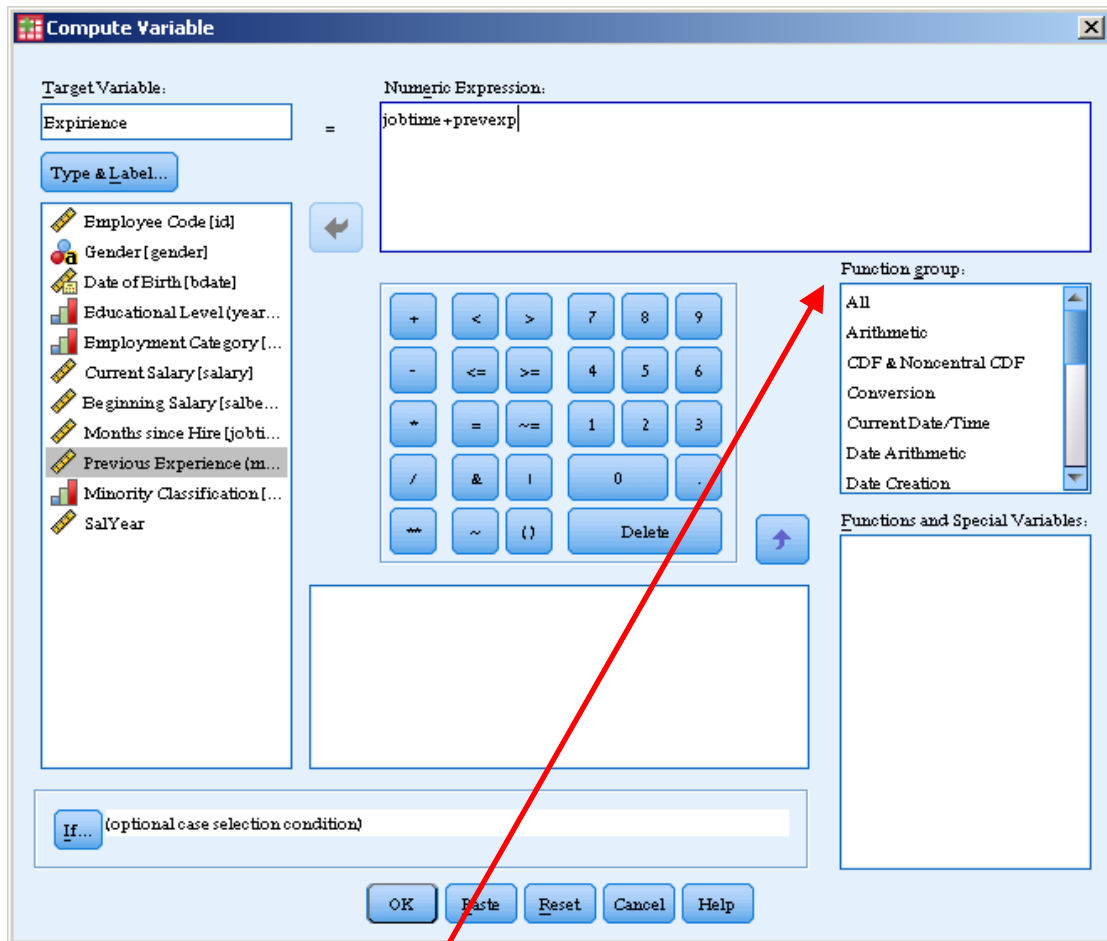
כך:



כעת, כדי לסיים את הפעולה, יש ללחוץ על המקש OK, בחלקו התחתון של החלון. כעת, תוכלו לראות בחלון הנתונים את המשתנה SalYear. משתנה זה מכיל עבור כל תצפית את המשכורת השנתית המתאימה לה, המחושבת על-ידי המשכורת החודשית כפול 12.

בדוגמה הקודמת, יצרתם משתנה חדש שחושב על-ידי הערכים של משתנה קיים אחד. אולם, ניתן לחשב משתנים חדשים גם על-ידי שקלול ערכים במספר משתנים. למשל, ניח שברצוננו ליצור משתנה בשם Experience (ניסיון), שהערכים שלו יהיו סכום הניסיון הקודם של העובד (בחודשים), ומספר חודשי ההעסקה שלו בחברה הנוכחית (בחודשים).

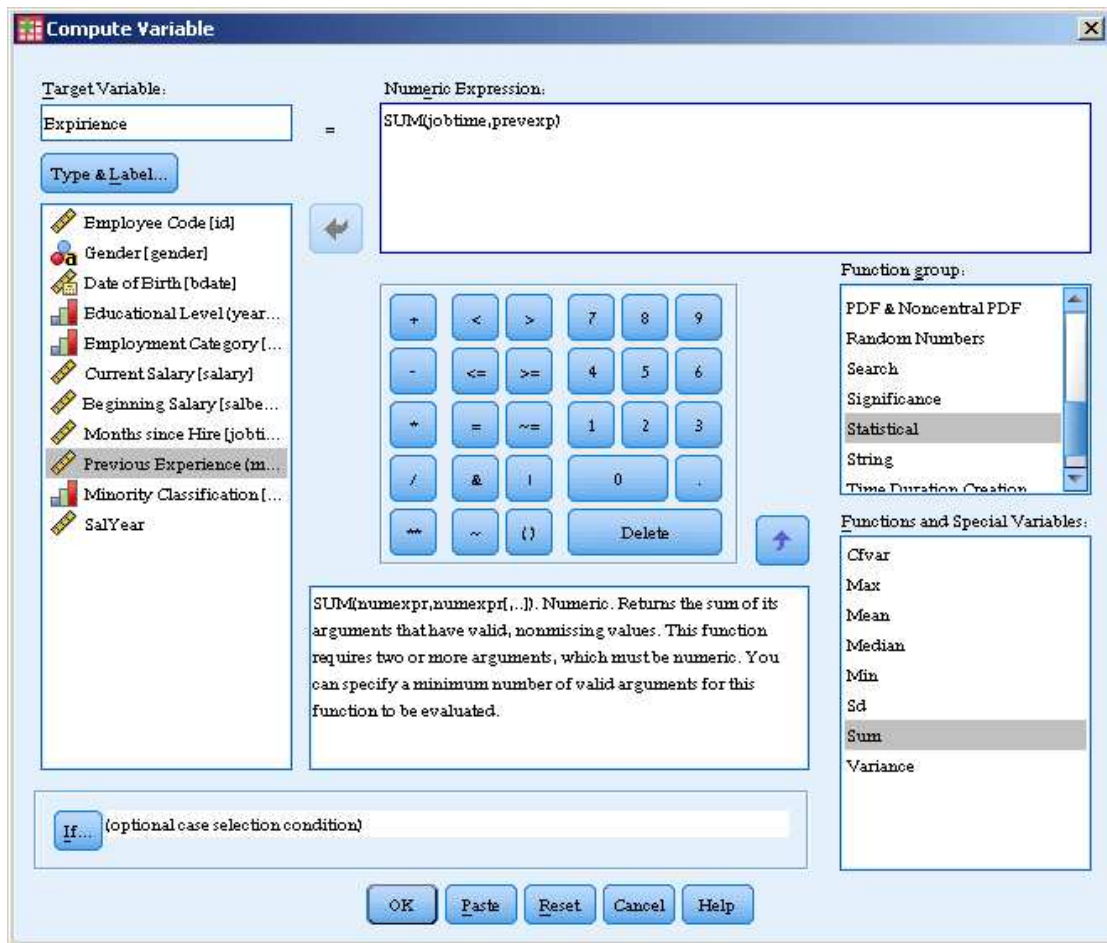
לשם כך, יש לפתוח שוב את חלון ה-Compute, וליצור את המשתנה Experience. ניתן לעשות זאת בשתי דרכים. הדרך הראשונה תהיה להגדיר את המשתנה החדש כסכום המשתנים הקיימים באופן הבא:



דרך זו הינה הפשוטה ביותר, אולם תהיה ארוכה אם יהיה צורך לחשב סכום של 10 משתנים, או ממוצע של 30 משתנים, וכן הלאה. כדי לבצע פעולות מורכבות מסוג זה, וכן כדי להגדיר משתנים חדשים באמצעות חישובים מורכבים יותר, ניתן להשתמש בפונקציות הקיימות ב-SPSS. את הפונקציות ניתן לבחור מתוך חלון הפונקציות.

בחלקו הימני של חלון ה-Compute נמצאות שתי חלונות. החלונת העליונה יותר, Function group, מאפשרת לבחור את קבוצת הפונקציות הרצויה: פונקציות מתמטיות, פונקציות סטטיסטיות, פונקציות המתאימות לעבודה עם משתנים מילוליים, פונקציות פיננסיות, וכו'. לאחר שבחרתם את הקטגוריה המתאימה (במקרה זה, פונקציות סטטיסטיות), תופיע רשימת הפונקציות הרלוונטיות בחלונת התחתונה יותר, Functions and special variables.

בחרו כעת את הפונקציה Sum (סכום) מתוך קטגוריית הפונקציות הסטטיסטיות. אל תוך הסוגריים של הפונקציה, העבירו את המשתנים שברצונכם לחבר. החלון אמור להיראות כעת כך:



כעת, ניתן לחוץ על OK. המשתנה Experience ייוצר בקובץ הנתונים. פונקציות שימושיות אחרות שתוכלו להזדקק להן לעיתים הן ממוצע (Mean), סטיית תקן (Sd), שונות (Variance), מינימום (הערך המינימלי מבין סדרה של משתנים, Min) ומקסימום (הערך המקסימלי מבין סדרה של משתנים, Max).

לדוגמה, נניח שלפניכם קובץ נתונים המכיל את משכורות העובדים לפי חודשים. בקובץ זה, לכל עובד יהיו נתונים ב-12 משתנים, Sal1 עד Sal12, עבור כל אחד מחודשי השנה:

id	Sal1	Sal2	Sal3	Sal4	Sal5	Sal6	Sal7	Sal8	Sal9	Sal10	Sal11	Sal12
1												
2												
3												
4												
5												

על מנת לחשב עבור כל עובד את המשכורת החודשית הממוצעת, יש ליצור משתנה חדש (בשם MeanSal, למשל). בחלון ה-Numeric Expression ניתן להגדיר משתנה זה באופן הבא:

Mean(Sal1 to Sal12)

(** את המילה to יש לכתוב באופן ידני).

באופן דומה, ניתן ליצור משתנה המכיל את המשכורת השנתית, כלומר Sum(Sal1 to Sal12), או משתנה המכיל את שונות המשכורות, המשכורת הגבוהה/הנמוכה ביותר, וכן הלאה.

קטגוריית הפונקציות האריתמטיות, Arithmetic, מכילה פונקציות מתמטיות וחשבוניות רבות, שגם הן עשויות להיות שימושיות. בכל פעם שתעמדו עם הסמן על פונקציה מסויימת, תיאור מילולי של הפונקציה יופיע בחלון האפור שמשמאל לחלון הפונקציות. ניתן להיעזר בתיאור זה כדי לדעת מה מבצעת כל פונקציה.

קידוד מחדש של משתנים קיימים - Recode

לעיתים נרצה לשנות את ערכיהם של משתנים קיימים. אפשרות זאת מתבצעת באמצעות התפריט Recode (קידוד מחדש). דוגמאות לחלק מהאפשרויות שפקודה זו מציעה:

- קידוד מחדש של סולם רציף (כגון משתנה גיל, עם ערכים רציפים בין 1-100) לסולם בדיד, בעל מספר מועט של רמות (כגון: גילאי 1-20 = רמה 1 'צעירים', גילאי 21-60 = רמה 2 'בוגרים', גילאי 61-100 = רמה 3 'מבוגרים').
- קידוד מחדש של ערכי סולם מדידה, כגון: הפיכת ערכי משתנה מסולם בסדר עולה (מ 1 הנמוך ביותר ל 5 הגבוה ביותר) לסולם בעל סדר יורד (מ 5 הנמוך ביותר ל 1 הגבוה ביותר).

