

מדריך לעיבודים סטטיסטיים ב-SPSS בסמינר מחקר ותזה

נוסח 8 – מרץ 2015

ד"ר עדו גל ©

מבוא

מדריך זה מיועד לתלמידי סמינרים ותלמידי מ.א. בחוג לשירותי אנוש, ומתאר צעדים בסיסיים בעיבוד נתונים שנאספו במסגרת סמינר או תיזה. ההנחה היא שתלמידים יבחרו עיבודים המתוארים כאן או יוסיפו ניתוחים מתקדמים (שאינם כלולים במדריך) בהתאם לשאלות המחקר שלהם, תוך העזרות בחומר שלמדו בקורס לסטטיסטיקה, ולפי ייעוץ מהמרצה. המדריך משתמש בלשון זכר אך כל האמור בו מתייחס באותה מידה לגברים ונשים.

המדריך כולל עשרה פרקים המאורגנים לפי סדר העבודה הצפוי על קובץ נתונים טיפוס. תחילה, תבדקו את תקינות הנתונים (פרקים א-ג), בהמשך, תבצעו צעדים מקדימים כדי ללמוד את הקובץ ולהכיר את מאפייני הנבדקים (פרקים ד-ה), תבנו משתנים מסכמים (המבוססים על תשובות הנבדקים לפריטים השייכים לסולמות שונים) או תייצרו משתנים חדשים המבוססים על קיבוץ או שינוי ערכים במשתנים רציפים קיימים (פרקים ו-ז). אחרי הכנה זו, תוכלו לבדוק את שאלות והשערות המחקר בדרכים מגוונות (פרקים ח-י).

שלבי עבודה עיקריים

פרק	נושא	עמ'
	מבוא	1
א.	קריאת קובץ הנתונים	3
ב.	בדיקת הנתונים הגולמיים בקובץ	3
ג.	תיקון נתונים שגויים והגדרת ערכים חסרים	4
ד.	הכרת מאפייני הנבדקים וההתפלגויות במדגם	6
ה.	ניתוח פריטים	7
ו.	חישוב ציונים ממוצעים/מסכמים בסולם (אינדקסים)	9
ז.	שינוי ערכי משתנים: קיבוץ, היפוך כיוון, ועוד	9
ח.	בדיקת התפלגויות וממוצעים בתת-קבוצות	12
ט.	בדיקת השערות לגבי קשר בין משתנים:	13
	1. בדיקת קשר בטבלאות (צילוח) ומבחן חי-בריבוע	15
	2. בדיקת קשר באמצעות מתאמים (פירסון, ספירמן)	15
י.	בדיקת השערות לגבי הבדלים בין קבוצות:	16
	1. מבחן t (בדיקת הבדלים בין שתי קבוצות)	17
	2. ניתוחי שונות (בדיקת הבדלים בין כמה קבוצות)	19
	נספח: טיפול בקבצים והעתקת נתונים ל-Word	26

לכל פרק מבנה דומה: תחילה מוצג רקע לגבי הנושא/שאלה/דילמה איתם תצטרכו להתמודד בשלב זה. אחרי כן מוסבר מה רצוי לעשות ואיזה פעולות לבצע בפועל ב SPSS. בסיום, ניתנות הצעות לפירוש הממצאים, ותזכורות לגבי דברים שכדאי לשים אליהם לב בתכנון או פירוש.

המדריך מותאם למהדורה 17 ומעלה של SPSS. יתכן כי פרטים מסוימים במהדורה חדשה יותר של SPSS יהיו שונים קצת מאלו הרשומים כאן. הרגישו נוח לבצע ניסוי וטעייה – תוכנת SPSS היא די ידידותית וניתן לחפש בתפריטים אפשרויות לפי צרכי המשתמש.

וכעת, כמה הצעות חשובות...לגבי ניתוח נתונים ודיווח מחקרי:

1. אמצו "ראש פתוח" בגישה לעיבוד נתונים: כנקודת מוצא, חשוב לציין כי מה שלמדתם בעבר בקורסי סטטיסטיקה נתן לכם רק חלק מהרקע והכלים לנתח קובץ נתונים אמיתי. אמנם למדתם על פרוצדורות סטטיסטיות שונות, אך על כל אחת בנפרד. כעת תעמדו בפני אתגרים חדשים: היות ועיבוד נתונים אינו תהליך המתרחש לפי תפריט מוכתב מראש, אלא "נתפר" לפי צרכי החוקרים, תצטרכו לבדוק שאלות מחקר והשערות ייחודיות לכם, לקבל החלטות "תוך כדי תנועה", ולהפעיל צורת חשיבה שונה חלקית מזו המוכרת לכם. לכן, המדריך מוליך אתכם דרך 10 פרקים המאורגנים לפי שלבי העבודה הצפויים בעיבוד נתונים, וכתוב כדי לסייע לכם בהחלטות ובטיפול במגוון של סוגיות טכניות העולות בכל שלב, כולל כאלו שלא למדתן בעבר.

2. נסו לבצע עיבודים מעניינים, תוך שימוש בהליכים סטטיסטיים פשוטים: מדריך זה מבוסס על ההנחה שעדיף כי תלמיד בסמינר יבצע בדיקות סטטיסטיות שהוא מבין ויכול לפרש היטב את תוצאותיהן, גם אם הן לא מתוחכמות מבחינה סטטיסטית. למרבה הצער, לעתים קרובות תלמידים אשר למדו בקורס מתקדם על עיבודים רב-משתניים כגון רגרסיה או ניתוח שונות (מבחן F) מנסים ליישם מבלי להבין אותם לעומק או לדעת איך לפרש את תוצאותיהם, ואז נמצאים במבוכה. לכן, עדיף להתמקד בהליכים סטטיסטיים המוכרים לכם. אך עם זאת, המדריך מתוכנן לתמוך בכם ברגעי החלטה שונים, כך שתוכלו "לפרוש כנפיים", לבצע עיבודים מעניינים ולבדוק הן השערות פשוטות והן שאלות רב-משתניות מורכבות יותר, על ידי שילוב מושכל ויצירתי של עיבודים שונים ושל הליכי עזר שהמוסברים בתוך הפרקים השונים.

3. השתמשו ב"תהליכי גישוש": המדריך מבוסס בחלקו על הגישה הטוענת שחלק מקדים חשוב בניתוח נתונים הוא ביצוע תהליכי גישוש (exploratory data analysis) שמטרתם לפתח תחושה ראשונית לגבי "מה קורה בנתונים שלי" ולהכיר את התפלגויות המשתנים, כשלב מקדים לביצוע בדיקות פורמליות. ראו במיוחד את ההסברים הקצרים בפרקים ד', ו', ח'.

4. השתמשו בגרפים: לעתים רצוי להשתמש בניתוחים גרפיים להכרת ההתפלגויות במשתנים שונים (כחלק מתהליכי גישוש שמוזכרים בסעיף קודם), או לבחינת אופי הקשרים בין משתנים שונים (האם הקשר הוא לינארי? או קשר U?). בנוסף, גרפים מסייעים להציג דפוסי ממצאים מורכבים שקשה להביןם כשהם בטבלאות דחוסות (למשל גרף המשווה ממוצעים במספר משתנים בכמה קבוצות נבדקים). התייעצו עם המרצה אם רצוי, ואיך, להציג גרפים בדו"ח.

5. נתחו מה קורה בתת-קבוצות: חשוב לזכור כי את העיבודים המתוארים בהמשך רצוי לבצע לא רק במדגם המחקר כולו אלא גם בתת קבוצות. למשל, אם אחת ההשערות היא כי יהיו הבדלים בין גברים ונשים במשתנה מסוים, ברור כי הגיוני לבדוק את השפעת המיגדר גם ביחס למשתנים הנבדקים בהשערות אחרות (גם אם אין לכם השערות ספציפיות בנושא). לכן, לאחר שחישבתם מדדי מרכז ופיזור או מתאמים, או הרצתם מבחנים סטטיסטיים (כגון מבחן t) במדגם כולו, יתכן ויש לחזור על חלק מהליכים אלו ולבדוק אם ממצאים שנמצאו במדגם כולו אכן יציבים ומופיעים גם בתוך תת-קבוצות. לעתים מגלים דפוסים מעניינים ובלתי צפויים בדרך זו, והדבר מאפשר להעמיק את העיבודים ולשפר את תרומת המחקר להבנה ולידע.

6. חישבו באופן ביקורתי על משמעות הממצאים, ושימו לב לדיווח/כתיבה אפקטיבית: סמינר מחקר הוא תהליך רב-שלבי שמטרתו לפתח ולחזק הבנה ומיומנויות בתחומים רבים. חשוב לא רק לבצע עיבודים סטטיסטיים מתאימים, אלא גם לפרש נכון את התוצאות, ולכתוב בצורה תמציתית וברורה את הממצאים ואת המסקנות הנובעות מהם. (שימו לב לחלק האחרון של כל פרק במדריך זה, אשר מרחיב לגבי פירוש פלטים או לגבי הזהירות הנדרשת בפירושם). שאלו עצמכם כל הזמן האם הכתיבה שלכם היא מספיק מאורגנת וברורה לקורא העבודה. היעזרו בדוגמאות שמספק המרצה שלכם לגבי תכנון, ניסוח וארגון פרקי עבודת המחקר/הסמינריון.

לסיום, אני מקווה כי מדריך זה יסייע לכם בהכנת עבודת המחקר. אנא שילחו תגובות והצעות לשינויים אלי: <iddo@research.haifa.ac.il>, כדי שאוכל להמשיך ולשפר את המדריך.

פרק א': קריאת קובץ הנתונים

1. רקע: ברוב הסמינרים תלמידים מקבלים קובץ גדול שנוצר מאיחוד קבצים צוותיים או ממסד נתונים חיצוני. לעתים, צוותי מחקר יוצרים קובץ עצמאי משלהם. אם קיבלתם קובץ ממורה הסמינר, אתם אמורים לקבל מידע מובנה ("ספר נתונים") לגבי מבנה הקובץ – כלומר מסמך המציין לאיזה פריט בשאלון מתייחס המשתנה בכל טור בקובץ, ומתאר את הערכים המוגדרים כחריגים או כ"ערך חסר" (missing values).

2. פתיחת קובץ הנתונים: דרך תפריט: Data ← open ← file

יופיע חלון Open file. נוודא כי אנו נמצאים על הכונן (Drive) הנכון, לדוגמא: C או G, בהתאם לפרטי הקובץ שניתנו. נחפש את הספרייה המתאימה, ונלחץ על שם הקובץ הרצוי. על המסך יופיע גליון הנתונים (Data editor). בשורה העליונה רשומים שמות המשתנים. (הטור הראשון מסמן את המספר הסידורי של כל נבדק/רשומה (case/record) בתוך הקובץ).

