

פרק 5. סטטיסטיקה תיאורית (Descriptive Statistics)

5.1 התפלגות שכיחויות (Frequencies)

5.1.1 כללי

התפלגות שכיחויות מתארת את הערכים הקיימים של המשתנים והשכיחות שלהם, ומאפשרת הפקה של סטטיסטיקה תיאורית נוספת כגון מדדי מרכז (ממוצע, חציון, שכיח, סכום), מדדי פיזור (שונות, סטית תקן, מינימום ומקסימום, מדדים למיקום יחסי כגון אחוזונים), ועוד.

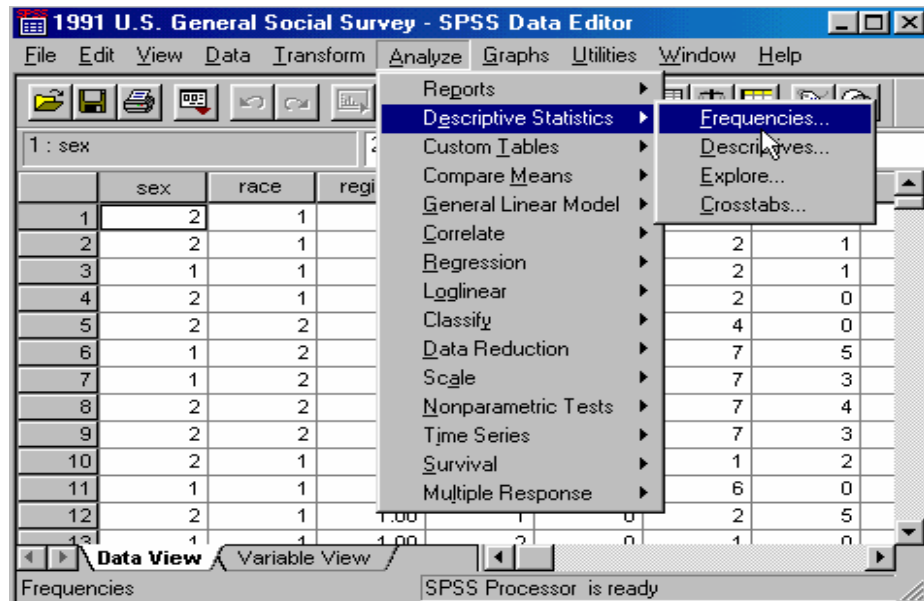
השימוש בהליך של התפלגות השכיחויות מיועד לשתי מטרות עיקריות:

1. בדיקה ומציאת טעויות בהקלדת הנתונים. הפלט מתאר את ערכי המשתנה כפי שהוזנו למחשב, והשכיחות (במספרים ובאחוזים) של כל ערך. לדוגמא, במשתנה כמו מין, אשר הערכים האפשריים שלו הינם 1 (זכר) או 2 (נקבה), לא יתכנו ערכים מספריים אחרים. במידה ומתקבלים בהתפלגות ערכים שונים מאלו, ניתן להבין שקיימת שגיאה בהקלדת הנתונים.
2. תיאור התפלגות המשתנה, הערכים האפשריים שלו ופיזורם, כיצד נראית ההתפלגות. ניתן לתאר את ההתפלגות באמצעות טבלה, דיאגרמה ומדדים תיאוריים (מדדי מרכז, פיזור).