למשל, בדוגמה זו, נניח שברצוננו להפוך את משתנה ההשכלה לקטגוריאל. במקום לדווח על מספר שנות ההשכלה, ברצוננו ליצור משתנה חדש, שיגדיר את ההשכלה באופן הבא:

הערך "1" – יציין עובדים בעלי עד 12 שנות השכלה (כולל)

הערך "2" - יציין עובדים בעלי 13-16 שנות השכלה (כולל)

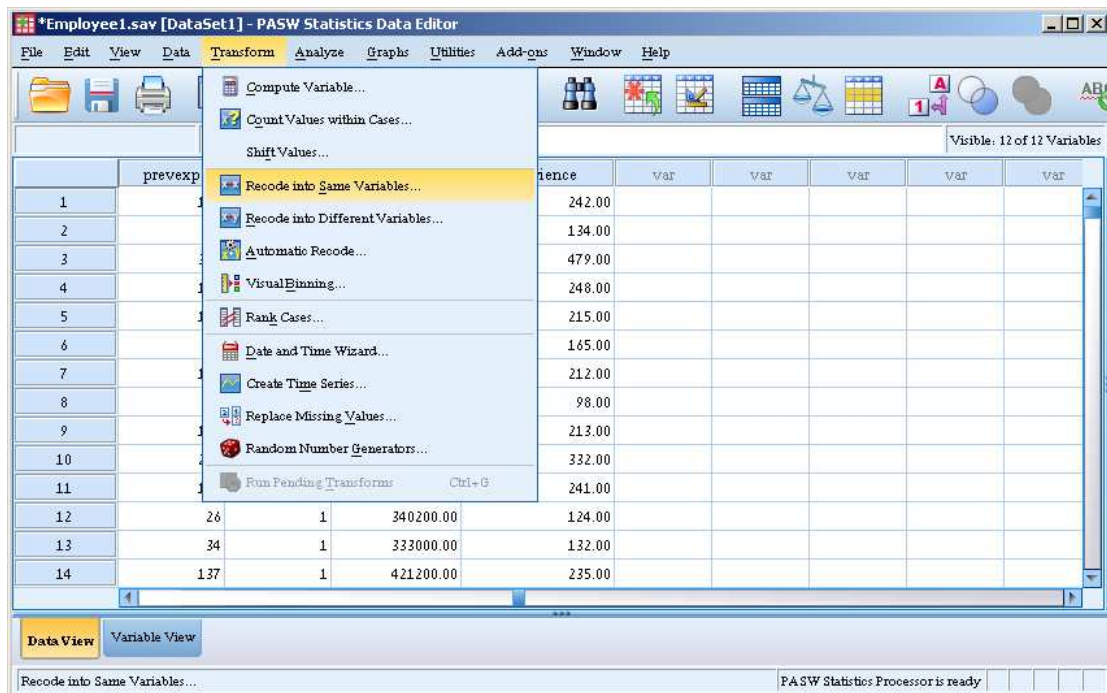
הערך "3" - יציין עובדים בעלי 17 שנות השכלה ומעלה.

את הקידוד החדש נוכל לבצע בשתי דרכים. הדרך הראשונה, שאינה מומלצת, היא לשנות את ערכי המשתנה educ עצמם. כלומר, משתנה זה לא יכיל מעתה מספר שנות השכלה אלא רק את קטגוריית ההשכלה. זהו למעשה קידוד מחדש של המשתנה עצמו. אפשרות זו אינה מומלצת מאחר והערכים המקוריים במשתנה יאבדו באופן זה.

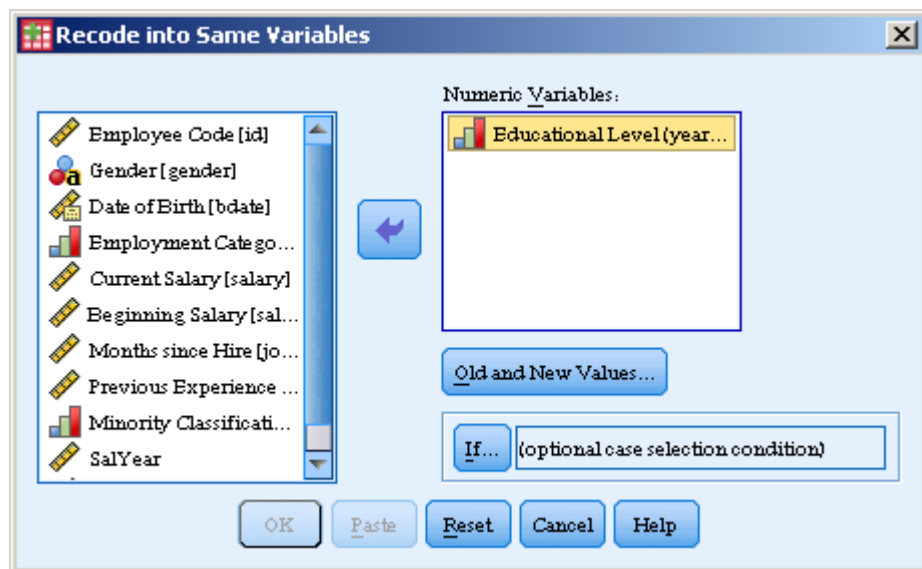
הדרך השנייה, והמומלצת, היא קידוד ערכי המשתנה educ אל תוך משתנה חדש. באופן זה, המשתנה המקורי יישמר, ובנוסף אליו ייוצר משתנה נוסף שיכיל את קטגוריית ההשכלה. כך, לא יאבדו הערכים המקוריים של המשתנה.

כעת, נעבור להדגים את השימוש בשתי השיטות. נתחיל בדרך הראשונה.

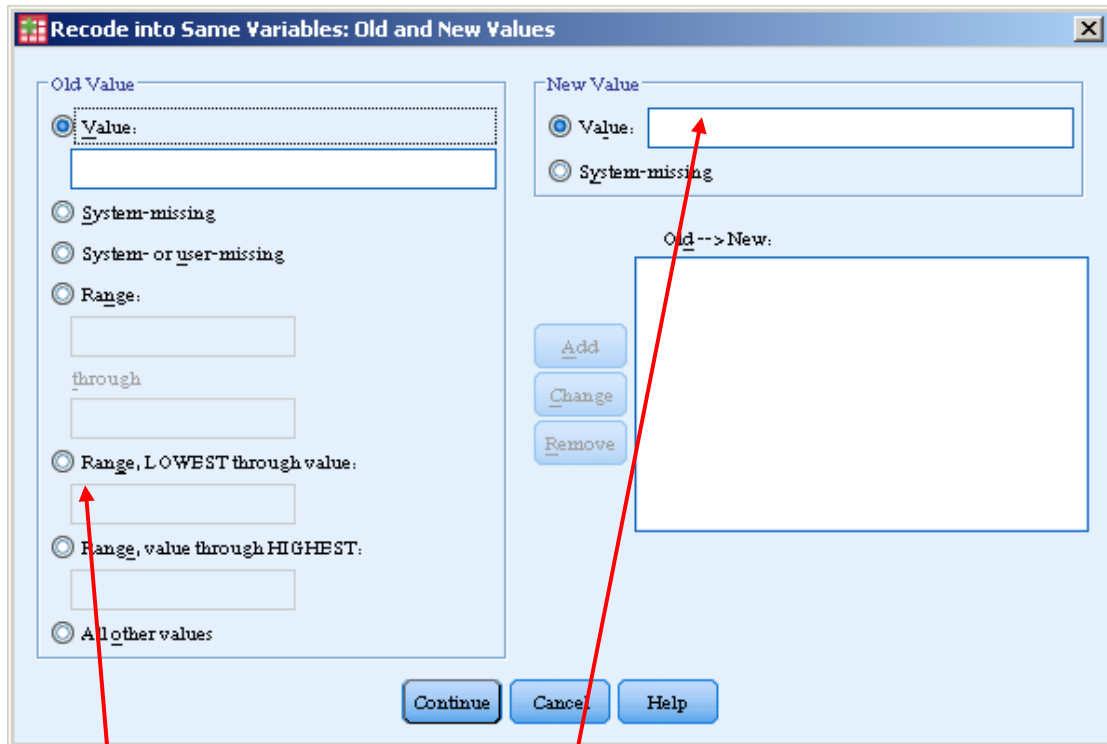
לשם כך, יש לבחור בתפריט Transform, ואז Recode Into Same Variables :



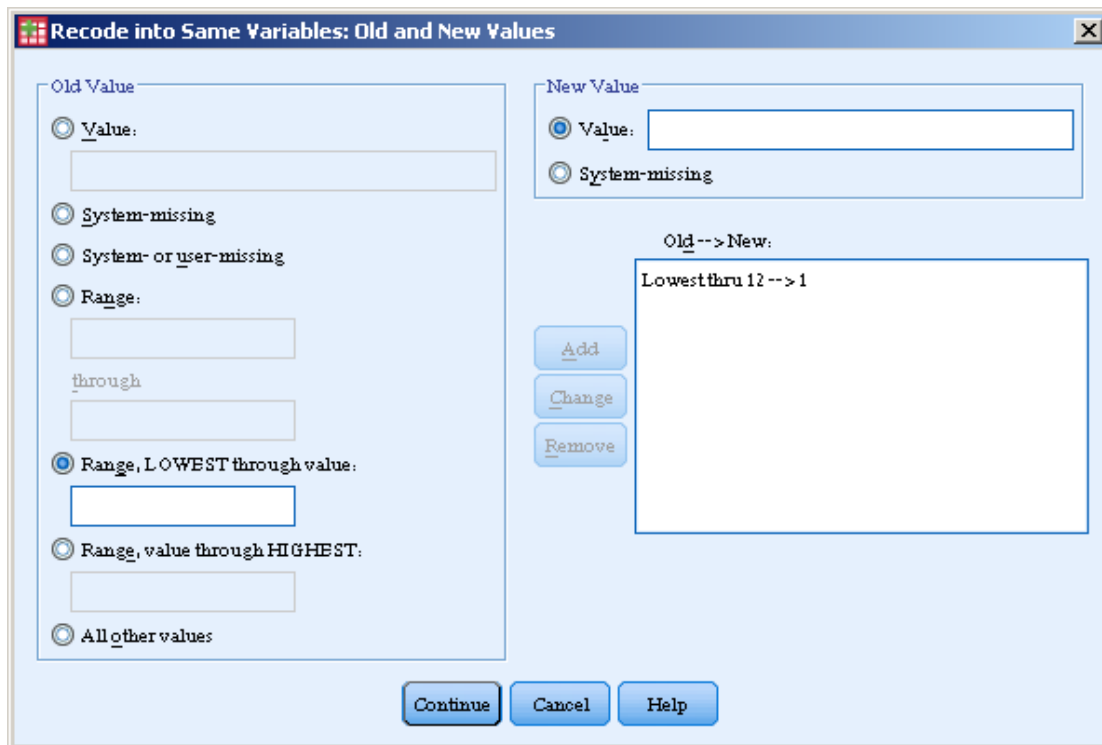
כעת, ייפתח חלון שבו יש לבחור את המשתנים שברצוננו לשנות את ערכיהם. במקרה זה, המשתנה הוא educ. יש לגרור אותו לחלון הימני (Numeric Variables) :



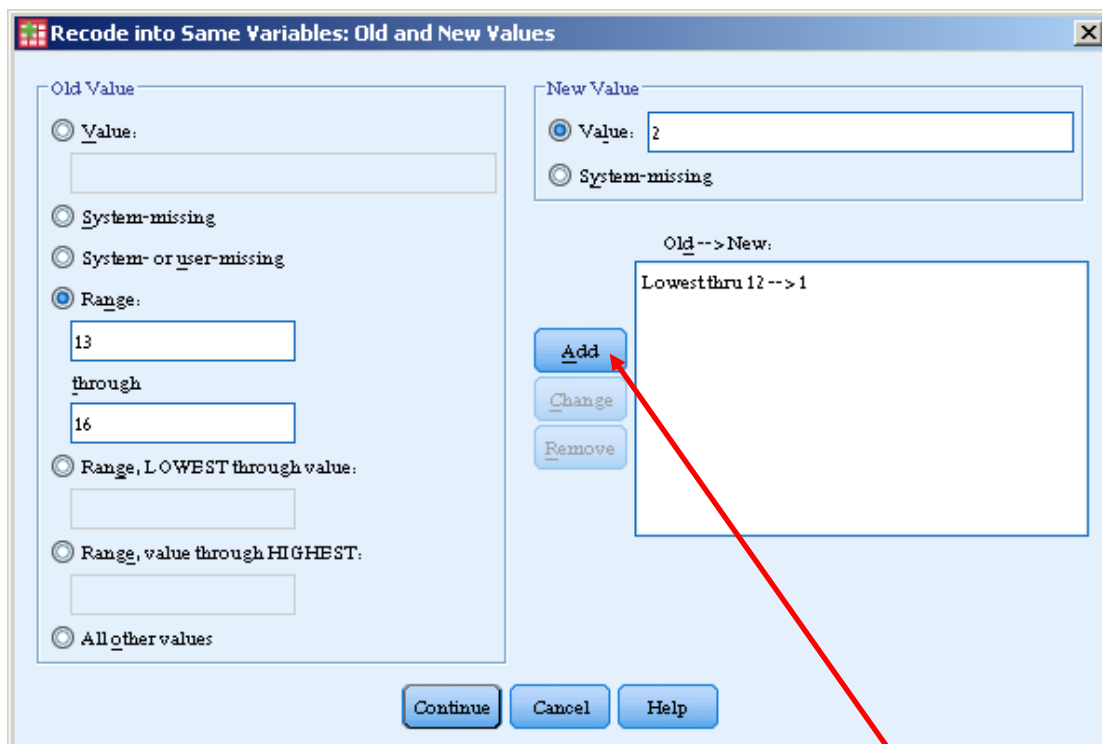
כדי להגדיר את האופן שבו ברצוננו לקודד מחדש את ערכיו של משתנה זה, יש ללחוץ על הכפתור Old and New Values. ייפתח החלון הבא :