פרק ב': בדיקת הנתונים הגולמיים בקובץ

1. רקע: לפני שנוכל להתחיל לבצע עיבודים סטטיסטיים, עלינו לוודא שהקובץ שלנו "נקיים" והנתונים תקינים - רק אח"כ ניתן לבצע עיבודים נדרשים. חיוני לבדוק את תקינות הנתונים, כי נתונים איכותיים הם תנאי הכרחי למחקר טוב! לפיכך, בפרק זה מוסבר איך לבדוק אם קיימים ערכים לא-תקינים או לא-הגייוניים בקובץ. בפרק הבא (פרק ג') מוסבר איך לתקן ערכים שגויים. (אולם אם יסתבר בבדיקה כי בקובץ שלכם אין בעיות - ניתן לדלג על פרק ג'). נתונים בעייתיים נוצרים בגלל טעויות של הנבדקים (שגו במילוי השאלון) או טעויות של מי שהקליד את הנתונים. כתוצאה, יתכן שקיימים בקובץ ערכים "לא-תקינים" שברור ממבנה השאלון כי הם שגויים – למשל בפריטים סגורים ידוע מראש מה טווח התשובה האפשרי (למשל 1-5) ולכן כל מספר שחורג מטווח זה הוא שגוי. בנוסף, יתכן שמולאו או הוקלדו ערכים לא-הגייוניים בפריטים "פתוחים", למשל הוכנס גיל או מספר שנות לימוד שחורג מטווח סביר.

2. מה עושים? ניתן לאתר ערכים בעייתיים בשתי דרכים (לרוב משלבים את שתיהן...):

(א) הרצת DESCRIPTIVE לבדיקת מינימום/מקסימום בכל משתנה, ו/או (ב) הרצת FREQUENCIES לבדיקת התפלגות שכיחויות בכל משתנה. להלן הסבר מפורט לגבי כל דרך.

א. בדיקת מינימום/מקסימום: (להסתכל במסך - מיותר להדפיס!!)

1. ANALYZE ← DESCRIPTIVE STATISTICS ← DESCRIPTIVE.

2. סמנו את כל המשתנים הרצויים בתיבה השמאלית ולחצו על החץ להעביר לתיבה מימין. (כדי לסמן את כולם, לחצו/קליק על הראשון ואז מישכו את הסמן עד לסוף הרשימה).

3. לחצו OK. SPSS יציג בחלון הפלט מינימום/מקסימום וכן ממוצע וסטיות תקן.

4. בידקו (במסך) כי המינימום/מקסימום בכל משתנה לא חורגים מהטווח המותר.

תוצאות חישוב מינימום/מקסימום מופיעות בטבלה פשוטה, אך היא **לא מפרטת** את כל הערכים השגויים (רק את המקסימום/המינימום - יתכן שיש עוד ערכים שגויים בין המקסימום למינימום בגלל טעויות הקלדה אחרות). לכן בדיקה כזו משמשת רק כשלב ראשון לאיתור נתונים בעייתיים. אם נמצא משתנה עם נתון בעייתי, נריץ בהמשך התפלגות שכיחויות (frequency) כדי לגלות מהם כל הערכים הבעייתיים, ולכמה אנשים הם קיימים.

ב. בדיקת שכיחויות: (זיכרו כי מספיק להסתכל על הפלט במסך - מיותר להדפיס!!)

1. ANALYZE ← DESCRIPTIVE STATISTICS ← FREQUENCIES

2. סמנו את כל המשתנים הרצויים בתיבה השמאלית ולחצו על החץ. (לא את מספר הנבדק) (כדי לסמן את כולם, לחצו/קליק על הראשון ואז מישכו את הסמן עד לסוף הרשימה).

3. לחצו OK. SPSS יציג בחלון הפלט את כל הערכים הקיימים בכל משתנה ושכיחות כל אחד.

4. בדיקת התפלגות השכיחויות שקיבלתם: (ראו המשך ההסבר בסעיפים 4/א4/ב עמוד הבא...)

4א': בדיקת בעיות תקינות במשתנים "סגורים": האם מופיעים ערכים שאינם תקפים?
 נניח שהערכים הלגיטימיים במשתנה מסוים אמורים להיות בין 1-5. אם תראו ליד VALID ערך אחר (למשל 6 או 55) זה אומר שערך לא-תקין הוקלד לנבדק אחד או יותר, אך SPSS עדיין מתייחס אליו כערך לגיטימי ויש צורך לתקן מצב זה (בשלב הבא).

4ב': בדיקת בעיות לוגיות במשתנים "פתוחים" - האם מופיעים ערכים לא-הגיוניים?

אם במשתנה יכול להיות טווח גדול של ערכים (כגון גיל, מס' חודשי עבודה, וכו'), אתם אמורים לקבוע מהו טווח הערכים הלגיטימיים/הגיוניים, ואז לאתר משתנים בהם יש חריגה מהערך הסביר. סורקים בעין את התפלגויות השכיחויות. אם הקובץ קטן, ניתן להסתכל בעין לאורך הטורים ב Data Editor. במקרים מסוימים ניתן גם להצליב משתנים (ע"י CROSSTABS) שאמור להיות ביניהם קשר לוגי (ראו דוגמה 2)

היו זהירים - ייתכן כי נבדקים רשמו ערכים לא-צפויים שהם בכל זאת הגיוניים. בנוסף, זיכרו כי חריגות במשתנים פתוחים קורות כאשר מקלידים ערכים המתייחסים ל"מספר שנים", כגון גיל, מספר שנות לימוד, או שנות ותק בארגון או בתפקיד. במחקרים שונים נקבעים כללי הקלדה שונים לערכים כאלו, למשל כשנים בלבד, או גם עם חלקי שנים, או עם הסבה למספר חודשים. בהתאם לכך משתנה הטווח הלוגי של המשתנה. אם מסבים למס' חודשים (לא שנים), הטווח האפשרי גדול מאוד.

דוגמה 1: הקלדתם את ה"ותק בארגון" של נבדק כשבר עשרוני (למשל "6.5" לסימון שש וחצי שנות עבודה, אם כך הוחלט) במקום להסב זאת לחודשים. תוכלו לגלות זאת על ידי איתור ערכים לא סבירים בהתפלגות השכיחויות.

דוגמה 2: יש סתירה בין משתנים עם קשר לוגי. למשל, אם מופיע במשתנה "וותק בתפקיד" מספר גדול יותר מהמשתנה "וותק בארגון". (מבחינתו של הנבדק, יתכן והוא חשב על כך שעבד באותו התפקיד בארגון קודם. אך אנו רצינו לדעת וותק בתפקיד בארגון הנוכחי, כלומר מבחינתנו הנבדק פירש לא נכון את השאלה).

פרק ג': תיקון נתונים שגויים והגדרת ערכים חסרים

1. רקע: יש שני הליכים מקובלים לטיפול בערכים שגויים בקובץ, המוסברים למטה, אחד בו מקלידים במקום הערך השגוי את הנתונים האמיתיים והנכונים שצריכים להופיע, ושני בו אתם מודיעים ל-SPSS להתעלם מערכים מסוימים שהם שגויים.

2. מה עושים? להלן תאור של שני ההליכים המקובלים לטיפול בערכים שגויים.

הליך 1 - הקלדת הערך הנכון: בחרו בהליך זה אם ביכלתכם לאתר את השאלון המקורי של הנבדק ולגלות מהו הערך הנכון שהיה צריך להיות מוקלד, ואם הקובץ קטן וקל לניווט. בפועל, אחרי שבדקתם לפי הפרק הקודם מינימום-מקסימום ו/או שכיחויות, אתם יודעים כעת איזה מספרים (ערכים) שגויים מופיעים בקובץ לגבי משתנה מסוים.

1. הכנסו ל Data editor והסתכלו בטור של המשתנה הבעייתי - מיצאו באיזו שורה יש ערך שגוי (לעתים יהיו כמה נבדקים שלהם ערכים שגויים - אל תעצרו בראשון...).

2. בידקו מה "מספר הנבדק" שלגבי אותה שורה (מספר הנבדק לרוב יופיע בטור הראשון, תחת משתנה המזהה את המספר הרשום על השאלון המקורי).

3. חפשו את השאלון המקורי, בידקו מה הערך האמיתי המופיע בו והקלידו את הערך הנכון הזה בתא המתאים ב- Data editor, במקום הערך השגוי.

4. בצעו Save.

הליך 2 - שינוי ערכים שגויים בקובץ המקורי והחלפתם בערך שיוגדר "חסר": גם אם איתרתם בבדיקות שביצעתם את קיומם של נתונים שגויים, לעתים קרובות לא ניתן או לא רצוי לתקן נתונים שגויים כאלו. לכאורה, ניתן למחוק כל ערך שגוי. אולם, פעולה זו גורמת לאיבוד מידע ובלבול - אם מוחקים נתונים (יוצרים תאים ריקים), החוקר לא יכול יותר להבחין בין מקרים בהם הנבדק דילג או לא ענה כלל על שאלות (ולכן התא ריק), לבין

מקרים בהם נבדקים ענו על השאלון אך רשמו ערכים חריגים (שנמחקו), לבין טעויות הקלדה של החוקרים (שנמחקו). בנוסף, לעתים קרובות תיקונים בקובץ אינם אפשריים או מסובכים מדי למשל אם הקובץ גדול מאוד או אין לכם גישה לשאלונים המקוריים. **במצבים כאלו, עליכם להודיע לSPSS להתעלם מערכים מסוימים שהם שגויים. בהתאם למצב ולהנחיות שתקבלו מהמורה, תוכלו לבחור באחד או שני הצעדים הבאים:**

א. **הקלדת ערך (מספר) מיוחד שיוגדר כ"ערך חסר" (missing value):** כבר בתהליך הכנת השאלונים להקלדה ניתן לקבוע ערך מיוחד שיוקלד במקום כל ערך שגוי/חריג (למשל אם נבדק דילג על פריט): בד"כ מחליטים שרירותית ש"ערך חסר" יהיה מספר כגון 9 או 999 או 9999 שהוא גבוה מכל ערך לגיטימי במשתנה כלשהו. (באופן דומה ניתן להקליד ערכים מיוחדים אחרים, למשל "888", לסימול מקרים מיוחדים (אחרים).