Analyze →

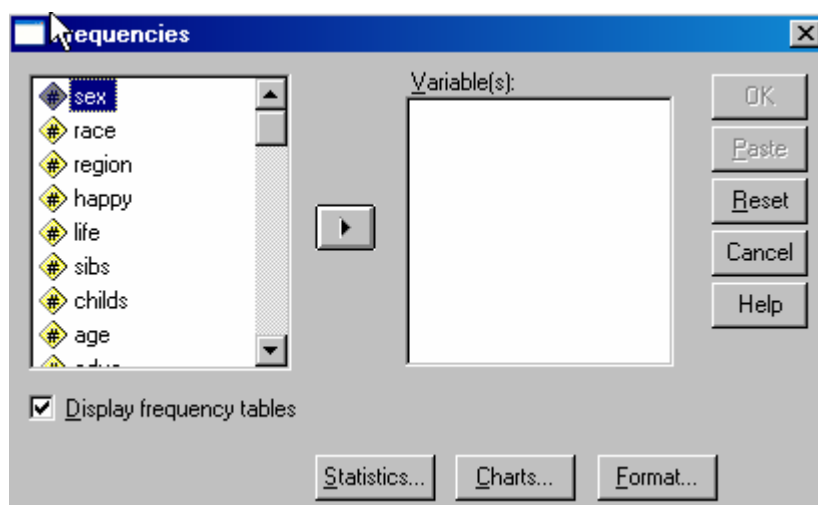
Descriptive Statistics →

Frequencies



תרשים 5.1. בחירת ההליך Frequencies מתוך התפריט הראשי.

לאחר שבחרת את ההליך, תתקבל תיבת השיח הבאה :



תרשים 5.2. תיבת שיח ראשית של ההליך Frequencies.

בצד שמאל רשימת כל המשתנים הקיימים בקובץ (שים לב שניתן לדפדף למטה ולמעלה, לא כל המשתנים מוצגים במסך הנוכחי). בצד ימין, תחת הכותרת **Variable(s)**, המשבצת אשר אליה מעבירים את המשתנים המיועדים לעיבוד.

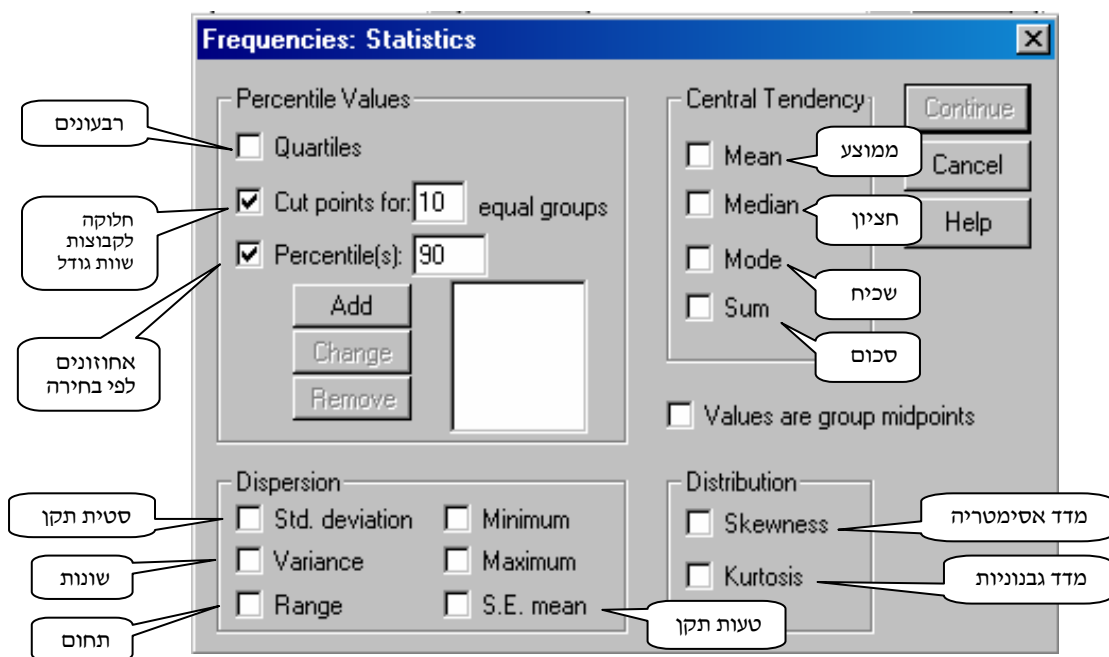
בחירת המשתנים הרצויים לעיבוד: בחרי/י על ידי לחיצה עם העכבר, את המשתנה הרצוי. המשתנה הנבחר יודגש בצבע כחול, והחץ בין שתי המשבצות, יופנה ימינה. העבר/י את המשתנה שבחרת, על ידי לחיצה על החץ הפונה ימינה. ניתן להעביר מספר משתנים בפעולה אחת.

לאחר העברת משתנים (לפחות אחד), תודגש המילה **OK** בתיבת שיח זו. ניתן לאשר ביצוע ההליך על ידי לחיצה עם העכבר על המילה **OK**.