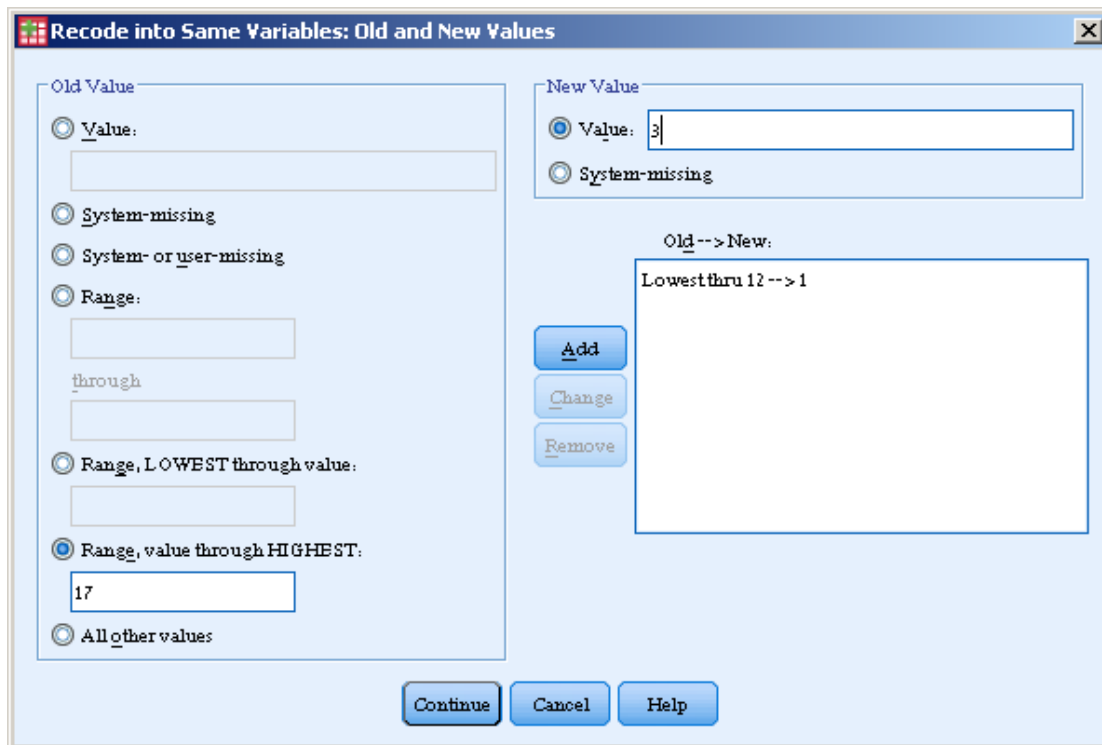
חלון זה מחולק לשניים. החלק השמאלי מיועד להגדיר את הערכים ה"ישנים" במשתנה, והחלק הימני מיועד להגדיר את הערכים ה"חדשים" המתאימים להם. לדוגמה, קטגוריית ההשכלה הנמוכה יותר בדוגמה שלנו היא עד 12 שנים. הערכים הישנים במשתנה, המתאימים לקטגוריה זו, הם ממספר שנות ההשכלה הקטן ביותר, ועד 12. לשם כך, נבחר באפשרות Range, LOWEST through value. אפשרות זו מאפשרת לבחור את הטווח שבין הערך הנמוך ביותר במשתנה (LOWEST) לבין הערך שברצוננו לקבוע. בחלונית שתיפתח יש לכתוב את הערך 12. כעת, לאחר שבחרנו את טווח הערכים במשתנה הישן, עלינו לכתוב מהו הערך החדש שייכתב במקומם. לפי האופן שבו הגדרנו קודם את הקטגוריות, קטגוריית השכלה זו תסומן באמצעות המספר 1. לכן, יש לכתוב את המספר 1 בחלונית New Value שבחציו הימני של החלון. לבסוף, יש ללחוץ על הכפתור Add. כעת, החלון ייראה כך :



לאחר שהגדרנו את קטגוריית ההשכלה הראשונה, יש להגדיר את הקטגוריה הבאה. בחלקו השמאלי של החלון נבחר באפשרות Range (טווח), נגדיר את הטווח כ-13 עד 16, ונכתוב את הערך 2 בחלקו הימני של החלון:



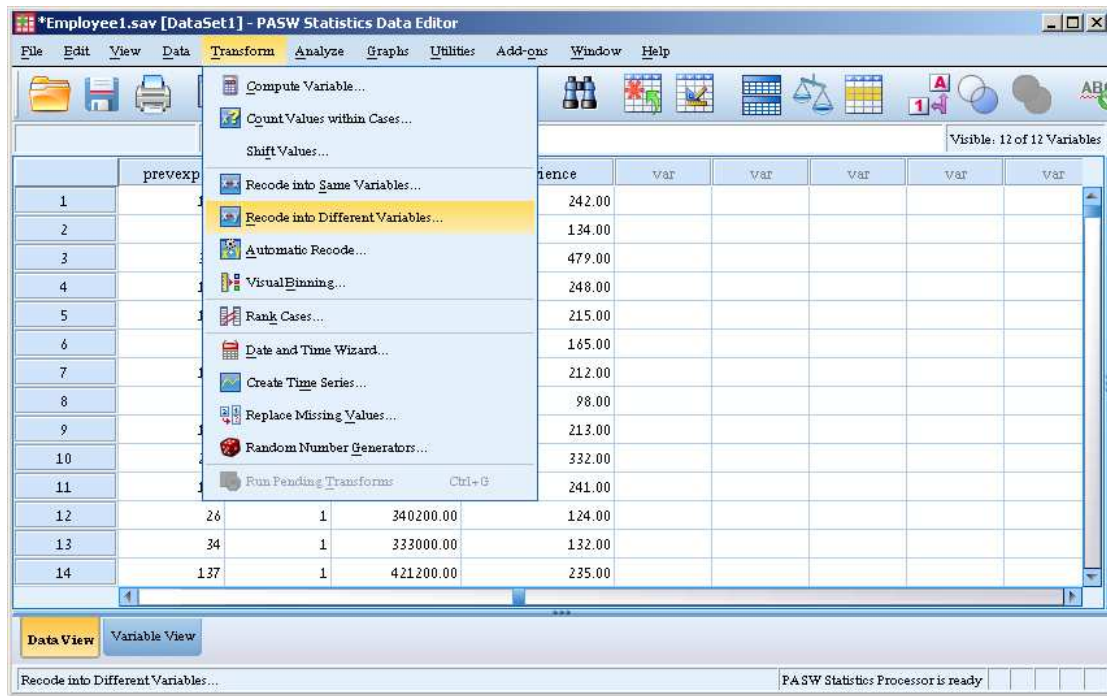
שוב, נלחץ על Add, ונעבור להגדיר את הקטגוריה השלישית:



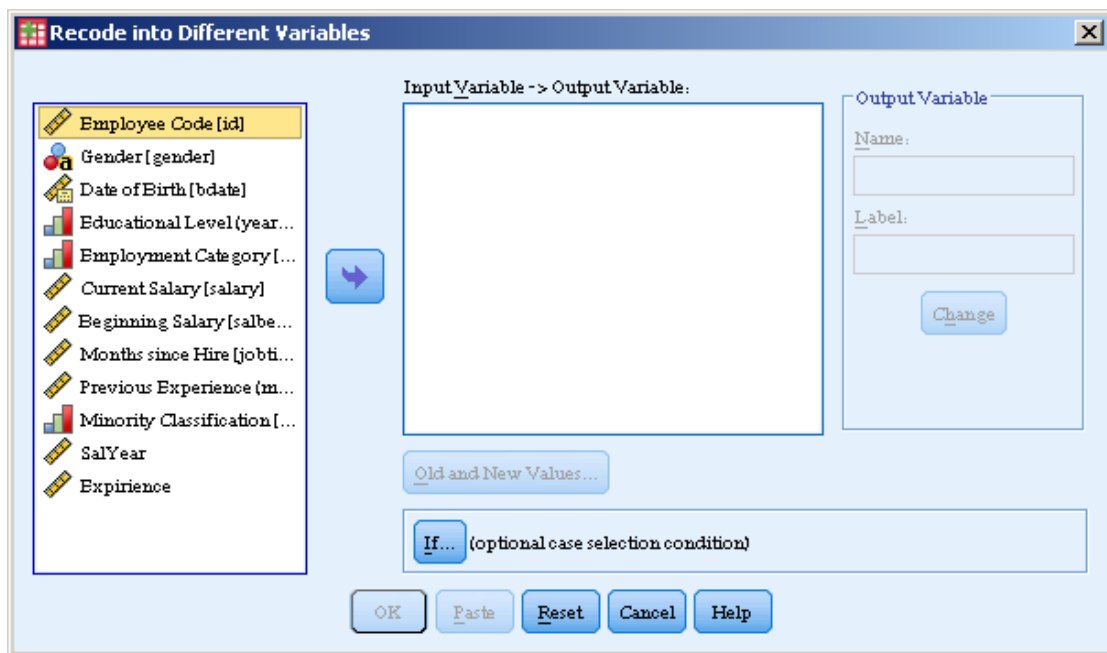
נעשה זאת באמצעות בחירה ב-Range, value through HIGHEST, מאחר וברצוננו להגדיר טווח שבין 17 לבין הערך הגבוה ביותר הקיים. שוב, נלחץ על Add. כעת, לאחר שהגדרנו את האופן שברצוננו לקודד מחדש את המשתנה, נלחץ על הכפתור Continue, ואז OK. כעת, תחת שם המשתנה educ שבחלון הנתונים לא יופיעו מספר שנות ההשכלה, אלא הערכים 1, 2 או 3 בהתאם לקטגוריה הרלוונטית.

כפי שכבר נאמר קודם, רצוי מאוד שלא לאבד את הערכים המקוריים של המשתנה, אלא לבצע את הקידוד אל תוך משתנה אחר. לשם הדוגמה, פתחו שוב את הקובץ Employee (המקורי). כעת, נרצה ליצור משתנה בשם EducCat (קטגוריית השכלה), שלתוכו נכניס את הערכים 1, 2 או 3 בהתאם להשכלתו של כל עובד.