ב. **מגדירים לSPSS מהם הערכים החסרים (missing values) בכל משתנה בו יש בעייה:**

- לחצו בתחתית כדי לעבור לחלון "variable view" בו מגדירים מאפייני משתנים.
- מצאו את השורה של המשתנה הראשון בו גיליתם קודם ערך שגוי.
- לחצו בתיבת "missing" בשורה של המשתנה השגוי (תחת הטור בשם זה).
- לחצו על הכפתור שהופיע בתיבה.
- **בתיבת הדיאלוג שתופיע, עדכנו את הערכים שהוגדרו כשגויים במשתנה.** למשל אם מדובר במשתנה מסוג ליקרט בטווח 1-5 והגדרתם "9" כערך חסר, הקלידו ערך זה כאחד הערכים הבדידים (discrete values). אפשר להכניס כמה ערכים בדידים, אם ידוע מבדיקת השכיחויות מהם כל הערכים השגויים, או להכניס טווח (range): למשל, אם מדובר במשתנה מסוג ליקרט בטווח 1-5 עם הרבה ערכים שגויים (כגון 6, 8, 44, 55) ניתן לרשום את הטווח 55-6 שיכלול את כל הערכים השגויים. שימו לב כי אפשר גם להכניס שילוב של טווח וערך בדיד.
- חזרו על התהליך לגבי כל משתנה בו זיהיתם שגיאה במשתנה כלשהו.

3. פירוש - מה SPSS עושה עבורכם אחרי שהגדרתם ערכים חסרים? כאשר ערכים מסוימים הוגדרו כ"missing" התוכנה יודעת כי הם שגויים ויש להתעלם מהם בעיבודים הסטטיסטיים, אך עדיין מסוגלת לספור אותם (ולחשב שכיחויות) כאשר בודקים מה מספר הנבדקים שענו (N) על פריט כלשהו. כלומר, SPSS מסוגל להבחין בין "valid cases" (רשומות עם ערכים תקינים והגיוניים) לבין סוגים שונים של "missing cases" (רשומות עם ערכים לא-תקינים/שגויים). זו גם דרך שמאפשרת לחוקר להשאיר בקובץ קודים מיוחדים לסימון מקרים שהחוקר רוצה להבחין ביניהם, כגון "לא יודע", "לא מבין", "לא רלוונטי", ולחשב שכיחויות.

4. בקרת טיב התיקונים:

- א. לאחר שהכרזתם על ערכים חסרים בחלופה 1 או 2, יש לוודא שלא "פספסתם" אף משתנה (תמיד קורות טעויות!). יש להריץ DESCRIPTIVE פעם נוספת (ראו צעד 1), הן לבדיקת מינימום/מקסימום והן לקבלת ממוצע וסטיית תקן בכל משתנה בקובץ. תצטרכו מדדים סטטיסטיים אלו להכרת הנתונים שלכם, לכן זה כדאי בכל מקרה.
- ב. בידקו בפלט (על המסך, אין צורך להדפיס) כי לאחר התיקונים המינימום והמקסימום בכל משתנה לא חורגים מהטווח התקיין/ההגיוני.
- ג. חזרו על צעדים קודמים אם גיליתם בעיה נוספת כלשהי.

תזכורת – לבצע גיבוי!! האם שיניתם משהו בקובץ? בצעו SAVE AS ותנו שם חדש לקובץ. הגנו על הנתונים שלכם – הקפידו תמיד לשמור גיבוי על אמצעי איחסון נוסף - למשל אפשר לשמור על זיכרון נייד וכן לשלוח לעצמכם במייל. תמיד צריך שיהיו שני עותקים נפרדים של הקובץ שלכם בשני מקומות שונים.

פרק ד': הכרת מאפייני הנבדקים במדגם והתפלגויות משתנים

1. רקע: לאחר תיקון הקובץ, כשלב ראשון בניתוח הנתונים, יש להכיר יותר מקרוב את מאפייני הנבדקים והמדגם ואת המשתנים המקוריים בקובץ. שאלות בשלב זה הן, למשל:

- כמה אנשים יש בתת-קבוצות במשתנים חשובים (כגון מיגדר, מצב משפחתי)?
 - איך נראית ההתפלגות של משתנים רציפים חשובים (כגון גיל, שנות השכלה)?
- למידע כזה שימושים חשובים. נוכל לראות אם ההתפלגויות במדגם תואמות את ציפיותינו, והאם יש חריגות משמעותיות (למשל מספר גדול בהרבה של נשים לעומת גברים). מידע על התפלגויות גם ישרת אותנו כשנתאר את מאפייני המדגם (בפרק "שיטה" או "ממצאים"). בנוסף, הכרת התפלגויות חשובה כי לפיה נחליט (בהמשך) איך לקבץ משתנים רציפים.

להלן דוגמאות לצעדים שיש לבצע בשלב זה כדי לענות על שאלות לגבי מאפייני נבדקים והתפלגויות. (כמובן, עליכם להחליט על השאלות שמעניינות אתכם לגבי הקובץ שלכם).

2. מה עושים? דוגמאות להרצת עיבודים פשוטים לשם מתן מענה לשאלות בסיסיות

א. כמה גברים ונשים יש במדגם? כמה בעלי השכלה אקדמית?
ניתן לענות על שאלות מסוג זה ע"י בדיקת ההתפלגות הפשוטה של המשתנים. נשתמש בפרוצדורת Frequencies (תחת Frequencies Analyze **B** Descriptive Statistics **B**). נבחר משמאל את המשתנה המתייחס למיגדר או להשכלה, נלחץ על החץ להכלילו בעיבוד (המשתנה יעבור לתיבה מימין). נלחץ OK להפעלת הפרוצדורה. בפלט שיופיע נבחן מספרים מוחלטים ואחוזים. (אם הכנסנו לתיבה הימנית משתנה שלא רצינו, נסמן אותו בתיבה הימנית ונלחץ על החץ ההפוך כדי להחזירו לרשימה הכללית בתיבה בצד שמאל).

ב. מה ממוצע שנות הלימוד? ניתן לחשב מדדי מרכז ופיזור במספר דרכים:

1. באמצעות Frequencies: ניתן לקבל מדדים מסוימים כ"עיבוד משנה" של Frequencies: נכנס לפרוצדורה כמוסבר למעלה. לאחר סימון המשתנה הרלוונטי והעברתו לתיבה מימין באמצעות החץ, באותו חלון נלחץ על Statistics ותופיע תיבת משנה. נבחר במדדים המבוקשים, כגון מדדי מרכז (ממוצע Mean, חציון Median, שכיח), מדדי פיזור (שונות, סטיית תקן, טווח) ואחוזונים (תחום בין-רבעוני). נלחץ Continue. נלחץ OK להרצת הפרוצדורה. בחלון הפלט יתקבלו המדדים המבוקשים בנוסף להתפלגות השכיחות.

2. באמצעות Descriptive: אם מעוניינים רק בממוצע ומדדי פיזור, ולא בהתפלגויות (שכיחות) גולמיות, נריץ Descriptive (תחת Descriptive Statistics Analyze **B**). נסמן כרגיל את המשתנים, ואז יש נלחץ על Options ונבחר מדדי מרכז ופיזור רצויים (שימו לב לברירת המחדל). נלחץ Continue. לאחר מכן נלחץ OK להרצת הפרוצדורה.

ראו הנספח להסבר טכני לגבי שמירת קובץ OUTPUT של SPSS או העתקה לקובץ WORD.

3. בדיקות שכיחות וממוצעים בתת-קבוצות

לעיתים רוצים לדעת כמה אנשים (במספרים מוחלטים, אחוזים) קיימים בתת-קבוצות שונות במדגם, או מה הממוצעים של משתנים כלשהן בתת-קבוצות שונות. מידע זה חיוני כדי שנוכל לתאר את גודלן של קבוצות עיקריות במדגם ולהבין את ההבדל בין ממצאים המתייחסים לכלל המדגם לבין ממצאים בקבוצות מסוימות (לעיתים לכך מתייחסות חלק מההשערות).

ניתן להשתמש במספר תהליכים לקבל מידע כזה. הנפוצים ביותר הם "צילוח" (הצלבת נתונים בלוחות שכיחות) ע"י פרוצדורת CROSSTABS או הרצת עיבודים בנפרד בכמה תת-קבוצות ע"י SPLIT FILE. שני הליכים אלו מוסברים במפורט בפרק ח' בהמשך - הם שימושיים בשלב מתקדם יותר, לאחר שתבנו (תחשבו) ציונים מסכמים במשתנים שונים.

פרק ה': ניתוח פריטים

1. רקע: הרבה שאלוני מחקר כוללים קבוצות פריטים שנועדו לבדוק משתנה מסוים. אך לרוב, איננו מתעניינים בתשובת הנבדק לכל פריט לחוד, אלא רוצים ליצור ציון מסכם (לעתים יקרא "אינדקס" או "סולם") ע"י חישוב ממוצע או סכום של התשובות בכל הפריטים הנכללים בסולם/אינדקס, כדי לבטא את מצבו של הנבדק במשתנה. אך לפני שנוכל לחשב אינדקס/סולם, עלינו את ה"עקיבות הפנימית" (מדד אלפא של קרונברך), כדי לדעת אם כל הפריטים הנכללים בסולם אכן בודקים את אותו משתנה, או אולי יש פריטים שיש קשר נמוך בינם לשאר הפריטים (לא נכנס כאן לדיון מעמיק בנושא, אשר נלמד במפורט בקורס שיטות מחקר. ההסבר בהמשך הוא טכני, ומטרתו להוליך אתכם דרך שלבי הביצוע ולהדריך על מה להסתכל בפלט).

2. מה עושים? נבצע "ניתוח פריטים", הכולל ארבעה צעדים המוסברים להלן:

(שימו לב: לפני תחילת העבודה, צריך להפוך את הכיוון של פריטים שמנוסחים בצורה הפוכה לשאר הפריטים.... ראו הסבר פשוט בנושא, בפרק ז' סעיף 5, בתחתית עמ' 11).

צעד 1: כניסה לפרוצדורה "RELIABILITY" ובחירת אופציות פלט

- א. תחת תפריט Analyze נבחר Scale : Reliability Analysis.
- ב. כדי שיהיו ברשותנו כל הנתונים הבסיסיים הנדרשים להערכת טיב השאלון ומאפייני הפריטים, עלינו לבחור מספר אופציות ש SPSS לא מפעיל באופן אוטומטי. לכן, לאחר שנפתח חלון הפרוצדורה, נלחץ על Statistics ונבחר באופן הבא (להקליק בריבוע):
 1. בתיבת Descriptive for נסמן את כל שלושת הריבועים (אלפא, אם מושמט, וכו')
 2. בתיבת Inter-item נסמן: Correlations. לחץ Continue.

צעד 2: בחירת משתנים והפעלת הפרוצדורה

נבחר את הפריטים (משתנים) לניתוח מרשימת המשתנים משמאל. לרוב נסמן בשלב זה כל פריט השייך לסולם (נשתמש ב"מפתח" או ב"ספר קידוד" כדי לדעת את שמות המשתנים בסולם). נלחץ על החץ כדי להכניסם לתיבת Items. נלחץ OK להריץ את העיבוד.