עם בחירת המשתנה / משתנים וביצוע **OK** תופק טבלה של התפלגות שכיחויות, לכל משתנה שהועבר למשבצת המשתנים.

במידה והינך מעוניין/ת להפיק בפעולה זו במדדים סטטיסטיים נוספים, עליך להיכנס למשבצת אשר מופיעה בה מילת המפתח **STATISTICS**.

5.1.2 אפשרויות בחירה של מדדים סטטיסטיים



תרשים 5.3. תיבת שיח של מדדים סטטיסטיים, מתוך ההליך Frequencies.

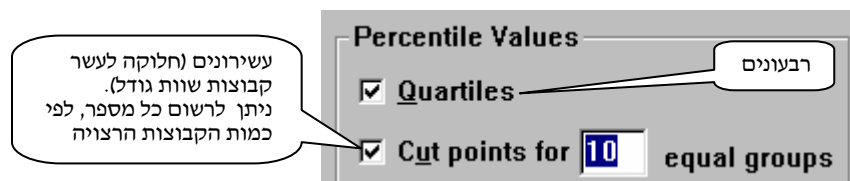
לצד כל אחד מהמדדים הסטטיסטיים משבצת לבנה. בחירת המדד הסטטיסטי נעשית על ידי הקשה עם הלחצן השמאלי של העכבר על המשבצת הרצויה. בעקבות זאת יופיע הסימן $\sqrt{\quad}$ בתוך המשבצת. ביטול: על ידי לחיצה נוספת.

בתוך תיבת השיח של מדדים סטטיסטיים נוספים, קיימות ארבע קבוצות של מדדים סטטיסטיים:

1. **מדדי מרכז (Central tendency)** - ממוצע, חציון, שכיח, סכום.
2. **מדדי פיזור (Dispersion)** - סטית תקן, שונות, טווח, מינימום ומקסימום, טעות תקן.
3. **מדדי התפלגות (Distribution)** - Kurtosis, Skewness.
4. **מדדי מיקום יחסי - אחוזונים (Percentile values)** - המדד רבעונים מפיך שלושה אחוזונים:
 - א. אחוזון 25 (25% מהתצפיות מתחתיו או שוות לו, ו- 75% מהתצפיות מעליו או שוות לו).
 - ב. אחוזון 50, שהינו גם החציון.
 - ג. אחוזון 75.אחוזונים אלו מחלקים למעשה את המדגם לארבעה חלקים שווים (מבחינת כמות נבדקים). ניתן להפיק אחוזון מדויק שמעוניינים בו, באמצעות Cut points או באמצעות Percentile(s).

אופן הפקת Cut points:

לאחר סימון $\sqrt{}$ במשבצת הני"ל, יש לרשום מספר כלשהו במשבצת הבאה:
(כרגע רשום בה המספר 10).

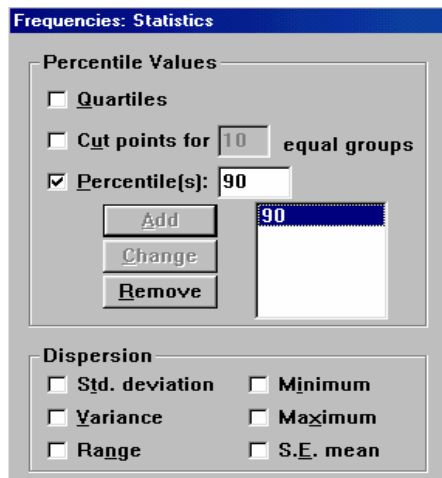


תרשים 5.4. הפקת רבעונים ועשירונים מתוך Statistics בהליך Frequencies.

כתיבת המספר בחלון המתאים יגרום להפקת אחוזונים בפלט, אשר לפיהם ניתן לחלק את קובץ הנתונים לקבוצות שוות גודל, בהתאם למספר שנרשם. לדוגמא, המספר 4 יפיק רבעונים (כפי שמתקבל באפשרות העליונה). המספר 10 יפיק עשירונים, וכך הלאה.