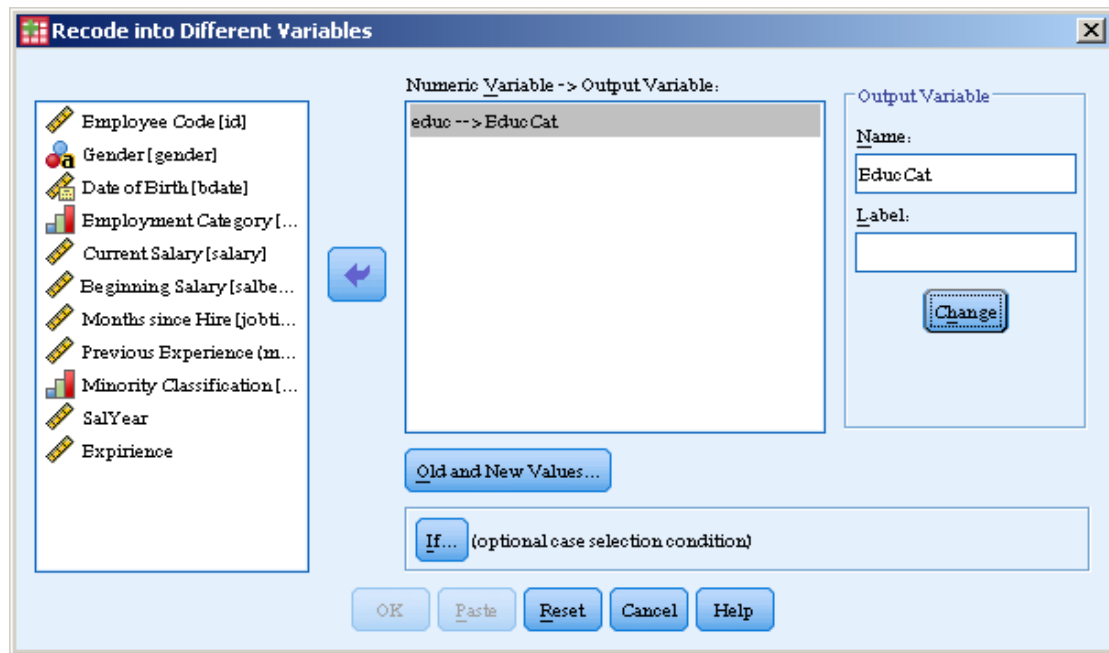
לשם כך, נבחר בתפריט Transform ובאפשרות Recode into Different Variables:



כעת, ייפתח החלון הבא :



בחלקו השמאלי של החלון מופיעה רשימת המשתנים בקובץ הנתונים. את המשתנה שאת ערכיו רוצים לשנות יש לגרור, באמצעות החץ, לחלון הלבן המרכזי. לאחר מכן, ייפתחו שתי החלוניות שבחלקן הימני של המסך (Output Variable). בתוכן, ניתן לכתוב את שם המשתנה החדש שברצונכם ליצור. יש ללחוץ על הכפתור Change, ואז ייכנס השם החדש אל תוך החלון המרכזי :



כעת, כל מה שנותר לעשות, הוא להגדיר את האופן שבו ייתבצע השינוי בערכי המשתנה. יש לעשות זאת באמצעות לחיצה על הכפתור Old and New Values, באופן זהה לזה שביצעתם קודם ב-Recode into Same Variables. בסוף התהליך, יש ללחוץ על הכפתור OK. כעת, קובץ הנתונים יכיל את המשתנה educ שערכיו יישמרו ללא שינוי. בנוסף, ייוצר המשתנה EducCat על פי ההגדרות שהגדרתם.

תרגיל הגשה מספר 5

3. השתמשו בקובץ World95.sav
4. באמצעות הפקודות שלמדתם בשיעור זה, והניתוחים הסטטיסטיים שלמדתם בשיעורים הקודמים, צרו קובץ תוצאות (output) שיכיל טבלאות, שבהן נתונים העונים על השאלות הבאות. הדפיסו את קובץ התוצאות (מתוך חלון ה-output ניתן לבחור בתפריט File ואז Print). לבסוף, צרפו דף נוסף, ובו התשובות המספריות בלבד לשאלות שלהלן (לדוגמה: א – 65, ב- 537, וכן הלאה).
 - א. צרו משתנה בשם lifeexp, שיכיל את תוחלת החיים הממוצעת במדינה. משתנה זה יחושב כממוצע של תוחלת החיים לגברים ושל תוחלת החיים לנשים
 - ב. חשבו ממוצע, שונות ורבעונים של המשתנה lifeexp שיצרים.
 - ג. הפכו את המשתנה urban לקטגוריות: 0-20%, 20-40%, 40-60%, 60-80%, 80-100%. צרו משתנה חדש שיכיל מידע זה, וייקרא UrbCat.
 - ד. הציגו טבלת שכיחויות למשתנה UrbCat.
 - ה. חשבו את השטח של כל מדינה באמצעות שימוש במשתנים populatn ו-density.
 - ו. הציגו טבלת שכיחויות למשתנה השטח שיצרתם.
 - ז. חלקו את המשתנה aids לשתי קטגוריות שוות, כך שמספר שווה של מדינות יהיה בכל אחת מהן. צרו משתנה חדש בשם AidsCat שיכיל, עבור כל מדינה, את הקטגוריה המתאימה שהיא שייכת אליה.
 - ח. חשבו את מקדם המתאם המתאים ביותר בין UrbCat ו-AidsCat.

שיעור מספר 6

שיעור מס' 6: מבחני T להשוואה בין ממוצעים והצגה גראפית של התוצאות

מבוא לשיעור

כאשר אנו עוסקים בבדיקת השערות ביחס לממוצע או ממוצעים, ברוב המקרים איננו יודעים למה שווה שונות האוכלוסייה ובכל זאת אנו רוצים להסיק מסקנות בקשר לממוצעים. כאשר נרצה להשוות בין שני ממוצעים נשתמש במבחן T.

התפלגות T היא סימטרית, חד שכיחית ודמוית פעמון כך שצורתה הכללית דומה להתפלגות נורמאלית. ממוצע ההתפלגות הוא 0 וכן גם השכיח והחציון. צורת ההתפלגות תלויה בדרגות החופש ולכן אין התפלגות אחת אלא משפחה של התפלגויות, אחת עבור כל מספר אפשרי של דרגות חופש. ככל שמספר דרגות החופש גדל כך מתקרב עקום T להתפלגות הנורמאלית. על מנת לבצע מבחן T עלינו להניח מספר הנחות הכרחיות לביצוע המבחן:

הנחות המבחן הן

דגימה מקרית ובלתי תלויה של התצפיות

השונויות בשתי האוכלוסיות מהן נלקחו המדגמים שוות/ הומוגניות על אף שאינן ידועות לחוקר.

המשתנה מתפלג נורמאלית באוכלוסייה

כמו שלמדתם בתחילת הסמסטר הנוכחי בקורס הסקה סטטיסטית יש הבדל בין השאלות הבאות:

- האם יש הבדל ב"מספר שעות הצפייה הממוצע" בימים שני וחמישי? (מדגמים תלויים)
- האם השתייכות לחברת הכבלים משפיעה על "מספר שעות הצפייה הממוצע" בימי ראשון? (מדגמים בלתי תלויים)

על מנת לענות על כל אחת מהשאלות הללו, צריך לבצע השוואה בין ממוצעים: או שמשווים בין ממוצעי הציונים של שתי קבוצות נבדקים שונות (לדוג' – מנויי חברת Yep ומנויי חברת Hot), או שמשווים בין שני ציונים שונים של אותה קבוצה. בכל מקרה יש לבצע השוואה בין ממוצעים

(t-test). השאלה היא אם מדובר במבחן T להשוואה בין ממוצעים של מדגמים תלויים או בלתי תלויים.

- השאלה הראשונה (האם יש הבדל במספר שעות הצפייה הממוצע בימים שני וחמישי?) דורשת השוואה בין תוצאות צפייה של יומיים עבור אותם אנשים – משום כך ההשוואה בין תוצאות הצפייה היא בין ממוצעים של מדגמים תלויים. (אותם נבדקים = מדגמים תלויים).

- השאלה השנייה דורשת השוואה בין שתי קבוצות נבדקים שונות, מנויים של חברות כבלים שונות, לגבי "שעות הצפייה הממוצעות ביום ראשון" – משום כך ההשוואה בין התוצאות היא השוואה בין ממוצעים של מדגמים בלתי תלויים. (נבדקים שונים בכל רמה של המשתנה הבלתי תלוי = מדגמים בלתי תלויים).

הדגמה ושימוש

SPSS

בתרגיל נשתמש בקובץ הנתונים (General data file_2007.xls). אנא ייבא/י את הקובץ לתוכנת ה-SPSS.

להלן תיאור המשתנים בקובץ:

נבדקו 80 נשאלים לגבי הרגלי הצפייה שלהם בטלוויזיה. קובץ הנתונים מכיל את המשתנים הבאים:

No – מספר הנשאל

Sex – מין הנשאל (0-זכר, 1-נקבה)

Age – גיל הנשאל

Cable – חברת הכבלים בה הנשאל מנוי (0-לא מנוי של חברת כבלים, 1-מנוי של חברת "Yes", 2-מנוי של חברת "Hot").

Marital status – מצב משפחתי (1-רווק, 2-נשוי, 3-גרש, 4-אלמן)

Net income – הכנסה נטו של הנשאל בחודש (בשקלים)

Children – מספר הילדים של הנשאל

Hours sun – מספר שעות הצפייה הממוצע בימי ראשון.

Hours mon – מספר שעות הצפייה הממוצע בימי שני.

Hours tues – מספר שעות הצפייה הממוצע בימי שלישי.

Hours wed – מספר שעות הצפייה הממוצע בימי רביעי.

Hours thur – מספר שעות הצפייה הממוצע בימי חמישי.

Hours fri – מספר שעות הצפייה הממוצע בימי שישי.

Hours sat – מספר שעות הצפייה הממוצע בימי שבת.

No. of favorite shows – מספר התכניות האהובות על הנשאל.

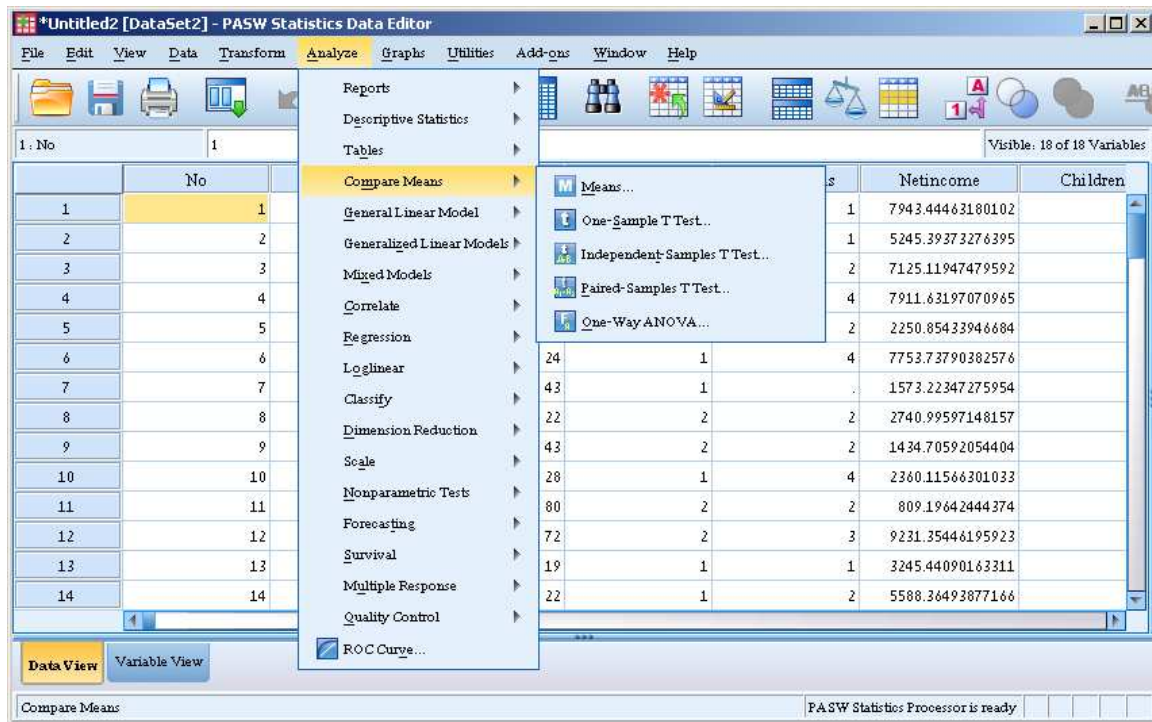
No. on – מספר הפעמים שהנשאל מדליק את הטלוויזיה ביום.

No. hobbies – מספר התחביבים האחרים שיש לנשאל, מלבד צפייה בטלוויזיה.

Advert – הקרנת פרסומות לנבדק (1-כן, 0-לא). לחלק מהנשאלים הוקרנו פרסומות ולחלק לא, לצורך בדיקת השפעת הקרנת הפרסומות על המשתנים האחרים.

כיצד משווים בין ממוצעים?