בחלון התוצאות נבדוק:

- א. ממוצעים וסטיות תקן: האם הם סבירים בכל המשתנים (אין ערכים חריגים)
- ב. מתאמים בין הפריטים: האם הם חיוביים וסבירים בגודלם. פריט עם מתאמים קרובים לאפס או שליליים לשאר הפריטים כנראה לא קשור לשאר הפריטים בסולם ויתכן שצריך להוציאו. (אך יתכן ששכחנו להפוך כיוון של פריט מסוים. ראו פרק ז' סעיף 5).
- ג. Alpha: (נתייחס בד"כ ל Cronbach's Alpha ולא ל Standardized). כלל אצבע מקובל הוא שעדיף כי ערך האלפא יהיה מעל 0.70. בתנאים מסוימים חוקרים מוכנים לקבל גם ערכים נמוכים במקצת מ 0.70. אם הערך נמוך או קרוב ל 0.70, נבדוק האם יש פריט שפוגם בעקיבות הפנימית, בעזרת הטבלה "Item-total statistics". נבדוק כמה טורים:
 - 1.1: Corrected item-total correlation: טור זה מציג את המתאם בין הפריט לציון הממוצע בכל שאר הפריטים, כשהממוצע מחושב בלי פריט זה. נחפש פריטים שלגביהם יש בטור זה מתאם שלילי או נמוך (כלל אצבע - מתחת 0.4). מתאם נמוך או שלילי מראה כי הקשר בין הפריט לשאר הפריטים בסולם הוא רופף או חלש--כלומר הפריט לכאורה לא בודק את המושג/נושא שבודקים שאר הפריטים בסולם.
 - 2.2: Alpha if item deleted: טור זה מציג מה יהיה ערך האלפא בסולם, אם הפריט היה מושמט/נזרק (אם לא היה נכלל בסולם). יש לפרש את ערך האלפא המופיע לכל פריט ע"י השוואתו לערך האלפא הכללי (שמוצג בתחילת או סוף הפלט). אם האלפא הרשום לפריט מסוים בטור זה הוא גבוה יותר מה Alpha שרשום בסוף הפלט, הרי שהאלפא היה עולה (משתפר...) אם הפריט היה יוצא מהסולם. כלומר, בלי פריט זה שאר הפריטים היו, בממוצע, במתאם גבוה יותר זה עם זה, והעקיבות הפנימית של הסולם הייתה משתפרת (האלפא היה נהיה גבוה יותר).

צעד 3: קבלת החלטות

אם מלכתחילה האלפא גבוה מ-0.70, יתכן והכל בסדר בסולם, ואפשר להשתמש בכל הפריטים לחישוב הציון המסכם בסולם. אולם כאשר פריט נראה בעייתי לאור המדדים הנ"ל (מתאם נמוך עם שאר הפריטים, או אלפא משתפר אם הפריט היה מושמט), יש מקום לשקול ניפוי הפריט הבעייתי כדי לשפר את העקיבות הפנימית. עם זאת, אסור לפעול כאן באופן מכני וסתם לזרוק פריט – יש להפעיל שיקול דעת ולקבל החלטות בהתבסס על כמה שיקולים, הן פרגמטיים והן תיאורטיים, המוסברים בקצרה להלן:

- א. נבדוק (1) האם יש ערכים שגויים בקובץ, או טעויות בהגדרת missing values? (2) האם שכחנו להפוך כיוון של פריט "שלילי"? (זה נושא פשוט - ראו סעיף 5 בסוף פרק ז').
- לשם כך נבדוק שוב את הערכים בהתפלגות הגולמית (frequency), ואת מפתח הסולמות. במיוחד, נעיין בתוכן/ניסוח הפריטים בשאלון ונחשוב אם מבחינה מושגית/תיאורטית כל פריט קשור למשתנה אותו הסולם אמור לבטא. לעיתים, עיון חוזר בשאלון מראה שניסוח הפריט היה מעורפל וזה יכול ליצור תגובות מוזרות של הנבדקים אליו, וזה יסביר את המתאם הנמוך בין הפריט ה"בעייתי" לכאורה לשאר הפריטים בסולם. נבדוק גם אם לפריט יש התפלגות מוטה או מוזרה (נבחן את ההתפלגות הגולמית ואת הממוצע וסטית התקן). גם זה עשוי להשפיע על המתאם עם שאר הפריטים.
- ב. רק אחרי בדיקה כזו נוכל להחליט מה לעשות: האם ל"זרוק" או להשאיר פריט בעייתי? בד"כ, לא נשאיר פריט שהמתאמים שלו עם שאר הפריטים נמוכים מאוד או שליליים. אולם, אין להוציא פריט באופן שרירותי רק כדי להכריח את הסולם להגיע לערך אלפא מעל 0.70. תמיד נפעיל שיקול דעת ונתייחס למספר סוגיות:

- **תוכן הפריט** (מה הוא מודד): נבחן את ניסוח הפריט ומה הוא מודד. לעתים נחליט להשאיר פריט שחשוב לנו לשמור בסולם למרות שהוא יוצר אלפא בעייתי במקצת.
- **הטרוגניות הסולם**: בסולמות המודדים משתנה בעל הגדרה מורכבת (למשל עם מספר תת-מימדים), ניתן לצפות למתאמים נמוכים בין חלק מהפריטים בסולם ולכן לערך אלפא נמוך מהמקובל, כי מדובר בסולם שאמור להיות יותר הטרוגני מעצם הגדרתו.
- **אורך הסולם**: בד"כ עדיף שבסולם יהיו 3-4 פריטים ומעלה כדי לאפשר "כיסוי" סביר של המשתנה הנמדד לעתים לא נזרוק פריטים אם הסולם שנשאר הוא קצר מדי.

לאור שיקולים כאלו, לעתים יתכן ונשאיר פריט שנתונו בעייתיים במקצת, ונקבל את העובדה כי במדגם המחקר הנוכחי לסולם האמור יש עקיבות פנימית נמוכה מהמצופה. להבהרה, ניתן להשתמש בסולם אפילו עם ערך האלפא שלו נמוך מ-0.70 (בדרך כלל בין 0.50 – 0.70), אולם במצב כזה נדרשת זהירות בפירוש משמעות הממצאים. אם האלפא נמוך, המשמעות המושגית של המשתנה הנמדד ע"י הפריטים בסולם אינה מגובשת מספיק, ולכך יהיה צורך להתייחס ב"דיון" במגבלות על פירוש והכללה.

- ג. אם לאחר צעד 3 לא מצאנו כל פריט בעייתי וערך האלפא הכללי סביר, ניתן להסיק כי לפחות על בסיס תהליך ניתוח הפריטים הראשוני שתואר כאן, כל הפריטים אכן "שייכים" לאותו סולם, וניתן לחשב מהם אינדקס (ראו הפרק הבא). **אולם, אם הנתונים מראים על בעיה, ולא נמצא הסבר סביר ל"התנהגות" הפריט שיצדיק השארתו בסולם, לרוב נחליט לנפות אותו – ואז נבדוק אם זה משפר את האלפא.**

צעד 4: חזרה על ניתוח הפריטים, לאחר ניפוי פריט, לשם בדיקת השיפור באלפא

אם החלטנו לנפות פריט, נריץ שוב RELIABILITY על כל הפריטים בסולם, פרט לפריט שהחלטנו להוציא – נסמן אותו בתיבה הימנית ונלחץ על החץ להחזירו לרשימה הכללית משמאל. אז נלחץ OK להריץ שוב. בפלט החדש, נבדוק את ה Alpha. אם הוא עדיין נמוך (בד"כ מתחת 0.70), נחזור על צעד 3 שתואר לעיל. אם נאתר פריט נוסף שראוי לנפותו, נחזור פעם נוספת על הפרוצדורה, בתקווה להגיע לסולם בעל עקיבות פנימית סבירה.

תזכורת: קיימות פרוצדורות נוספות שבאמצעותן ניתן לבדוק יותר לעומק את הקשרים בין פריטים בסולם, במיוחד "ניתוח גורמים" (Factor Analysis). התייעצו עם המורה.

פרק ו': חישוב ציון מסכם בכל סולם ("אינדקס")

1. רקע: לאחר ניתוח הפריטים, ניתן לחשב בכל סולם ציון מסכם ("אינדקס") כדי לבטא את מצבו הכולל של הנבדק במשתנה הנמדד ע"י הסולם. לרוב ציון זה יהיה ממוצע או סכום התשובות בפריטים הנכללים בסולם. (עדיף להשתמש בממוצע כי הוא באותו טווח ערכים כמו הפריטים המרכיבים אותו, לכן קל יותר לפרש אותו ולהבין את משמעות הציונים בו.)

2. מה עושים: נבחר בהליך COMPUTE תחת Transform. נניח לצורך הדוגמה שאנו רוצים לחשב אינדקס של "שביעות רצון" כממוצע של שלושה פריטים בשאלון: s1, s3, s6

Compute Variable	
Target Variable:	Numeric Expression:
satis	=(s1+s3+s6)/3

א. בתיבה משמאל (target variable) נכתוב שם קצר למשתנה החדש שניצור (למשל "satis"). בתיבה מימין לסימן השוויון (=) נכתוב נוסחה אלגברית פשוטה לחישוב המשתנה, למשל: $(s1+s3+s6)/3$

ב. נלחץ OK. כעת SPSS יחשב לכל נבדק את ערך המשתנה החדש - לא נראה כלום על המסך - אך המשתנה יצטרף ב Data editor (בסוף המשתנים הקיימים) ונוכל כעת להשתמש בו.

3. פירוש - הכרת המשתנה החדש שנוצר: יש לבדוק שהחישוב נעשה נכון, ע"י הרצת DESCRIPTIVE או התפלגות שכיחויות (FREQUENCIES). נבדוק מינימום/מקסימום, ממוצע, סטיית תקן, וכו' כדי להיות בטוחים כי חישוב האינדקס נעשה נכון. אם הטווח אינו סביר, או הממוצע נמוך מהמצופה, או המשתנה החדש קיים רק לחלק מהנבדקים, החישוב כנראה שגוי! (שימו לב – יתכן גם ששכחתם להפוך כיוון של פריט "שלילי". ראו טיפול בכך בפרק ז' סעיף 5, עמ' 11.)

4. בדיקת גרפים: היות והאינדקס שחישבתם יהיה לרוב משתנה רציף עם מספר ערכים רב, מומלץ להסתכל על גרפים (או התפלגות שכיחויות) כדי להכיר את התפלגותו ואופיו של המשתנה החדש. תחת תפריט Graphs ניתן להפיק גרף Histogram או Area.