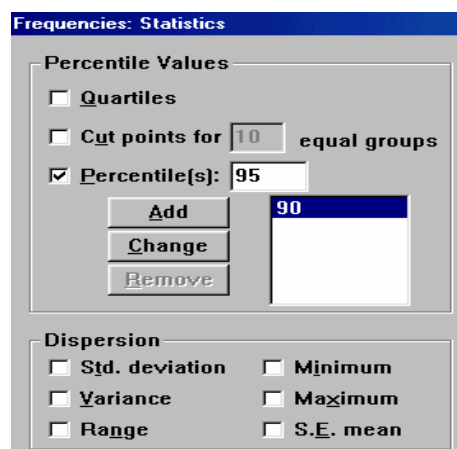
הפקת אחוזונים באמצעות מילת המפתח Percentile(s): לאחר סימון $\sqrt{}$, עליך לרשום במשבצת הרצויה את האחוזון המדויק. לדוגמא, המספר 90 יגרום להפקת האחוזון ה-90. לאחר רישום המספר, יש לבחור על המילה **Add** אשר תודגש בצבע שחור. בחירתך תועבר למשבצת הגדולה אשר מתחת ל Percentile(s). ניתן לחזור על הפעולה כמספר האחוזונים שרוצים.

שינויים בערכים שרשמת: בחר באמצעות הלחצן השמאלי של העכבר, את האחוזון בתוך המשבצת הגדולה אליה הועבר (ראה תרשים 5.5). למחיקה, בחר את המילה Remove המודגשת.



תרשים 5.5. מחיקת האחוזונים הרצויים מתוך Statistics בהליך Frequencies.

לשינוי, יש להביא את העכבר אל המספר 90 אשר רשום מול המילה Percentile(s), ולרשום במקומו מספר אחר (בדוגמא, 90). בחרי את מילת המפתח Change (עדכן שינוי) או Add (הוספה).



תרשים 5.6. שינוי האחוזונים הרצויים מתוך Statistics בהליך Frequencies.

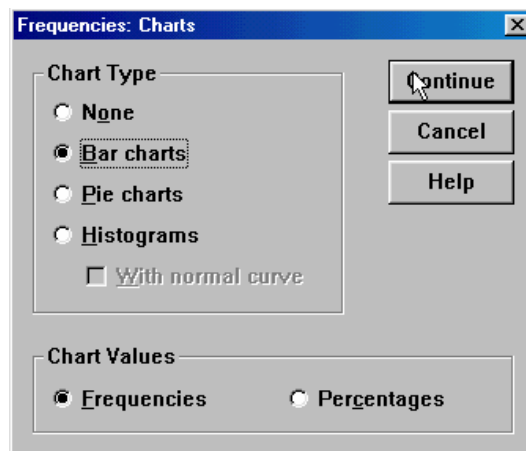
קיימת מידה רבה של חפיפה בין האפשרויות השונות של ערכי האחוזונים (Percentile values). לדוגמא, ערכי הרבעונים המופקים באמצעות מילת המפתח Quartiles, ניתן להפיק גם ב Cut points. כל אחד מהאחוזונים המופקים באמצעות מילות מפתח "מוכנות" (Cut points, Quartiles), ניתן למעשה להפיק גם על ידי הגדרה מדויקת משלך ב Percentile(s).

5.1.3 הפקת דיאגרמות

במסגרת ההליך Frequencies ניתן להפיק שלושה סוגי דיאגרמה, אשר מטרתם להציג גראפית את ההתפלגות של המשתנים.

- **היסטוגרמה – Histogram** – מיועדת למשתנים כמותיים רציפים. ניתן לסמן את האפשרות הרשומה מתחת לסוגי הדיאגרמות: Display Normal Curve, ועל גבי ההיסטוגרמה תוצג עקומת ההתפלגות הנורמאלית. (אפשרי רק בהיסטוגרמה).
- **דיאגרמת עמודות ("מקלות") – Bar chart** – מיועדת למשתנים כמותיים בדידים, או משתנים בסולם מדידה סידורי (אורדינלי). ניתן לקבל דיאגרמת עמודות עם שכחות מוחלטת במספרים (מילת המפתח Frequencies), או באחוזים (מילת המפתח Percentages).
- **דיאגרמת "עוגה" – Pie chart** – מיועדת למשתנים שמיים וסידוריים (אורדינליים – דרוגיים).

יש שים לב, שסימון כל אחת מסוגי הדיאגרמות, הינו באמצעות עיגול ליד כל דיאגרמה, ולא ריבוע כפי שהיה בתפריט קודם. משמעות צורת סימון זו היא שניתן לסמן אך ורק אפשרות אחת בו זמנית.