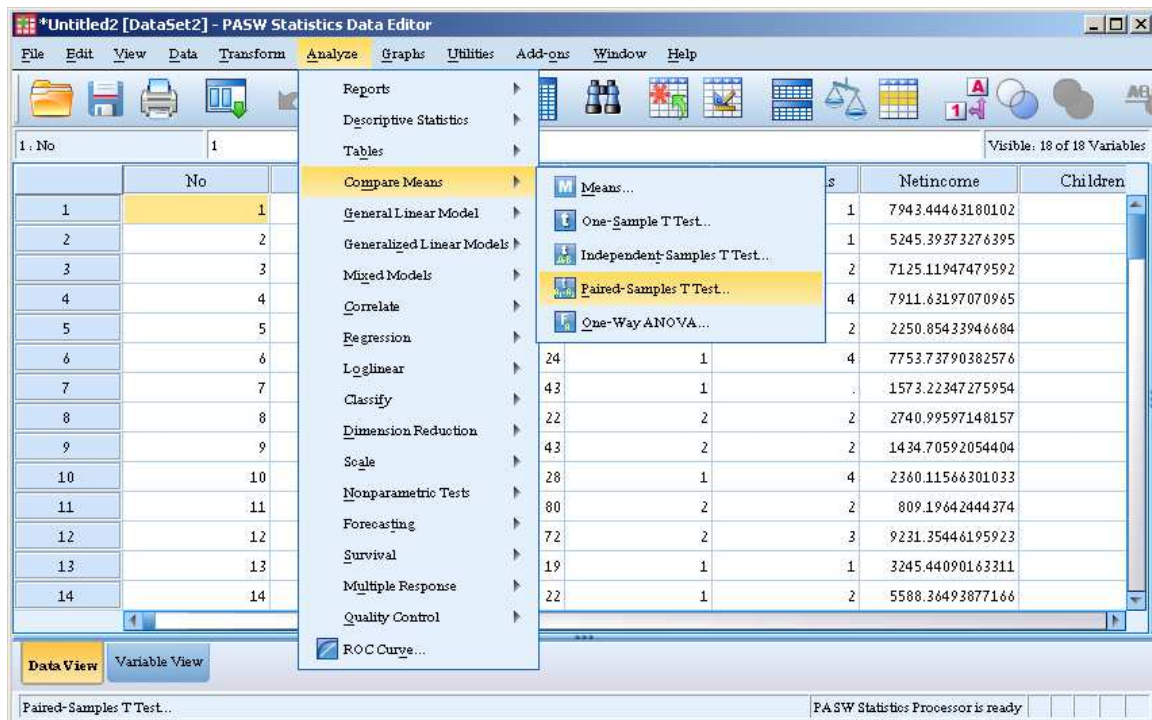
Analyze > compare means



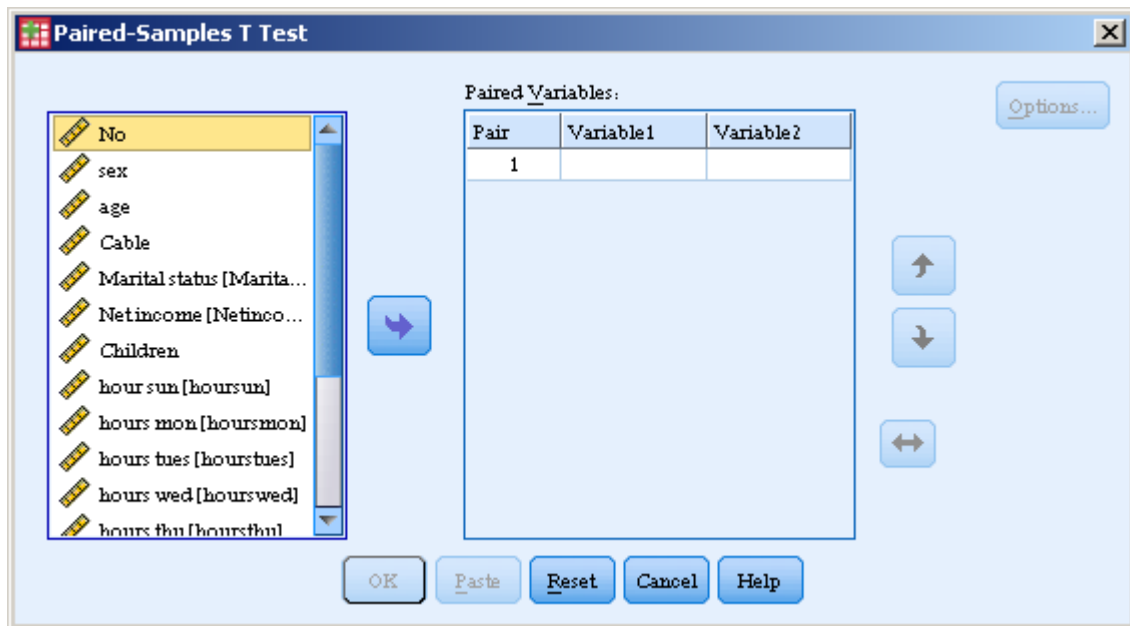
ישנן חמש אפשרויות להשוואות ממוצעים תחת אפשרות זו. האפשרות הראשונה (means) היא היחידה שרק מציגה ממוצעים, ללא ביצוע מבחן מובהקות סטטיסטי. האפשרות האחרונה ('One Way Anova') מאפשרת לנו לבצע ניתוח שונות חד גורמי (מבחן F). שלוש האפשרויות האחרות מאפשרות לנו לבצע מבחני T שונים.

נתחיל בשאלה הראשונה הבוחנת האם יש הבדל ב"מספר שעות הצפייה הממוצע" בימים שני וחמישי? (מדגמים תלויים). מכון שמדובר במדגמים תלויים – האפשרות המתאימה היא:

.Paired-Samples T-test



כעת יפתח החלון הבא:

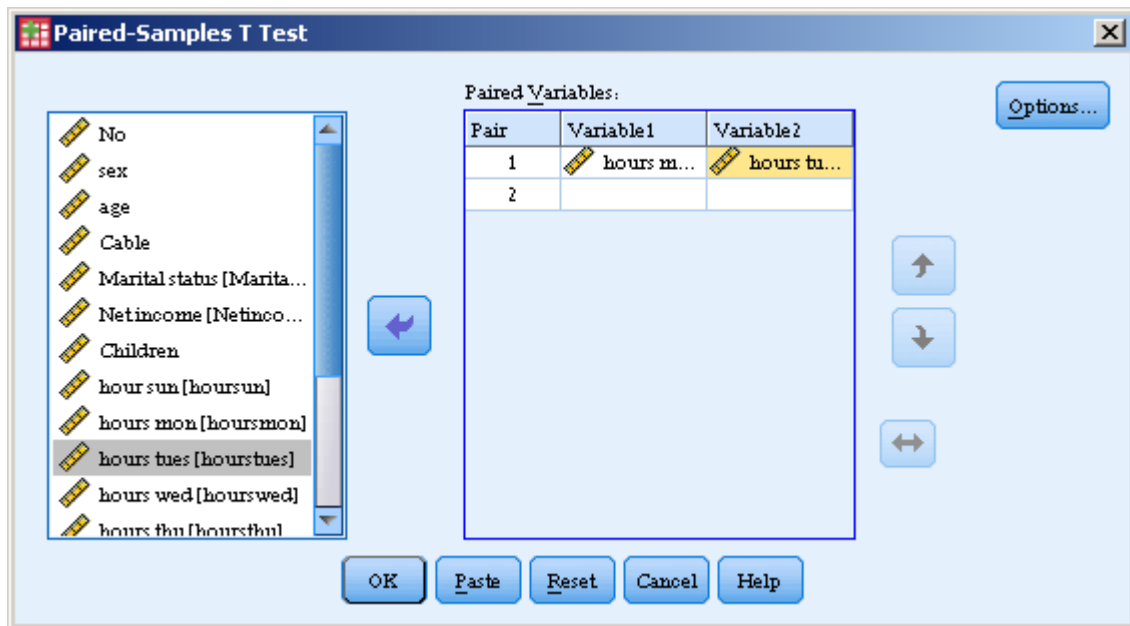


על מנת לבדוק אם אכן קיים פער ב"מוצע הצפייה בין יום שני וחמישי", נערוך השוואה בין ממוצעי הציונים בשניהם.

מסמנים 'Hours mon'

בחר 'Hours thur',

לאחר בחירת כל משתנה לחץ על החץ השחור במרכז.



לחצו על OK לביצוע

במקש ה- 'options' אפשר להגדיר גם את גודל הטעות (מסוג אלפא) שאנו מעדיפים, כאשר ברירת המחדל היא 0.05.

כעת נפתח בפנינו דף הפלט המחולק למספר טבלאות.

הטבלה הראשונה מציגה סטטיסטיקה תיאורית בסיסית – ממוצעים, סטיות תקן וכד'.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 hours mon	7.43	77	4.792	.546
hours tues	6.39	77	3.167	.361

הטבלה השנייה מציגה את המתאם בין המשתנים ואת המובהקות שלו.

("ממוצע הצפייה" בשני הימים חושב עבור אותם אנשים, לכן סביר שיהיה ביניהם מתאם והוא נלקח

בחשבון בחישוב הסטטיסטי).

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 hours mon & hours tues	77	.291	.010

הטבלה השלישית מציגה את מבחן ה-T עצמו.

בתחילה יש הצגה של נתונים תיאוריים לגבי הפער בין הממוצעים. בהמשך מוצג מבחן המובהקות:

התוצאה של מבחן ה-T, דרגות החופש שבהם נעשה שימוש והמובהקות.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	hours mon - hours tues	1.039	4.916	.560	-.077	2.155	1.854	76	.068

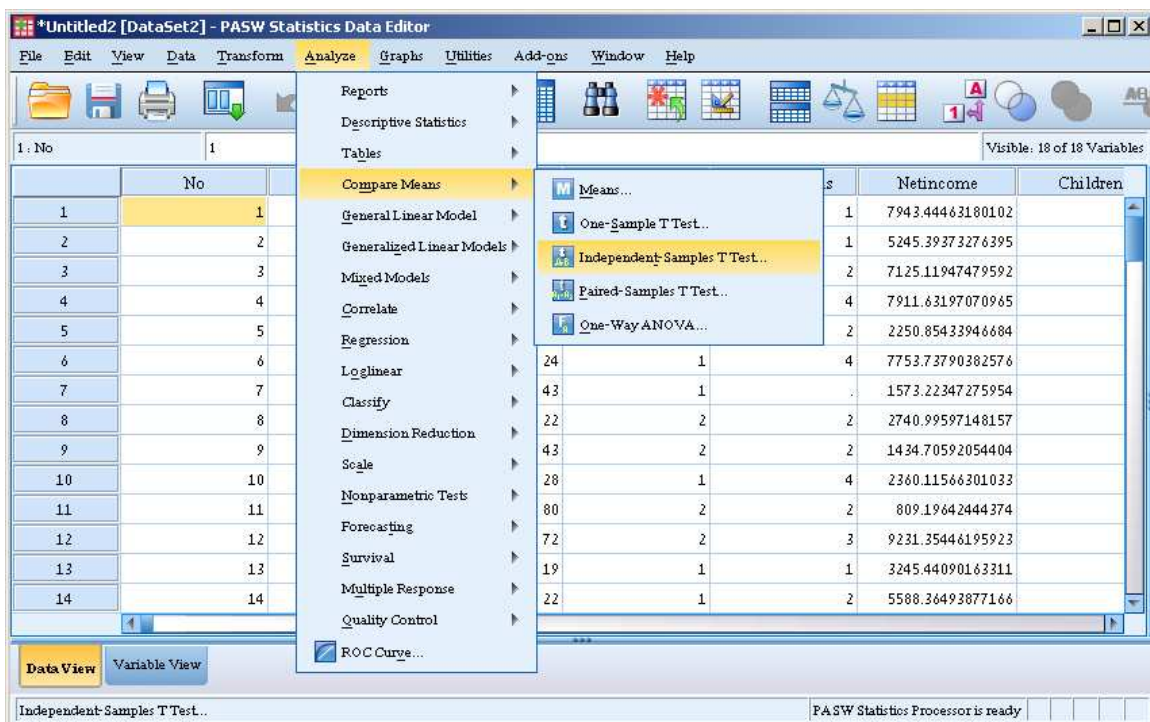
*ש1. האם על פי הטבלה יש הבדל ב"מספר שעות הצפייה הממוצע בין הימים שני וחמישי" ?