קריאה נכונה וחשיבה ביקורתית לגבי גרפים: יש לבחון את צורת ההתפלגות: האם היא נורמלית, או מוטה (רוב האנשים מקבלים ציון גבוה או נמוך, מרוכזים בטווח צר מאוד? יתכן וכך תזוהו נושאים חדשים שרצוי לבדוק בהמשך (למשל אילו משתנים משפיעים על צורת ההתפלגות? האם יש הבדל בין תת-קבוצות כגון בין גברים ונשים?). בנוסף, יתכן כי התפלגות מוטה מאוד תגביל אתכם בבדיקת שאלה מסוימת – התייעצו עם המרצה.

פרק ז': שינוי ערכי משתנים: הקבצה, היפוך כיוון, ועוד

1. רקע: לפני שתבדקו את השערות המחקר, לעתים יש צורך לקבץ משתנים רציפים בקובץ שלכם לתוך מספר קטן של "קבוצות" או ערכים חדשים. למשל, המשתנה "גיל" לרוב מקבל מספר רב של ערכים (במדגם טיפוסי גיל יכול לנוע בין 16-80), אולם יתכן ששאלות המחקר דורשות רק השוואת שלוש קבוצות גיל (למשל 16-24, 25-64, 65 ומעלה). מצב דומה עשוי להיות לגבי משתנים כגון ותק או שנות השכלה. בנוסף, יתכן שבניתם אינדקס רציף (חישבתם ממוצע הציונים בכמה פריטים) אך בפועל אתם רוצים רק לבדוק את ההבדל בין שתי קבוצות, כגון אנשים שהם "גבוהים" (למשל מעל החציון) או "נמוכים" (מתחת לחציון) במשתנה זה. בכל המצבים המתוארים למעלה, עליכם לבצע הקבצה – בעזרת פקודת Recode. שימוש במשתנים מקובצים מאפשר לכם לעשות דברים מעניינים, למשל לבדוק השערות לגבי קשר בין משתנים על ידי קרוס-טבולציות (ראו פרק ח'), או לבדוק הבדלים בין קבוצות (למשל בין אנשים שהם "נמוכים" או "גבוהים" בשביעות-רצון) בעזרת מבחני t או F וכו' (ראו פרק י'). בנוסף, שינוי מקובל אחר בערך של משתנה הוא "היפוך" הכיוון שלו – ראו סעיף 5 בהמשך.

2. איך עושים הקבצה: פעולות מקדימות לפני שימוש בפקודת RECODE

לפקודת RECODE מספר שימושים: לרוב נשתמש בה כדי לקבץ משתנה רציף, ושלבי העבודה לשם כך מוסברים בהמשך סעיף זה (יש מספר שלבי-משנה – עיינו תחילה בכותרות, לפני קריאת הפרטים!). בנוסף, בעזרת פקודה זו ניתן לשנות ערכים של משתנה קיים לערכים אחרים, או להפוך כיוון של משתנה המנוסח באופן הפוך או "שלילי" (ראו סעיף 5 בהמשך).

א. פעולות מקדימות - קביעת גבולות הקטגוריות: לפני השימוש ב RECODE לקיבוץ משתנה רציף, יש להחליט מה יהיו גבולות הקטגוריות. כדי להבין את הנושא, להלן דוגמה: נניח שחישבנו משתנה חדש שהוא ממוצע התשובות בסולם עם כמה פריטים מסוג ליקרט בטווח 1-5 ("1" "לא מסכים כלל" עד "5" "מסכים מאוד"). ממוצע זה גם הוא בטווח 1-5. נניח שכעת אנו רוצים לקבץ/לחלק את האינדקס (הממוצע) לשלוש רמות: 1 (נמוך), 2 (בינוני), 3 (גבוה). **תחילה אנו חייבים להחליט מה גבולות הקטגוריות** (כלומר בין איזה ערך לאיזה ערך נקרא לציון כלשהו "נמוך", "בינוני", או "גבוה").

יש שתי גישות עיקריות לקביעת גבולות הקטגוריות, לכל אחת יתרונות וחסרונות משלה:

1. גישה "לוגית" לגבי גבולות הקטגוריות: נקבע גבולות לפי משמעות הסולם המקורי.

למשל – נוכל לקבוע טווח שווה לכל מחלקה לפעי ערכי הסולם המקורי. למשל, אם הסולם המקורי נע בין 1-5, נוכל לקבוע שכל קבוצה תהיה ברוחב של 1/3 מהציון: מי שהממוצע שלו בטווח 1.00 - 2.30 יקבל ערך מקובץ חדש "1" (נתייחס לכך כ"נמוך"), בין 2.31 - 3.70 יקבל ערך "2" ("בינוני"), ובין 3.71 - 5.00 יקבץ "3" ("גבוה"). לכאורה גישה כזו הגיונית, אולם בפועל היא תגרום לכם לבעיה אם ההתפלגות באינדקס מוטה או לא נורמלית (במובן הסטטיסטי...). נזכור כי לעתים קרובות ציונים באינדקסים נוטים כלפי מעלה (כלומר מתקבצים קרוב לערכים 4 או 5 בסולם 1-5). במצב כזה, אם נקבע גבולות של קטגוריות באופן לוגי, יתכן שיהיו מעט מדי נבדקים בקטגוריה הנמוכה, ויותר מדי נבדקים בקטגוריה הגבוהה. הדבר יוצר חוסר פיזור במשתנה המקובץ ואז הוא אינו שימושי. (אם באחת המחלקות יהיה אחוז קטן מדי של הנבדקים, זה עשוי להטות תוצאות של מבחנים מסוימים!). לכן, לעתים קרובות קובעים את גבולות הקטגוריות על בסיס אמפירי, המוסבר בסעיף הבא.

2. גישה "אמפירית" לגבי גבולות הקטגוריות: נקבע גבולות בין קטגוריות כך שבכל מחלקה יהיה אחוז די דומה של נבדקים. למשל, יתכן ונרצה לבנות שלוש קבוצות בעלות גודל יחסי דומה (בערך 35%-30% מהנבדקים). לשם כך, תחילה נריץ

FREQUENCIES, ונבדוק בהתפלגות מהם הערכים שמתחת להם נמצאים בערך 33% ו-66% מהנבדקים. כדי לקבץ משתנה לשתי קבוצות, אפשר להשתמש באותה שיטה כדי למצוא את החציון, או לחשב את הממוצע בעזרת descriptive. נוכל לקבוע כי מי שמעל לממוצע (או לחציון) יחשב כ"גבוה" ומי שמתחת יחשב "נמוך". **את ערכי הגבול שנקבע נכניס לפקודת recode ליצירת המשתנה המקובץ.**

לגישה זו יתרון וחסרון ביחס לגישה הקודמת. היתרון: בשיטה זו נוכל להשוות בין קבוצות בעלות גדלים דומים ונתגבר על הבעיה שנוצרת אם פועלים בגישה "לוגית" שתוארה למעלה. החסרון: גבולות אמפיריים בין מחלקות יכולים לעוות את התמונה האמיתית. הנבדקים שהוכנסו במשתנה המקובץ לקטגוריה "נמוכה" אינם בהכרח נמוכים במשתנה המקורי, הם רק נמוכים יחסית לאחרים. (בפועל, יתכן שהם בטווח בינוני או אפילו גבוה בהתייחס לטווח הערכים בפריטי המקור).

ב. בחירה בין שתי הגישות (או שילובן): לאור האמור, ברור שאין פתרון אלגנטי לקיבוץ משתנים - לכל שיטה יתרונות וחסרונות. תצטרכו להחליט בעצמכם, בהתאם להתפלגות המשתנים ואופי שאלות המחקר. תחילה הריצו FREQUENCIES כדי לראות איך מתפלג המשתנה (אל תדפיסו - רק הסתכלו במסך על ה Cumulative frequency). לפי השכיחות המצטברת מחליטים אם יתקבל מספר סביר של נבדקים בכל קטגוריה כאשר נחלק באופן "לוגי" למספר קטן של קבוצות. אם נראה שאחוז הנבדקים בקבוצות השונות יהיה שונה מאוד (למשל בקבוצה אחת יכללו 75% מהנבדקים, בשניה 24%, ובשלישית רק 1%), כדאי לבחור גבולות אחרים כך שיהיה אחוז סביר (מספיק גדול) של נבדקים בכל קבוצה.

תזכורת: בזמן קביעת גבולות הקטגוריות - יש לדייק בספרות עשרוניות! משתנים רציפים שחושבו כממוצעי מספר פריטים לרוב מכילים שבר עשרוני (למשל הממוצע של אדם באינדקס המבוסס על 7 פריטים עשוי להיות 4.153). רצוי לא להקליד ל RECODE ערכי גבול קטגוריות במספרים "עגולים" כגון: 1-2, 2.1 - 3, כי SPSS יאבד נבדקים בעלי ערכים שהם בין הגבולות שבחרתם (למשל מישו עם 2.05 נופל "בין הכסאות" של הגבולות הני"ל). הפתרון - לרשום יותר ספרות עשרוניות, כגון: 1.00 - 2.100, 2.101 - 2.999, וכך, כדי למזער את הרווח בין הקבוצות.

3. איך עושים הקבצה - הפעלת פקודת RECODE - הנמצאת תחת Transform

כשתלחצו על פקודת RECODE, SPSS יבקש אתכם לבחור האם לקדד ערכים למשתנה חדש "into different variable" (כלומר לא תשנו את המשתנה המקורי, אלא תיצרו משתנה חדש מקובץ) או לקדד לתוך המשתנה הקיים "into same variable" (ואז ימחקו הערכים המקוריים ובמקומם יכתבו החדשים). **עדיף לבחור ב "into different variable" כדי למנוע מחיקת המשתנה המקורי (הרציף),** אותו תצטרכו כדי לבצע עיבודים שונים כגון מתאמים.

א. "into different variable" - קידוד למשתנה חדש (האופציה המועדפת): בחלון שיפתח נבחר המשתנה/ים אותם רוצים לשנות ונעבירם לתיבה הימנית בלחיצה על החץ.