תרשים 5.7. תיבת שיח של הפקת דיאגרמות (Charts) בהליך Frequencies.

קריאת פלט

הדוגמא מתוך קובץ הנתונים : 1991 U.S General Social Survey
 לפניך דוגמא של טבלת התפלגות שכיחויות למשתנה : באיזו מידה הינך מאושר באופן כללי
 (General Happiness - Happy). לשאלה זו היו שלוש תשובות אפשריות :
 1. מאד מאושר (Very happy) 2. די מאושר (Pretty happy) 3. לא כל כך מאושר (Not too happy).
 NA הינו נתון חסר (No Answer).

General Happiness

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very Happy	467	30.8	31.1	31.1
	Pretty Happy	872	57.5	58.0	89.0
	Not Too Happy	165	10.9	11.0	100.0
	Total	1504	99.1	100.0	
Missing	NA	13	.9		
Total		1517	100.0		

פלט 5.1. פלט התפלגות שכיחויות המופק בהליך Frequencies.

הטורים משמאל לימין :

- **Valid** – תשובות אפשריות. (שלוש תשובות בדוגמא להלן).
 ה Total אשר רשום תחת הכותרת Valid הינו סה"כ המשיבים (בדוגמא 1504, בלי נתונים חסרים Missing), ואילו ה Total מתחת ל Missing הוא סה"כ הנבדקים שעובדו בהליך זה (1517).
- **Missing** – כמות הנבדקים שחסרה להם תשובה.
- **Total** – סה"כ משיבים לשאלה, או סה"כ נבדקים במדגם.
- **Frequency** – שכיחות במספרים. לדוגמא, 467 נבדקים ציינו שהם מאד מאושרים.
- **Percent** – שכיחות באחוזים, מתוך כל הנבדקים שעובדו בהליך זה (מתוך 1517). לדוגמא, 30.8% מתוך כל הנבדקים ציינו שהם מאד מאושרים.
- **Valid percent** – שכיחות באחוזים רק מתוך הנבדקים אשר יש להם תשובה, ללא נתונים חסרים (ללא Missing values) (מתוך 1504). לדוגמא, 31.1% מתוך המשיבים לשאלה, הינם מאושרים מאד. 0.9% נבדקים – חסרה להם תשובה.
- **Cumulative percent** – אחוז מצטבר. למספר זה משמעות רק במשתנים כמותיים או דרוגיים. האחוז בכל שורה הוא אחוז הנבדקים אשר נמצאים עד לערך המספרי אשר בשורה זו, ומטה. לדוגמא, 89% מהנבדקים הם מאושרים מאד או די מאושרים.

Statistics

Age of Respondent

N	Valid	1514
	Missing	3
Mean ממוצע		45.63
Std. Error of Mean טעות תקן של הממוצע		.458
Median חציון		41.00
Mode שכיח		35
Std. Deviation סטית תקן		17.808
Variance שונות		317.140
Range טווח		71
Minimum מינימום		18
Maximum מקסימום		89
Sum סכום		69078
Percentiles	25 - רבעון ראשון	32.00
	50 - רבעון שני - חציון	41.00
	75 - רבעון שלישי	60.00
	90 - רבעון תשיעי	73.00

פלט 5.2. פלט מדדים סטטיסטיים נלווים להתפלגות שכיחויות המופק בהליך .Frequencies (המושגים בעברית רשומים בתוך הטבלה).

תרגיל 1 בנושא Frequencies

קובץ העבודה: 1991 U.S. General Social Survey
הקובץ שלפניך מוגדר באנגלית. יש להפוך את כל ה Variable labels, value labels לעברית.

1. הפק/י התפלגות שכיחויות למשתנה מין.

א. מהו אחוז הגברים ומהו אחוז הנשים בהתפלגות זו, מתוך המשיבים

ב. אחוז גברים _____ אחוז נשים _____

2. הפק/י התפלגות שכיחויות למשתנה השכלה (educ) וכן גם מדדים של רבעונים, ממוצע, חציון. השב/י:

א. מהו ממוצע שנות ההשכלה _____

ב. מהי רמת ההשכלה ש 25% נמוכים ממנה _____

ג. מהי רמת ההשכלה ש 25% גבוהים ממנה _____

ד. מהו החציון של רמת השכלה? _____

ה. מה מספר הנבדקים שחסרים להם נתונים על רמת השכלה _____

3. הפק/י התפלגות שכיחויות Frequencies למשתנה (life).