נעבור לשאלה השנייה: האם יש הבדלים בין מנויי חברות הכבלים השונות (Hot , Yep) ב"שעות הצפייה הממוצעות ביום ראשון"? מכיוון שמדובר במדגמים בלתי תלויים האפשרות המתאימה היא

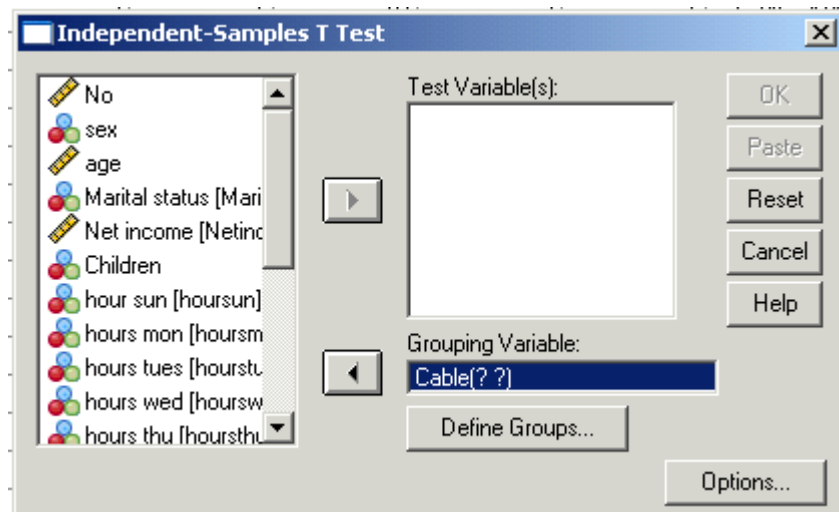
Independent-Samples T Test

כדי לבצע את המבחן צריך לחזור בסרגל הכלים לפקודה והפעם לבחור מבחן T למדגמים בלתי תלויים:

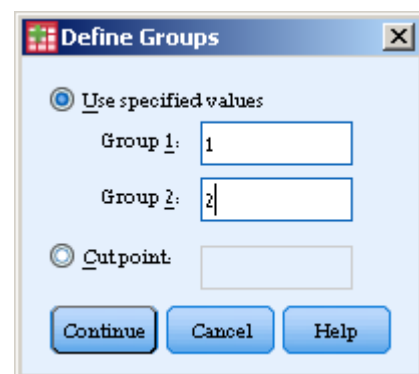
...analyze > compare means > Independent-Samples T Test



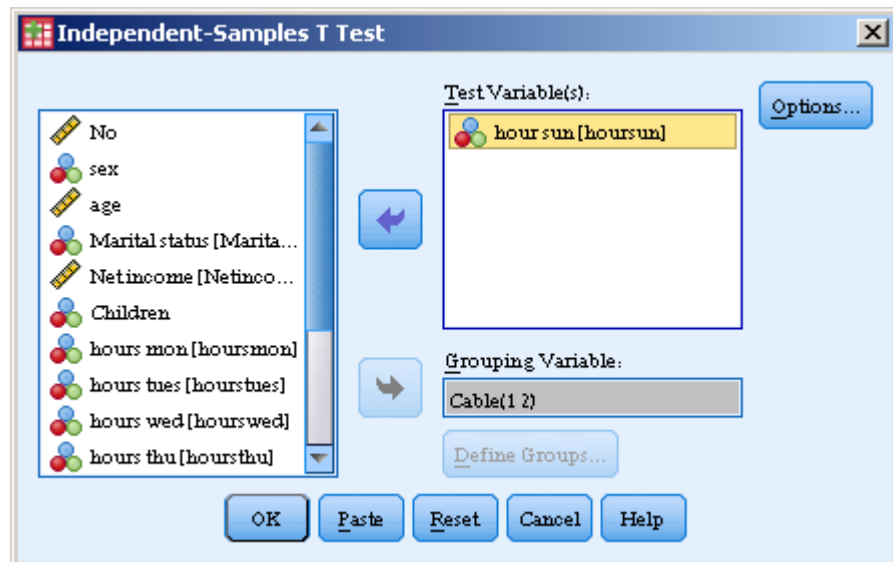
במקרה הזה אנחנו רוצים להשוות בין קבוצות שתי קבוצות, מנויי חברת Hot ומנויי חברת Yep, ב"שעות הצפייה הממוצעות שלהם ביום ראשון".
 בחלון הפקודה שנפתח יש להגדיר שני משתנים:
 'Grouping variable' – זהו למעשה המשתנה הקטגורי המחלק את הנבדקים לשתי קבוצות נבדקים שונות. תחילה נבחר להשוות בין מנויי חברת Hot ומנויי חברת Yep ולכן נבחר במשתנה "Cable" ונכניס אותו לתיבה.



כפי שאמרנו מבחן T משווה בין שתי קבוצות בלבד ואילו משתנה ה- Cable מכיל שלוש רמות. ניתן לראות כי ליד משתנה ה- Cable מופיעים שני סימני שאלה, סמני השאלה למעשה מודיעים לנו כי התוכנה אינה יודעת אילו מבין רמות המשתנה עליה להשוות וכי עלינו להגדיר זאת לה. עלינו להגדיר את הערכים שמייצגים את שתי הקבוצות שאנו רוצים להשוות, הקש על 'Define Groups': והגדר:
 Hot -2, Yep -1



לחץ על 'continue', ובחלון הראשי של הפקודה. כעת ניתן לראות כי במקום סמני השאלה מופיעים הרמות שבחרנו למשתנה Cable. כעת הגדר את ה- Test variables. אלו הם למעשה הימים השונים עבורם אנו מעוניינים לבצע את ההשוואה. בחר בחלון זה את "Hours Sun".



הרץ את הפקודה.

כעת קבלנו קובץ תוצאות המכיל שתי טבלאות. בחלק הראשון של חלון התוצאות אפשר לראות את הסטטיסטיקה התיאורית הקשורה להשוואות השונות.

Group Statistics

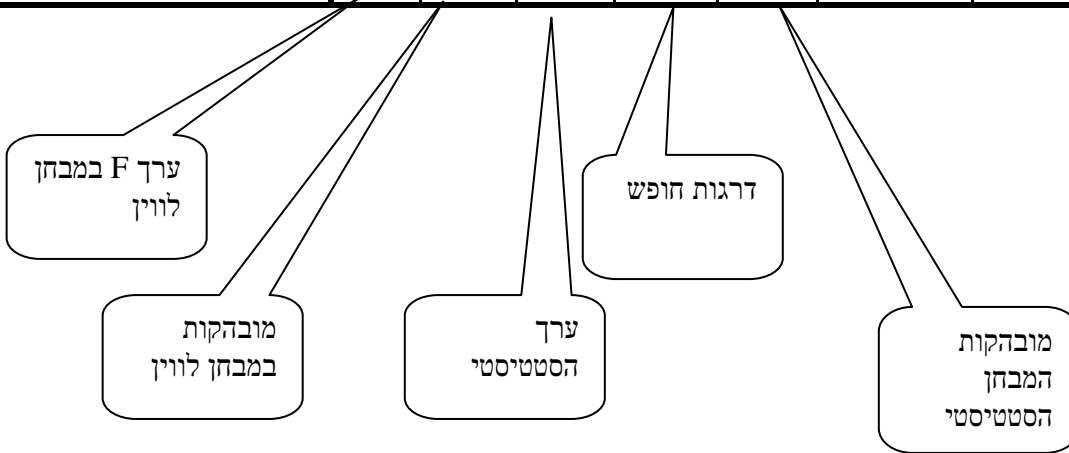
	Cable	N	Mean	Std. Deviation	Std. Error Mean
hour sun	1	38	3.95	2.856	.463
	2	29	5.24	3.291	.611

***ש2. מי מבין הקבוצות צופה בממוצע יותר שעות בטלוויזיה ביום שני?**

הטבלה השנייה קצת יותר גדולה. הטבלה גדולה יותר מכיון שהיא מפרטת תוצאות של שני מבחנים. כזכור – במבחן T להשוואות בין מדגמים בלתי תלויים, אחת ההנחות היא של הומוגניות של שונות. אז במקרה זה התוכנה מספקת לנו מבחן נוסף על מנת שנדע האם אנחנו יכולים להניח הומוגניות כזו או לא. שתי העמודות הראשונות הן חלק ממבחן Levine שבוחן את הנקודה הזו בדיוק. העמודה הראשונה מספקת את הסטטיסטי (במקרה זה F) ואת המובהקות שלו. אם התוצאה היא מובהקת (p קטן מ-0.05) – סימן שאי אפשר להניח הומוגניות של שונות (משמע, יש הבדל מובהק במידת ההומוגניות של השונות בקבוצות).

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
hour sun	Equal variances assumed	2.037	.158	-1.720	65	.090	-1.294	.752	-2.797	.209
	Equal variances not assumed			-1.687	55.552	.097	-1.294	.767	-2.831	.243



התוכנה מתגברת על בעיית הנחת ההומוגניות של שונויות בכך שהיא מבצעת שני מבחני T. אחד כאשר היא מניחה הומוגניות כזו (השורה העליונה בכל התאים המתייחסים למבחן T) ומבחן מקביל (עם תיקון) כאשר היא מניחה שהשונויות הן שונות (השורה התחתונה בכל התאים המתייחסים למבחן T). ואז כרגיל היא מציגה את הסטטיסטי, דרגות החופש והמובהקות.

ש*3. האם במבחן המוצג יש להשתמש במבחן T רגיל או במבחן מתוקן למצב בו לא

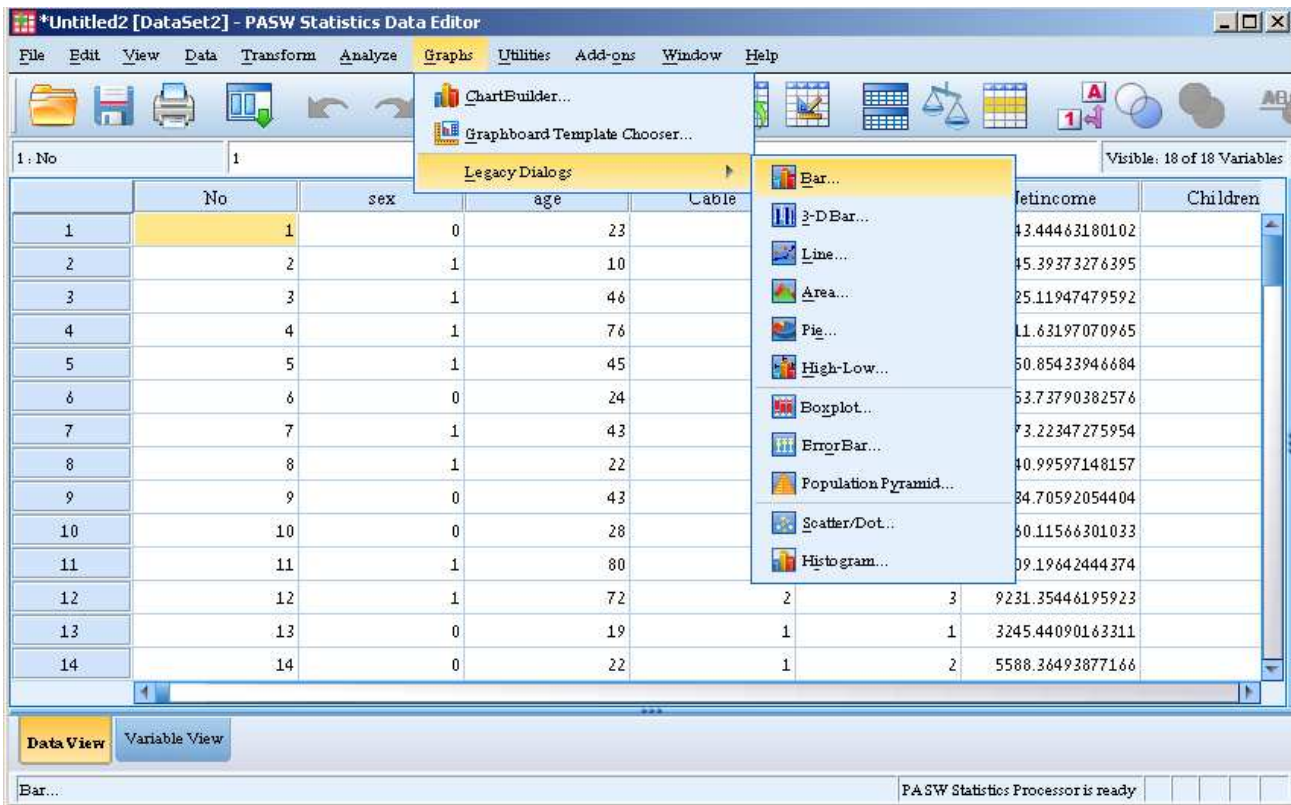
ניתן להניח הומוגניות של שונויות?

ש*4. האם המבחן מובהק?