1. בתיבה Name נכתוב את שמו של המשתנה החדש (המקובץ) שיווצר.
2. נבחר Old and new values על מנת להגדיר ערכים חדשים בהתבסס על הערכים הישנים. בחלון שיפתח קיימות כמה אופציות. בד"כ כשרוצים לקבץ משתנה רציף למספר קטגוריות בדידות, נבחר ב Range. אך ניתן גם לבחור ב: שינוי ערך בודד, הפיכת ערך חסר לערך אחר, קיבוץ מהערך הנמוך ביותר של המשתנה עד למספר אותו נגדיר, או קיבוץ ממספר אותו נגדיר עד לערך הגבוה ביותר של המשתנה.
3. נבחר Range ונקליד בשתי התיבות את טווחי הערכים הישנים שיתקבצו בכל ערך חדש.
4. נקליד ל Value את הערך החדש (למשל ב Range נכניס 1.00 - 2.40 וב Value נכניס 1).
5. נלחץ Add בסיום קידוד כל טווח (ואז SPSS יכניס מידע זה לחלון באמצע) (נחזור על שלבים 4-5 מספר פעמים בהתאם למספר הקטגוריות שברצוננו ליצור)
6. נלחץ Continue בסיום קידוד כל הטווחים.
7. נלחץ Change. רק כעת SPSS יבצע את כל ההנחיות שפירטנו בחלונות השונים בפקודה זו ויצור את המשתנה החדש בקובץ.

ב. "into same variable" - קידוד לתוך משתנה קיים: הפעולה דומה למה שהוסבר למעלה אך הערכים החדשים ימחקו את הישנים ברגע שנלחץ על CHANGE, לכן, עדיף לא להשתמש בפקודה זו אלא אם אתם משוכנעים שאינכם צריכים את המשתנה המקורי.

4. בדיקת איכות: חובה לוודא שה recode (או ההיפוך) בוצעו נכון! יש להריץ FREQUENCIES או DESCRIPTIVE על המשתנה המקובץ. מספר הנבדקים שהם VALID במשתנה המקובץ (החדש) חייב להיות זהה למספר הנבדקים עם ערכים VALID במשתנה הרציף המקורי. אם המספרים שונים, זה אומר ששגיתם בהגדרת הגבולות של הקטגוריות, ונבדקים "נעלמו" (כלומר מבחינת SPSS הם כעת מוגדרים כבעלי "ערך חסר" במשתנה המקובץ החדש). אחת הסיבות למצב זה הוא שימוש במספרים עגולים מדי בהגדרת הטווח. כאמור, עדיף להשתמש בשברים עם כמה ספרות עשרוניות (ארחי הנקודה) כדי למנוע "רווח" מיותר בין המחלקות.

5. איך הופכים את הכיוון של משתנה? לעתים פריט מנוסח בכיוון שלילי או הפוך לשאר הפריטים בסולם. במצב כזה נדרש להפוך ערכים. למשל, בפריט לגביו הגיבו נבדקים בסולם 1 עד 7, יתכן ונרצה כי 7 יהפוך ל1, כי 6 יהפוך ל2, וכו'. דרך פשוטה לבצע היפוך היא ע"י פקודת COMPUTE (נמצאת תחת Transform). בחלון שנפתח, כותבים נוסחה שתהפוך בבת אחת את

Compute Variable	
Target Variable:	Numeric Expression:
V1new	= 8-v1

כל הערכים. למשל, במשתנה v1 עם ערכים בין 1-7, נכתוב: $V1new = 8 - v1$ (זה כמו אלגברה בתיכון... ואם הסולם בטווח 1-5, נכתוב $6 - v1$. עם נוסחה כזו, 5 נהייה 1 כי שש פחות חמש = 1, ו-4 הופך ל-2, וכו').

לחילופין, ניתן לבצע היפוך ע"י פקודת RECODE המוסברת למעלה. למשל לגבי פריט עם סולם תגובה בין 1 ל-7, נכתוב בפקודה כי הערך 7 יהפוך ל-1, 6 יהפוך ל-2, וכו'. זו פקודה מסורבלת יותר כאשר נדרש להפוך מספר רב של ערכים, לכן משתמשים בה רק כשיש להפוך מעט ערכים.

פרק ח': בדיקת התפלגויות וממוצעים בתת-קבוצות

1. רקע: אחרי שביצענו את השלבים המתוארים בפרקים א'-ז', אנו מוכנים להתחיל לבדוק לעומק את שאלות והשערות המחקר. אולם, כשלב מקדים בחקירת קובץ הנתונים, לרוב עלות שאלות תיאוריות המחייבות הצלבה של כמה משתנים זה עם זה, למשל: כמה גברים במדגם עובדים בתפקידי ניהול, בהשוואה לנשים? איזה אחוז של הנבדקים מועסק בארגונים עסקיים לעומת מלכריים? האם עובדים בעלי וותק שונה נבדלים בממוצע שחיקה?
התשובות לשאלות כאלו ואחרות דורשות שנחשב שכיחויות, אחוזים או ממוצעים בתת-קבוצות. זאת כי עוד לפני בדיקת ההשערות עצמן, אנו רוצים לדעת כמה אנשים (במספרים מוחלטים, באחוזים) קיימים בתת-קבוצות שונות במדגם ומה המאפיינים שלהם, או מה ערכי המרכז והפיזור של משתנים כלשהם בתת-קבוצות שונות. מידע על גדלי תת-קבוצות חיוני לנו כדי שנוכח לתאר את היקפן של קבוצות עיקריות בפרק ה"תוצאות", וכדי שנוכל לתכנן איך לבדוק את ההשערות שלנו, ולהבין באיזו מידה ממצאים (למשל ממוצעים, מתאמים) שחושבו בכלל המדגם, אכן מתקיימים גם בתת-קבוצות.

בפרק ד' כבר הוצגו פרוצדורות כגון Frequencies, Descriptive אשר מאפשרים חישוב התפלגויות וממוצעים. **כאן בפרק ח'** נתייחס לשתי פרוצדורות נוספות ב SPSS מאפשרות ביצוע פעולות בתת-קבוצות או בדיקת גודלן של תת-קבוצות במספרים מוחלטים או אחוזים: "צילוח" (הצלבת לוחות) ע"י CROSSTABS, או הרצת עיבודים בתת-קבוצות ע"י SPLIT FILE.

2. צילוח (הצלבת לוחות/טבלאות) עם CROSSTABS: פרוצדורה זו מאפשרת בדיקת המספר או האחוז של המקרים בהצלבות של משתנים עיקריים שהם בדידים (למשל מה אחוז הגברים או הנשים שהם בעלי השכלה מעל או מתחת ל 12 שנות לימוד). (הערה: בעזרת לוחות שכיחויות של CROSSTABS ניתן גם לבדוק קיום קשר בין משתנים מסוגים מסוימים (כתחליף לבדיקת מתאם) ולבדוק מובהקות הקשר - ראו פרק ט' בהמשך לגבי מבחן יחי-בריבוע).

איך מריצים crosstabulation: נבחר Analyze → Descriptive statistics → Crosstabs. בתיבה שנפתחה, נסמן את המשתנה הראשון בו נרצה להשתמש ונעביר לתיבה Row(s), את המשתנה השני נעביר לתיבה Column(s). כעת נלחץ על Cells - בחלון שנפתח נסמן את המדדים שאנו מעוניינים בהם - לרוב נסמן (א) counts-observed (כדי לקבל את מספר הנבדקים שנמצאים בכל תא, כלומר בהצלבה של כל ערך בשני המשתנים), ובנוסף (ב) נסמן Percentages (לפי הצורך, נבחר מתוך: Row, Column, Total). נאשר Continue, ונלחץ OK. בחלון הפלט תופיע טבלת צילוח, עם נתונים בהתאם למה שביקשנו - למשל אחוזים של כל תא מתוך השורה או הטור. (כדי להקל על קריאת הפלט, לעתים עדיף להריץ בנפרד כמה צילוחים נפרדים, למשל אחד עם שכיחויות, ואחד עם אחוזים בשורות או בטורים).

הצעה: אם ברצונכם להצליב משתנה אחד עם כמה משתנים אחרים (כל אחד לחוד): הכניסו מספר משתנים לתיבת Row. כדי לבחון התפלגויות משותפות מורכבות יותר, למשל להצלבת שלושה משתנים, הכניסו את המשתנה השלישי כ"שכבה" חדשה לתיבת Layer המצויה מתחת לתיבות Row, Column.

3. בדיקות בתת-קבוצות בעזרת הוראת SPLIT FILE לפיצול זמני של הקובץ: הוראה זו מאפשרת לנו להריץ כל פרוצדורה שנבחר בכל אחת מתת הקבוצות של משתנה בדיד כלשהו.

דוגמאות: נניח שאתם רוצים לבדוק האם גברים ונשים נבדלים בממוצע של השתנה "שחיקה". כדי לבצע זאת, תעבדו בשני שלבים: תחילה תתנו ל SPSS הוראת SPLIT FILE (מוסבר בהמשך) בהתייחס למשתנה "מיגדר", ובהמשך תריצו פקודת Descriptive (שכבר הוזכרה בפרק קודם). השילוב הזה יגרום לכך ש SPSS יריץ את ה- Descriptive בנפרד בקרב גברים ונשים. באופן דומה, ניתן לבדוק בעזרת שילוב זה בין הוראת SPLIT FILE לבין פרוצדורה סטטיסטית את ההתפלגות (שכיחות) של משתנה מסוים בכל אחת מתת-הרמות של משתנה אחר, למשל לבדוק התפלגות "השכלה" בכל סוג/רמה של "מוצא".

3א. מה עושים כדי להריץ עיבודים בתת-קבוצות דרך SPLIT FILE? עובדים בשני שלבים:

1. **מודיעים SPSS מהו המשתנה שעליו צריך "לחלק" או לפצל את הקובץ.**
תחת תפריט DATA נבחר ב SPLIT FILE. נסמן את Compare Groups. נסמן בתיבת המשתנים ונעביר ימינה (עם החץ) את המשתנה בו אנו מעוניינים להשתמש לחלוקת הקובץ לתת קבוצות. (ניתן לעשות SPLIT על משתנה אחד או כמה משתנים).
2. **בוחרים בפרוצדורה סטטיסטית** בה אנו מעוניינים (תחת ANALYZE) ומריצים אותה.
דוגמה: רוצים להשוות התפלגות השכלה בין אנשים בשלוש רמות ותק (נמוך=1, בינוני=2, גבוה=3): בצעד 1 נגדיר את המשתנה "ותק" כמשתנה עליו מתבצעת חציית המדגם, בצעד 2 נבחר ב FREQUENCIES ונסמן את המשתנה "השכלה". נריץ את הפרוצדורה. נקבל פלט בן 3 חלקים: תחילה תופיע התפלגות לרמה הנמוכה ביותר של ותק (נניח שהיא "1"), אח"כ הרמה "2", ובסוף "3".
ניתן גם לבצע SPLIT על שניים או יותר משתנים. למשל אם אחד עם 3 רמות (ותק) ואחד עם שתי רמות (מין), נקבל פלט עם עד שישה חלקים (לכל הצלבה אפשרית, אם קיימים ערכים בכל רמה בקובץ). **שימו לב:** אם יש קבוצת נבדקים בעלי ערכים חסרים – הם יופיעו כקבוצה נוספת ונפרדת.