א. מהו אחוז הנבדקים שלא השיבו על השאלה _____

ב. מהו אחוז הנבדקים, מתוך המשיבים, שסבורים שהחיים מלהיבים _____

ג. מהו אחוז הנבדקים, מתוך כל המדגם, שסבורים שהחיים שגרתיים _____

ד. מהו אחוז הנבדקים, מתוך המשיבים, שסבורים שהחיים משעממים _____

4. הפק/י התפלגות שכיחויות למשתנה מספר אחים ואחיות (Sibs), וכן מדדי ממוצע, חציון, רבעונים, אחוזון 80.

א. מהו הממוצע _____

ב. מהו החציון _____

ג. מהו ערכו של רבעון ראשון _____

ד. מהו ערכו של רבעון שלישי _____

ה. מהו ערכו של האחוזון ה 80 _____

5. הפק/י התפלגות שכיחויות למשתנה מספר ילדים Childs.

א. מהו אחוז הנבדקים, מתוך כל המדגם, שיש להם 8 ילדים ויותר _____

ב. מהו אחוז הנבדקים, מתוך המשיבים, שאין להם ילדים כלל _____

ג. מהו אחוז הנבדקים שיש להם עד 3 ילדים כולל _____

6. הפק/י Frequencies למשתנה רמת השכלה של האימא (maeduc), והשב/י:

א. מהו האחוזון ה 90 _____

ב. מהו הממוצע _____

ג. מהו אחוז הנבדקים אשר לאמהותיהם עד 12 שנות השכלה (כולל) _____

פתרון תרגיל 1 בנושא Frequencies

1. הפק'י התפלגות שכיחויות למשתנה מין.

- ג. אחוז הגברים ומהו אחוז הנשים בהתפלגות זו, מתוך המשיבים ?
ד. אחוז גברים : 41.9% אחוז נשים : 58.1%

2. הפק'י התפלגות שכיחויות למשתנה educ (השכלה) וכן גם מדדים של רבעונים, ממוצע, חציון. השב'י :

- (1) ממוצע שנות ההשכלה : 12.88
(2) רמת ההשכלה ש 25% נמוכים ממנה : 12 שנים
(3) רמת ההשכלה ש 25% גבוהים ממנה : 15 שנים
(4) החציון של רמת השכלה : 12 שנים
(5) מספר הנבדקים שחסרים להם נתונים על רמת השכלה : 7 נבדקים

3. הפק'י התפלגות שכיחויות Frequencies למשתנה LIFE.

- ה. אחוז הנבדקים הכולל שלא השיבו על השאלה : 35.4%
ו. אחוז הנבדקים, מתוך המשיבים, שסבורים שהחיים מלהיבים : 44.3%
ז. אחוז הנבדקים, מתוך כל המדגם, שסבורים שהחיים שגרתיים : 33.3%
ח. אחוז הנבדקים, מתוך המשיבים, שסבורים שהחיים משעממים : 4.2%

4. הפק'י התפלגות שכיחויות למשתנה מספר אחים ואחיות Sibs, וכן מדדי ממוצע, חציון, רבעונים, אחוזון .80.

- (1) הממוצע : 3.93 (2) החציון : 3
(3) ערכו של רבעון ראשון : 2 ערכו של רבעון שלישי : 5 ערכו של האחוזון ה 80 : 6

5. הפק'י התפלגות שכיחויות למשתנה מספר ילדים Childs.

- ד. אחוז הנבדקים, מתוך כל המדגם, שיש להם 8 ילדים ויותר : 1.1%
ה. אחוז הנבדקים, מתוך המשיבים, שאין להם ילדים כלל : 27.8%
ו. אחוז הנבדקים שיש להם עד 3 ילדים כולל : 83.8%

6. הפק'י Frequencies למשתנה רמת השכלה של האמא maeduc, והשב'י :

- ד. האחוזון ה 90 : 15
ה. הממוצע : 10.79
ו. אחוז הנבדקים אשר לאמהותיהם עד 12 שנות השכלה (כולל) : 78.3%