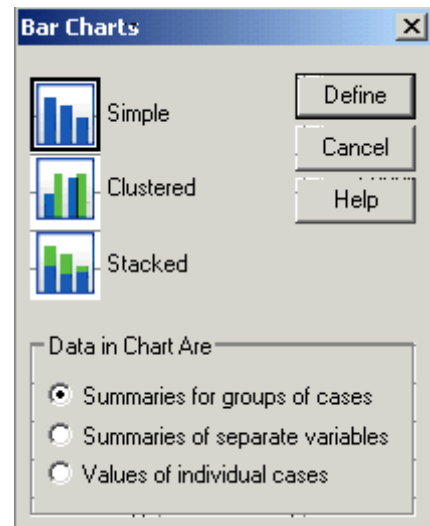
הצגה גראפית

הגרף המתאים להצגת השוואות בין ממוצעים הוא גרף מסוג Bar או Box. על מנת להציג גרף כזה היכנסו ל:

graphs > legacy Dialog > bar



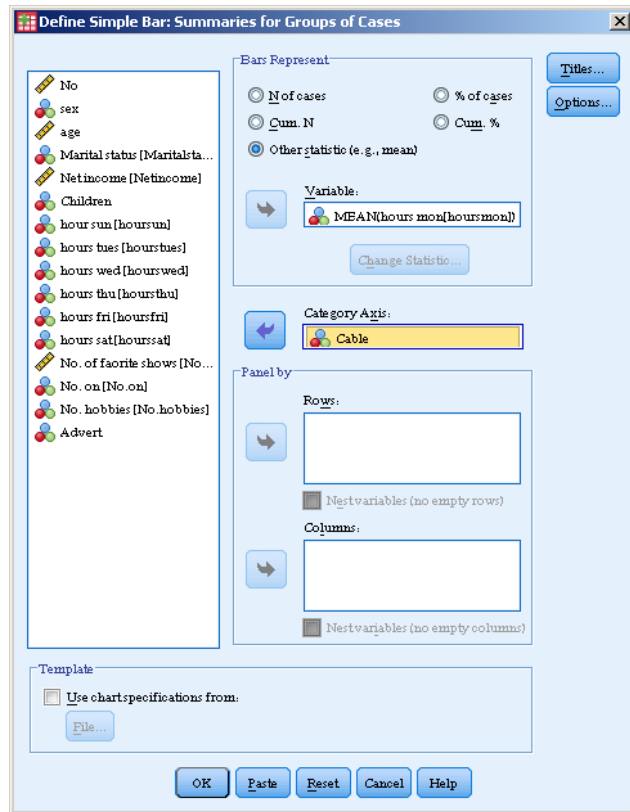
בחלון שיפתח יש לסמן Summaries for groups of cases במקרה ואנו מעוניינים להציג גרף של משתנים קטגוראליים (שתי קבוצות נבדקים שונות) ו Summaries of separate variables כאשר נרצה להשוות שני משתנים המכילים אותם נבדקים. כעת ניצור גרף לשאלה הראשונה וכן נבחר באופציה Summaries for groups of cases.



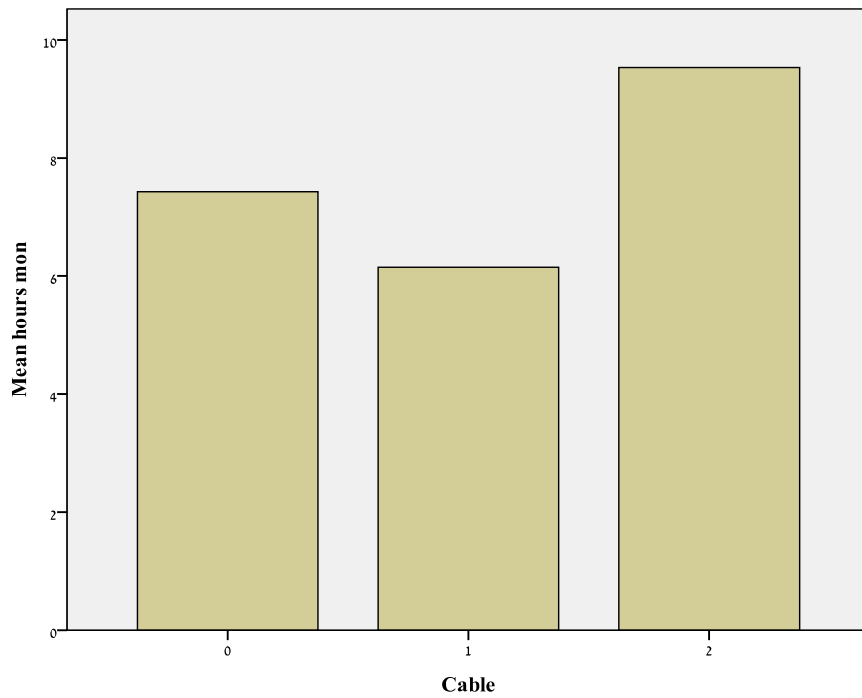
בחלון שיפתח, בחלק העליון של התיבה מסמנים את מה שאנו מעוניינים שעמודות הגרף ייצגו. האפשרויות לסימון הן: מספר מקרים, מספר מקרים מצטבר, אחוז מקרים, אחוז מצטבר של מקרים, ו 'other summary functions': תחת אופציה זו אפשר לבחור פונקציה מתמטית לסיכום הנתונים (ממוצע, שונות, סטיית תקן, ערך מקסימלי, ערך מינימלי ועוד).

בחלון 'variable' נכניס את המשתנה התלוי.

חלון 'category axis' נכניס את המשתנה הבלתי תלוי (המשתנה שמגדיר את החלוקה לשתי קבוצות).

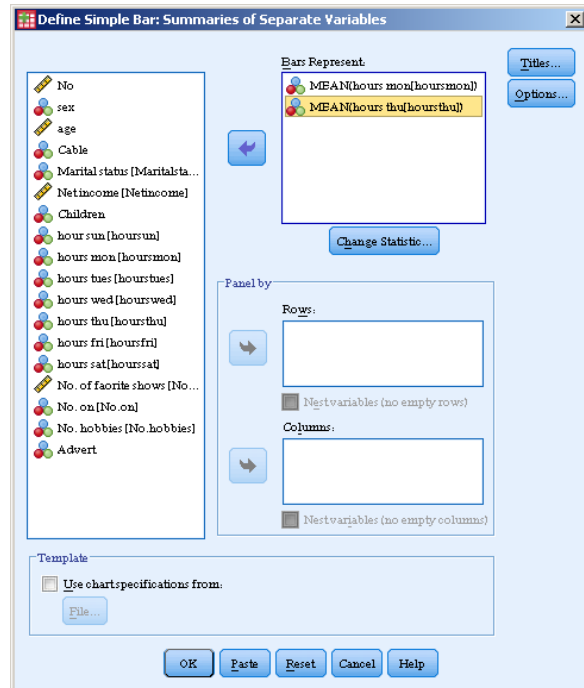


כעת נלחץ על OK על מנת ליצור את הגרף:

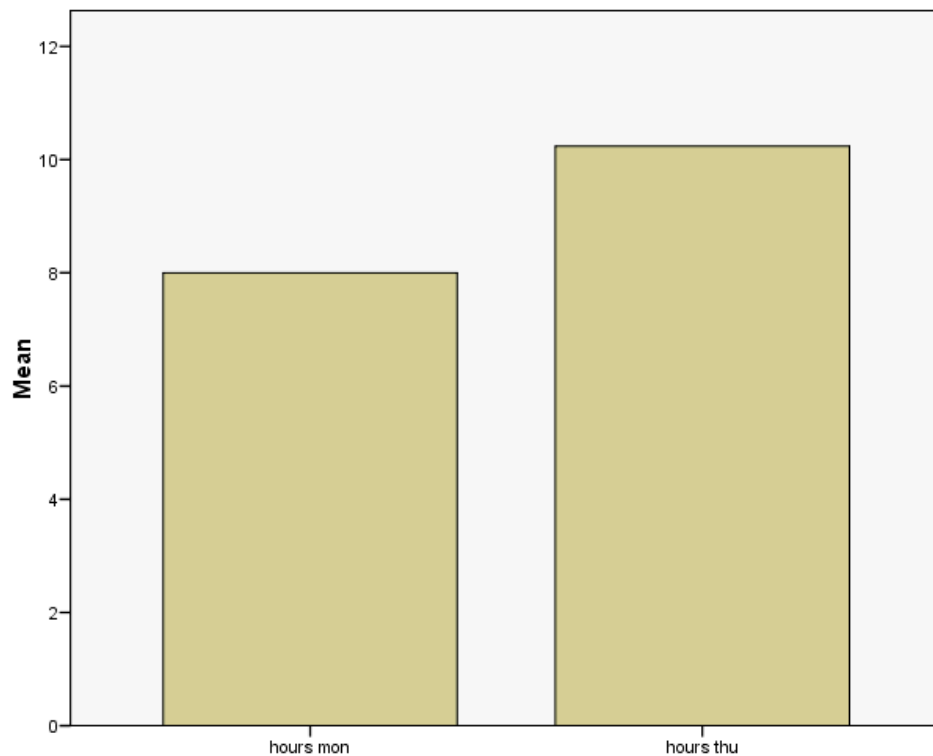


על מנת ליצור גרף לשאלה השנייה נחזור לאופציה של Graphs בסרגל הכלים הראשי. שוב נבחר גרף Bar. כעת נסמן את האופציה: Summaries of separate variables משום שאנו רוצים גרף למדגמים תלויים.

נבחר את המשתנים שעות צפייה ביום שני ושעות צפייה ביום חמישי ונעבירם לחלון האמצעי:



כעת נלחץ על OK ויתקבל הגרף.



תרגיל הגשה מספר 6

1. השתמשו בקובץ General data file_2007.sav
2. באמצעות הניתוחים הסטטיסטיים שלמדתם, צרו קובץ תוצאות (output) שיכיל טבלאות, שבהן נתונים העונים על השאלות הבאות. הדפיסו את קובץ התוצאות (מתוך חלון ה-output ניתן לבחור בתפריט File ואז Print). לבסוף, צרפו דף נוסף, ובו התשובות המספריות בלבד לשאלות שלהלן:
 - א. בצע מבחן שיבדוק האם קיים הבדל מובהק בין נשים לגברים במספר התוכניות האהובות עליהם. הוסף הצגה גראפית של התוצאות.
 - ב. בצע מבחן שיבדוק האם קיים הבדל מובהק בין ממוצע שעות הצפייה בימים ראשון ורביעי.
 - ג. האם יש הבדל בין נשאלים שצפו בפרסומת לאלו שלא צפו בה, בממוצע שעות הצפייה שלהם ביום שלישי? בצע את המבחן המתאים והוסף הצגה גראפית של התוצאות.

שיעור מספר 7

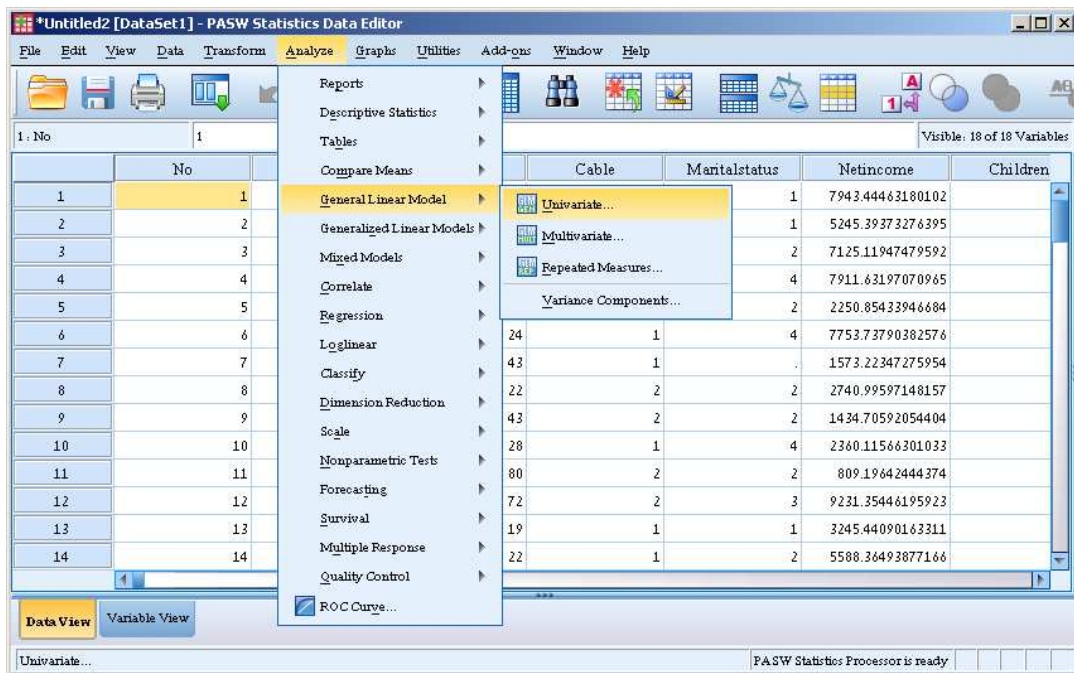
שיעור מס' 7: ניתוח שונות חד-גורמי (מבחן F) SPSS

ניתוח שונות חד-גורמי (מבחן F להשוואת ממוצעים) נועד לבדוק האם קיים הבדל בין הממוצעים של 2 אוכלוסיות או יותר.