3ב. שימוש נכון בהוראת SPLIT FILE - צריך לזכור שני דברים:

- **הוראת SPLIT FILE היא "הנחיה מקדימה" ולכשעצמה אינה מבצעת דבר** (כשתסיימו להקליק במסך של ההוראה הזו, לא יקרה כלום...) – היא רק מודיעה ל SPSS שמרגע זה ואילך, כל פרוצדורה שתבחרו בה (תחת תפריט ANALYZE) צריכה להתבצע בנפרד בכל הקבוצות הקיימות במשתנה שאותו ציינתם במסך של הוראת SPLIT FILE. לכן, תראו שמהו קורה בפועל רק כשתריצו פרוצדורה סטטיסטית אחרי כן.
- **חייבים לבטל הוראת SPLIT FILE אם רוצים לחזור להריץ על כל המדגם:** זיכרו כי ההוראה בתוקף כל זמן שלא הודעתם אחרת - כלומר SPSS יבצע SPLIT על כל פרוצדורה שתריצו עד שפעולה זו תבוטל. לכן, אם רוצים לחזור ולהריץ על כל המדגם, יש לבטל את הוראת SPLIT FILE - מבצעים זאת על ידי חזרה לתפריט של SPLIT והחזרה של המשתנה שנבחר אל התיבה השמאלית, או ע"י לחיצה על reset.

פרק ט': בדיקת השערות לגבי קשר בין משתנים

1. רקע: הרבה שאלות מחקר מתייחסות אל המידה בה קיים קשר (חיובי או שלילי) בין משתנים. אם אתם מעוניינים לבדוק את קיומם של קשרים בין משתנים (בין אם יש או אין לכם השערות פורמליות בנושא), אתם צריכים לבחור בדרך מתאימה לכך. כבר הוסבר בפרק קודם כי ניתן לבדוק קשר בין משתנים ע"י הסתכלות על צילוח (טבלאות דו-כיווניות), או ע"י הפקת דיאגרמות פיזור (scatter) תחת תפריט Graphs. **אולם, כדי לקבל אומדן פורמלי לקיום קשר, גודלו, כיוונו, ומובהקותו, יש להשתמש בטכניקות נוספות.** בפרק זה נתייחס בקצרה לשלוש אפשרויות לבדיקת השערות לגבי קשר בין משתנים, מתוכן עליכם לבחור, בהתאם לסולם המדידה ומאפייני התפלגות המשתנה. להלן תאור כללי של שלוש האפשרויות - פרטים טכניים בהמשך.

- **חי בריבוע:** כפי שהוסבר בפרק קודם, כשרוצים לבדוק קיום קשר בין שני משתנים בדידים בעלי מספר קטן יחסית של ערכים (למשל 2 X 2 או 5 X 5) ניתן לבצע צילוח (Crosstabs) של המשתנים. הסתכלות על טבלת הצילוח, ובמיוחד בדיקת השכיחות בתאים השונים (בערכים מוחלטים או במיוחד אחוזים בשורה/בטור) נותנת תחושה לגבי המידה בה המשתנים קשורים אחד לשני. (למשל אם הרבה נבדקים מקובצים בתאים מסוימים לאורך האלכסון). אולם, לבדיקת מובהקות הקשר שרואים לכאורה בצילוח, צריך לבצע מבחן סטטיסטי נוסף – כאן נתייחס למבחן "חי-בריבוע" (Chi-Square).
נשתמש במדד קשר זה אם לפחות אחד מהמשתנים הוא נומינלי (בסולם שמי).

• **מתאם פירסון**: נשתמש במדד זה אם ברצוננו לבדוק קשר בין משתנים בסולם אינטרוולי או מנה (ratio). לרוב מדובר במשתנים רציפים או עם מספר רב של ערכים בדידים.

• **מתאם ספירמן**: נשתמש במדד זה לבדיקת קשר בין שני משתנים, אם לפחות אחד משני המשתנים הוא אורדינלי (בסולם סדר). חשוב במיוחד להשתמש במתאם ספירמן, ולא פירסון, אם במשתנה יש מספר קטן של רמות/ערכים או אם רוב הנבדקים מרוכז רק בחלק מהקטגוריות/ערכים (כלומר, יש "ties") בעוד השאר "ריקות" יחסית.

שימו לב להתלבטות אופיינית: אם אתם משתמשים במשתנים (אינדקסים) שמחושבים כממוצע של כמה פריטים בסולם ליקרט מסוג 1-5 או 1-7, אפשר לרוב להתייחס אל המשתנים הללו כרציפים ולהשתמש במתאם פירסון, למרות שהם מורכבים מפריטים שכל אחד לחוד הוא בסולם אורדינלי והוא באופיו בדיד. זאת כי אינדקס שמחושב מכמה מרכיבים אורדינליים לרוב מתנהג כמשתנה אינטרוולי אם הוא כולל יותר מ 2 פריטים ואם בפריטים שמרכיבים את האינדקס יש מספר רמות (ערכים) ופיזור סביר של נבדקים ברמות השונות. אם יש התלבטויות, התייעצו עם המרצה...או בחרו במתאם ספירמן.

2. חישוב חי בריבוע (χ^2 , Chi-Square): כאמור, מבחן זה מאפשר לבדוק את הקשר בין משתנים בדידים בעלי מספר קטן יחסית של ערכים. לא מקובל להשתמש בו לבדיקת הקשר בין משתנים רציפים או בעלי מספר רב של ערכים בדידים.

בהמשך למוסבר בפרק ח'2, להרצת המבחן, תחת Analyze נבחר: Crosstabs ← Descriptive. בתוך Crosstabs נבחר במשתנים שיוכנסו לשורה ולטור בצילווח, ואז נלחץ על Statistics. **אך כעת, בתיבה שתופיע נסמן את הריבוע Chi-Square ונלחץ Continue ואח"כ OK להרצה.** נקבל פלט ובו צילווח בין המשתנים שבחרנו ובהמשכו תוצאות מבחן Chi-Square. נסתכל בערך של Pearson Chi-Square וברמת המובהקות שלו (תחת Sig). כמו במבחני מובהקות אחרים, ערך שווה או נמוך מ 0.05 מראה על חי-בריבוע מובהק. **אך נזכור כי חי-בריבוע מובהק, לכשעצמו, איננו מסביר מהו הקשר.** זהו רק מבחן מובהקות האומר לנו האם קיים (או לא קיים) קשר (תלות) בין שני משתנים בלוח. **כדי להבין את מאפייני הקשר וכיוונו, חייבים גם להסתכל בצילווח ולפרש את הנתונים הגולמיים: נבדוק את השכיחויות (counts) בתאים בטבלה הדו-כיוונית, ונסתכל על האחוזים של כל תא מתוך הסה"כ בשורה ובטור.**

שימו לב! שימו לב בפלט להערות שמתחת לטבלה! מקובל שהשכיחות המצופה בתא כלשהו (expected count) לא תהיה קטנה מחמישה אנשים. אך אם מצוין שיש תא (cell) אחד או יותר בהם יש פחות מ 5 תצפיות, יש בעיה בתוצאות המבחן ויתכן שערך החי-בריבוע מטעה.

3. חישוב מקדם מתאם (פירסון או ספירמן): אם רוצים לבדוק מתאם בין משתנים - תחת Analyze נבחר: Bivariate ← Correlate. בחלון שיפתח, נבחר תחילה את המשתנים אשר ביניהם אנו מעוניינים לבדוק את המתאם ונעביר לתיבה Variables. (אפשר לבחור יותר משני משתנים). מתחת לחלון בחירת המשתנים, נבחר את סוג המתאם (פירסון, ספירמן וכו'), לפי סוג המשתנים בהם מדובר (רציפים או בדידים). בנוסף ניתן לבקש ממוצעים וסטיות תקן לכל אחד מהמשתנים אם נלחץ על Options. בסוף הפעולה נלחץ OK, ונקבל פלט בחלון Output.

התראה - בחישוב מתאמים יש להשתמש במשתנים רציפים, לא מקובצים!
 כאשר בודקים קיום מתאם בין משתנים, **לא** נשתמש במשתנים מקובצים (אלו שהליך יצירתם הוסבר בפרק ז') כי פעולת הקיבוץ גורמת לנו לאבד מידע - היא מביאה לכך שאנשים שהיו בעלי ערכים שונים זה מזה במשתנה הרציף מקבלים את אותו ערך בדיד (במשתנה המקובץ). לכן יש לחשב מתאמים (או לבצע עיבודים אחרים כגון חישוב ממוצעים או סטיות תקן) רק לגבי משתנים רציפים.
מתי נשתמש במשתנים המקובצים? רק כאשר אנו רוצים לבדוק התפלגות משותפת (ע"י ביצוע צילווח) או כאשר רוצים לבדוק ממוצעים של משתנה רציף כלשהו בתת-קבוצות של משתנה רציף אשר בו אנו יוצרים תת-קבוצות על ידי קיבוץ.

א. הבנת המידע הסטטיסטי בפלט מתאם : בפלט יופיעו לגבי כל מתאם שלושה פריטי מידע. (להלן הסבר לא-פורמלי, שאינו תחליף לקריאת חומר מהקורס בסטטיסטיקה. נא לעשות שיעורי בית!).

- תחילה יהיה רשום מקדם המתאם (שכידוע יכול להיות חיובי או שלילי ובעל ערכים שונים בין 1- עד 1+). ליד המתאם עשויות להופיע כוכביות שמספקות התראה ויזואלית כאשר מתאם מובהק ברמה מסוימת (המקרא מופיע בתחתית הטבלה).
- מתחת למקדם המתאם תהיה רשומה רמת המובהקות (Sig.) של המתאם המסוים
- מתחת לרמת המובהקות רשום מספר הנבדקים N עליהם חושב המתאם.

רמת מובהקות מוצגת בד"כ כהסתברות (למשל $P < .02$). במדעי החברה וההתנהגות מקובל לראות מתאם כמובהק אם ההסתברות שלו שווה/נמוכה מ 0.05 (כלומר, P בערכים של 0.05, 0.01, 0.001, וכו')

ב. פירוש ביקורתי של מקדם המתאם : להלן הסבר קצר וראשוני אשר מחלק את פירוש מקדם המתאם לשלושה שלבים. (זהו הסבר לא-פורמלי, שאינו תחליף לקריאת חומר מהקורס בסטטיסטיקה. נא לעשות שיעורי בית!).