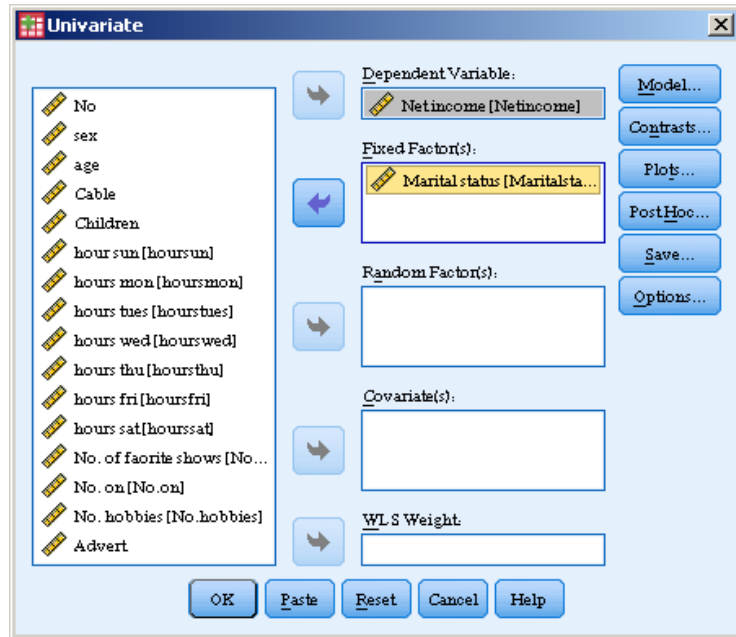
בתרגיל נשתמש בקובץ הנתונים (General data file_2008.xls). אנא ייבא/י את הקובץ לתוכנת ה-SPSS.

נניח שהחוקר מעוניין לבדוק האם יש הבדלים בהכנסה הממוצעת בין אנשים עם מצב משפחתי שונה? כדי לבדוק שאלת מחקר זו, עליו להשוות את ההכנסה הממוצעת (Netincome) בין הרמות השונות של המצב המשפחתי (Maritalstatus) בסרגל הכלים יש לבחור:

'Analyze' > 'general linear models' > 'univariate'



בחלון שיפתח יש לבחור את המשתנה התלוי והבלתי-תלוי.



יש להכניס את המשתנה התלוי בחלון ה-Dependent Variable, ואת המשתנה הבלתי-תלוי בחלון Fixed Factor(s), ואז ללחוץ OK. מתקבלות שתי טבלאות. הטבלה התחתונה היא טבלת הניתוח הסטטיסטי:

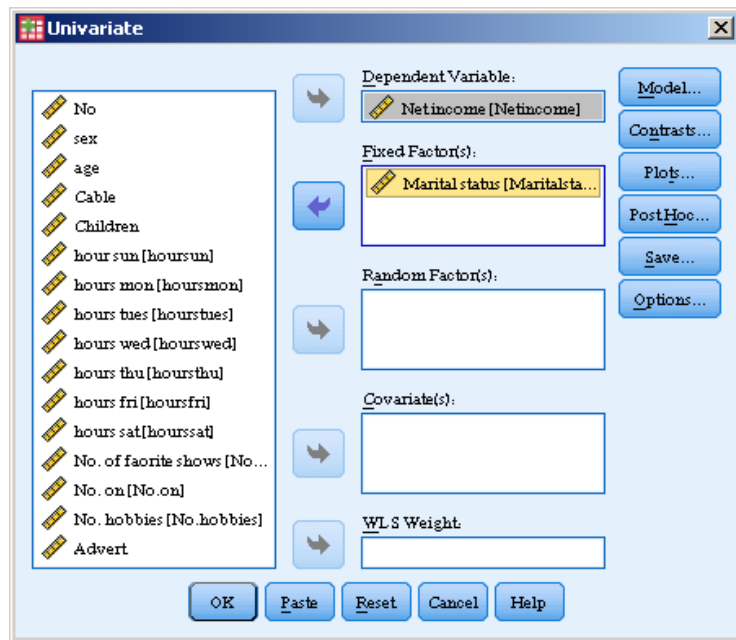
Tests of Between-Subjects Effects

Dependent Variable:Net income

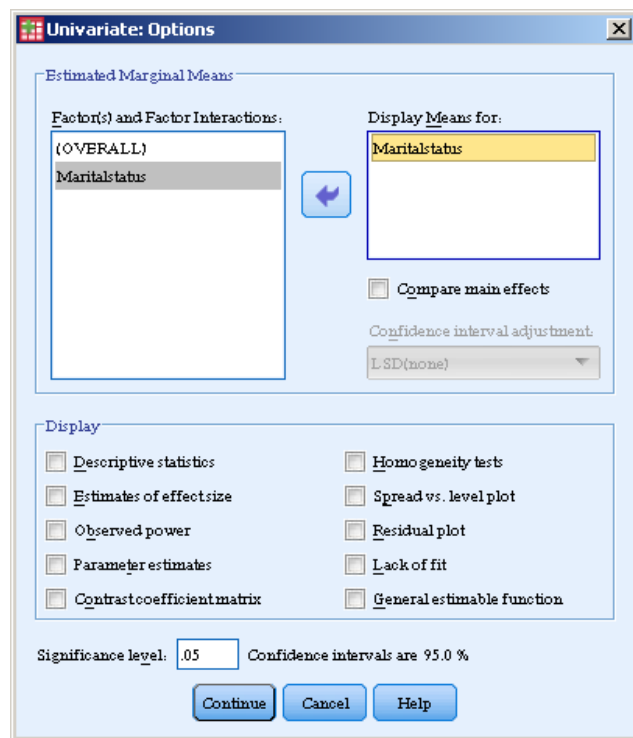
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	21054653.155 ^a	3	7018217.718	1.021	.389
Intercept	1305709355.573	1	1305709355.573	189.919	.000
Maritalstatus	21054653.155	3	7018217.718	1.021	.389
Error	474381041.979	69	6875087.565		
Total	2241609052.016	73			
Corrected Total	495435695.134	72			

a. R Squared = .042 (Adjusted R Squared = .001)

בשורה השלישית מתקבל המידע על המשתנה הבלתי-תלוי. ניתן לראות כי ערך ה-SS הוא 21054653.2. בנוסף מוצג ערך ה-F (1.021), ערך המובהקות (0.389) ודרגות החופש (3). בשורה הרביעית מופיעים הערכים של גורם הטעות: ערך ה-SS (474381042), MS (6875087.565) ודרגות החופש (69). הדיווח המלא יראה כך: [F (3,69) = 1.021, p=0.389]. כעת נרצה לראות את הממוצעים: לשם כך, יש לחזור לחלון זה:



יש להיכנס ל 'options' ותחת 'factor(s) and factor interactions' לבחור במשנה הבלתי-תלוי ולהעביר העבר אותו בעזרת החץ לחלון שממולו (חלון הנקרא 'display means for').



לאחר מכן יש ללחוץ Continue ואז OK.

כעת קיבלנו טבלה נוספת הכוללת את הממוצעים של המשתנה התלוי עבור הקבוצות השונות.

Marital status

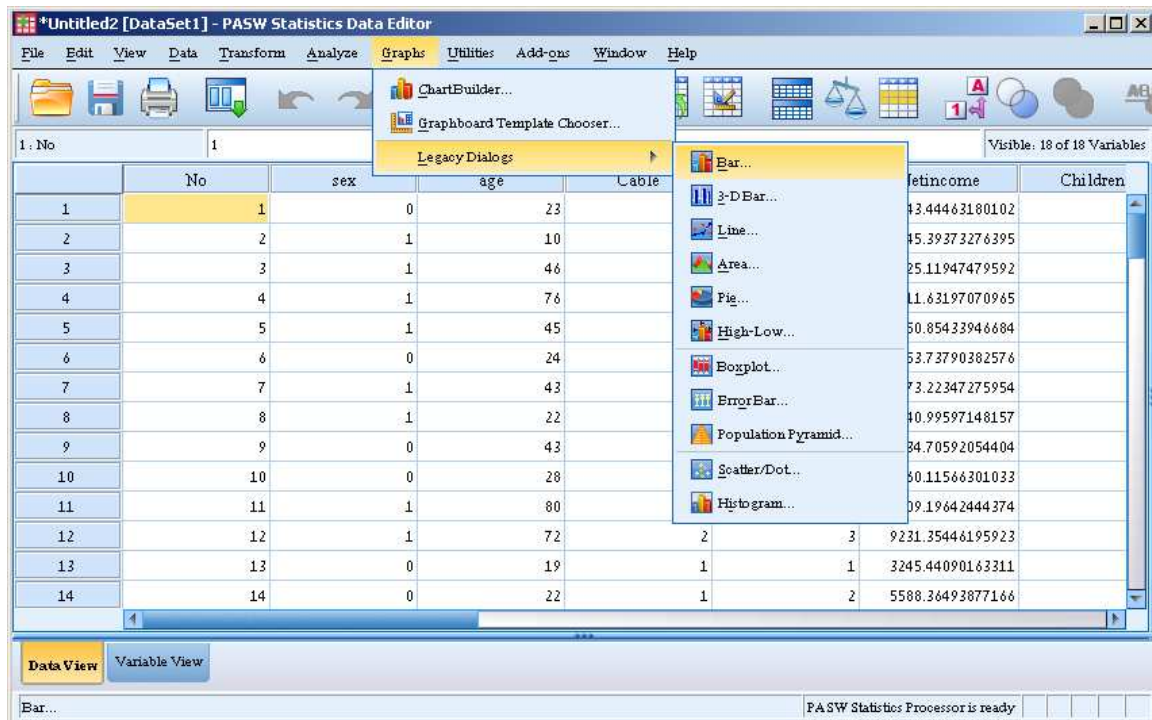
Dependent Variable: Net income

Marital status	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	5215.678	559.021	4100.462	6330.893
2	4435.261	443.205	3551.090	5319.431
3	4872.945	874.013	3129.337	6616.554
4	6170.684	991.038	4193.618	8147.750

* הצגת התוצאות בגרף.

ניתן לבחור בין שני סוגי גרפים 'bar' בו התוצאות יוצגו בעמודות ו'line' בו התוצאות יוצגו בקו רציף (יש לזכור כי הרצף חסר משמעות ומה שמשנה זה הערך על ציר ה Y בכל נקודה ספציפית על ציר ה X).

הצגה באמצעות גרף מסוג "bar".



בסרגל הכלים בחר 'simple' ולחץ על 'define', סמן את האפשרות 'other summary function', וב 'variable' בחר במשתנה התלוי, וב 'category axis' במשתנה הבלתי-תלוי ולחץ o.k.

הצגה באמצעות גרף "line"

The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with columns: No, sex, age, Cable, netincome, and Children. The 'Line...' option is selected in the 'Legacy Dialogs' menu.

No	sex	age	Cable	netincome	Children
1	1	0	23	13.44463180102	
2	2	1	10	15.39373276395	
3	3	1	46	25.11947479592	
4	4	1	76	11.63197070965	
5	5	1	45	30.85433946684	
6	6	0	24	33.73790382576	
7	7	1	43	73.22347275954	
8	8	1	22	40.99597148157	
9	9	0	43	34.70592054404	
10	10	0	28	30.11566301033	
11	11	1	80	19.19642444374	
12	12	1	72	9231.35446195923	2
13	13	0	19	3245.44090163311	1
14	14	0	22	5588.36493877166	2

בחרו 'simple' ולחצו על 'define', סמן את האפשרות 'function' 'other summary', ב 'variable' בחר במשתנה התלוי, וב 'category axis' במשתנה הבלתי-תלוי ולחץ o.k.

תרגיל הגשה מספר 7

1. השתמשו בקובץ General data file_2008.sav
2. באמצעות הניתוחים הסטטיסטיים שלמדתם, צרו קובץ תוצאות (output) שיכיל טבלאות, שבהן נתונים העונים על השאלות הבאות. הדפיסו את קובץ התוצאות (מתוך חלון ה-output ניתן לבחור בתפריט File ואז Print). לבסוף, צרפו דף נוסף, ובו התשובות המספריות בלבד לשאלות שלהלן:
 - א. בצע מבחן שיבדוק האם קיים הבדל מובהק במספר שעות הצפיה הממוצע בטלביזיה בימי שישי, בין גברים לנשים. בצע מבחן זה בשתי דרכים שונות.
 - ב. צור משתנה חדש, שיכיל את המספר הממוצע של שעות הצפיה השבועי של כל נשאל. לאחר מכן, בדוק האם יש הבדל במשתנה זה בין הרמות השונות של המשתנה Cable.
 - ג. הצג את התוצאות של סעיף ב' בשתי דרכים שונות.