- 1. משמעות מובהקות מתאם : אם הפלט מציין כי המתאם מובהק – מה זה אומר? מה תוכלו לכתוב בפרק התוצאות ו/או הדיון?** כאשר מתאם הוא מובהק כפי שמוסבר למעלה, לחוקר מותר לטעון כי יש קשר מובהק בין המשתנים או לחילופין שהקשר בין המשתנים שונה מאפס. (כלומר מתאם מובהק משמעו שהמתאם אינו "מקררי" אלא כנראה מייצג קשר אמיתי בין המשתנים – קשר שהוא שונה מאפס). עם זאת, חובה לזכור שמובהקות מקדם מתאם, כמו במדדים סטטיסטיים אחרים, תלויה בגודל המדגם. מתאם גבוה יחסית לא יהיה מובהק אם התקבל במדגם קטן (כי יתכן שהוא התקבל במקרה בגלל מאפייני הנבדקים המועטים במדגם). מנגד, גם מתאם נמוך יחסית יהיה מובהק אם חושב במדגם גדול (כי זה אומר שיש תופעה יציבה של קשר בין משתנים). לפיכך, זה שהמתאם נמצא מובהק אינו אומר שיש קשר חזק או משמעותי בין המשתנים – וכאן עוברים לשלב הבא.
- 2. עוד על פירוש משמעות מקדם המתאם : כאשר חוקרים מנסים להבין את משמעות תוצאות בדיקת מתאם בין משתנים, הם מתייחסים לכמה נושאים במקביל: גודל, כיוון, ומובהקות המתאם, וגודל המדגם.** כלומר, פירוש מקדם המתאם לא מתייחס רק לרמת המובהקות אלא גם לגודלו וכיוונו של מקדם המתאם (המייצג את כיוון וגודל הקשר בין המשתנים). אולם, חובה לזכור כי העובדה שמתאם מובהק אינה אומרת שיש בהכרח קשר חזק בין משתנים. למשל אם המדגם גדול מאוד (כמה מאות נבדקים), מתאם יכול להמצא מובהק למרות שהוא חלש או בינוני (בין אם הוא שלילי או חיובי). בנוסף חשוב גם להתייחס בזמן פירוש המתאם, במיוחד אם נמצא מתאם חלש או לא מובהק, לאפשרות שהקשר בין המשתנים הוא לא-לינארי (למשל שיש עקומת U). לסיום, יש לזכור שלעיתים תמצאו מתאם שאינו מובהק אך בכל זאת גודלו וכיוונו תומכים בהשערות שלכם. לכך עשויה להיות (אולי...) חשיבות ויתכן שתבחרו להציג זאת (בזהירות ועם ההסתייגויות הנדרשות) בפרק התוצאות ו/או בפרק הדיון כאשר תתייחסו למשמעות הממצאים.

- 3. קשר וסיבתיות :** לסיום, יש לזכור כי מתאם (מובהק) בין משתנים אינו מייצג בהכרח קשר סיבתי בין משתנים אלו. יש להזהר באופן בו תציגו התוצאות ומפורשים הממצאים – רצוי לא לטעון שמשנתנה מסוים גורם לשינוי במשתנה אחר או משפיע על משנתנה אחר, אלא אם כן יש לכם בסיס נוסף (תיאורטי או אמפירי) לטענה לגבי השפעה סיבתית, למשל בהתייחס לאופיים וטבעם של המשתנים, סדר הזמנים, שלילה של הסברים חלופיים (כולל השפעת משתנים מתערבים או מתווכים), וכד', כפי שלמדתם בקורס שיטות מחקר.

פרק י': בדיקת השערות לגבי הבדלים בין קבוצות

1. רקע: פרק זה עוסק בבדיקת הבדלים בין קבוצות. הבדלים בין קבוצות עשויים להיות מכמה סוגים – להלן תאור של כמה מצבים אופייניים בהם נעסוק בפרק זה:

א. הבדלים בין שתי קבוצות (במשתנה הבלתי-תלוי): למשל, נרצה לבדוק את הקשר בין רמת שכר לבין מגדר באמצעות בדיקת ההבדל בין השכר הממוצע בקבוצת הגברים לבין השכר הממוצע בקבוצת הנשים.

ממוצע שכר	ממוצע שכר
גברים	נשים

ב. הבדלים בין שלוש או יותר קבוצות: כאן יתכנו מספר מצבים:

ב1. כאשר ההשערה מתייחסת למשתנה בלתי תלוי אחד בו יש שלוש או יותר קבוצות: למשל, נרצה לבדוק את הקשר בין גיל לבין הנכונות לאמץ חידושים, כאשר המשתנה גיל מחולק לשלוש רמות (צעירים, בוגרים וקשישים). בפועל, נבדוק את ההבדלים בין הנכונות הממוצעת לאמץ חידושים של אנשים הנכללים בכל אחת משלוש רמות הגיל.

ממוצע נכונות לאמץ	ממוצע נכונות לאמץ	ממוצע נכונות לאמץ
-------------------------	-------------------------	-------------------------

צעירים בוגרים קשישים

ב2. כאשר ההשערה מתייחסת לכמה משתנים בלתי-תלויים: למשל, נרצה לבדוק את הקשר בין המגדר והוותק של עובדים (שני משתנים בלתי-תלויים) לבין מסוגלות עצמית (משתנה תלוי). לכל אחד מהמשתנים הבלתי-תלויים מוגדרות שתי רמות: מגדר (גבר/אשה), ותק העובד (חדש/ותיק). כך, נוצרות ארבע קבוצות, כמוצג בתרשים למטה. בפועל, נרצה לבדוק את ההבדלים מבחינת מסוגלות עצמית בין הקבוצות: נשים חדשות, נשים ותיקות, גברים חדשים, גברים ותיקים.

	צעירים	ותיקים
נשים	ממוצע מסוגלות עצמית	ממוצע מסוגלות עצמית
גברים	ממוצע מסוגלות עצמית	ממוצע מסוגלות עצמית

פרק זה יסביר כיצד לבדוק את ההבדלים בין קבוצות מהסוגים שהוזכרו. עבור כל סוג של הבדל בין קבוצות, קיימת שיטה סטטיסטית מועדפת לבדיקת הבדלים. (יודגש כי ההסברים בפרק הם לא-פורמליים ואינם תחליף לחזרה על חומר מקורס בסטטיסטיקה. מומלץ לעיין בספר סטטיסטיקה מתאים או להתייעץ עם מורה הסמינר כדי לאתר מידע נוסף שסייע בפירוש הפלטים והבנת הממצאים שלכם).

תזכורת לגבי רגרסיות: דרך נוספת לבדיקת השערות תלת-משתניות היא הרצת רגרסיה

(regression) בה בודקים את המידה בה שני המשתנים הבלתי-תלויים (למשל בהשערה שמתוארת בסעיף מס' 3: מגדר, ותק) מנבאים את המשתנה התלוי (למשל: מסוגלות עצמית). חלקכם למד על רגרסיה ויודע להבין איך לפרש פלטים הקשורים לרגרסיה - אולם הניסיון מראה כי למרות שהרבה תלמידים למדו על רגרסיה, קשה להם להסביר את משמעות הממצאים. לכן, אנחנו נתמקד בפרק זה רק במבחני t וניתוח שונות (שנקרא גם מבחן F ו-ANOVA).

2. בדיקת הבדלים בין שתי קבוצות: מבחן t למדגמים בלתי תלויים

א. רקע: הרבה שאלות מחקר מתייחסות להבדל בין שתי קבוצות במוצעים של משתנה כלשהו. למשל חוקרים רוצים או צריכים לבדוק את ההבדלים בין גברים לנשים, או בין בעלי תפקיד ניהולי ולא-ניהולי. אם מתעניינים בהבדל בין שתי קבוצות (בלבד), אפשר להשתמש בפרוצדורת t-test למדגמים בלתי-תלויים. (המונח "בלתי-תלויים" לא צריך לבלבל - כאשר משווים, למשל, בין ממוצע של גברים ונשים במשתנה כלשהו, הנחת המוצא היא שיש לנו "מדגם" נפרד של גברים ו"מדגם" נפרד של נשים. נתייחס אל שני המדגמים כנלקחים משתי אוכלוסיות שונות, לא תלויות אחת בשניה - כלומר בעלות ממוצע שונה, סטיית תקן שונה ולפיכך שונות (variance) שונה).

ב. כיצד נפעיל פרוצדורת t-test ? Analyze ← Compare means ← Independent samples t-test

1. **בחירת משתנים:** לתיבה העליונה (Test variable) נכניס את המשתנה התלוי (שאת הממוצע בו אנו רוצים להשוות בשתי הקבוצות). לתיבה התחתונה (Grouping variable) נכניס את שמו של המשתנה הבלתי-תלוי.
2. **נלחץ Define groups** ונגדיר את רמות המשתנה הבלתי-תלוי. המבחן משווה בין שתי קבוצות. נרשום בשתי התיבות מהו הערך של Group 1 ומהו הערך של Group 2. לדוגמה, אם רוצים להשוות גברים לנשים ובמשתנה GENDER יש שני ערכים, 1 = גברים, 2 = נשים, נכתוב "1" בקבוצה 1 ו-"2" בקבוצה 2.

הערה טכנית לגבי ההכנה הנדרשת להוצאת פרוצדורת t-test: צריך שהמשתנה הבלתי-תלוי יהיה בדיד (לא רציף) ויכלול מספר קטן של קבוצות, לרוב שתיים. אם במשתנה הבלתי-תלוי יש יותר קבוצות (למשל אם קיבצתם את הנבדקים לפי 3 רמות השכלה), פרוצדורת מבחן t מאפשרת לכם לבחור את שתי הרמות שביניהן נשווה את ממוצעי המשתנה התלוי: בתיבת Groups תוכלו לבחור להשוות אנשים שיש להם ערך 1 לעומת 2, או 2 לעומת 3, וכד'. אולם, אם המשתנה הבלתי תלוי הוא רציף, יש תחילה לבצע הקבצה – ראו פרק ז'.

ג. פירוש התוצאות: הפרוצדורה מציגה תחילה ממוצעים וסטיות תקן בשתי הקבוצות (כמו שקיבלנו כשהפעלנו Descriptive אחרי split file), ורק אז מציגה נתונים לגבי מובהקות ההבדל. (ראו בהמשך דוגמא לפלט של מבחן t למדגמים בלתי תלויים).

1. **ממוצעים:** תמיד נבחן תחילה - למי משתי הקבוצות ממוצע גבוה יותר, והאם ההבדל ביניהם הוא בכיוון ההשערה או בכיוון מנוגד לה? נסתכל על ההבדל בין הממוצעים, במונחי סטיות תקן. הבדל קטן בממוצעים (למשל 3.50 לעומת 3.60) עם סטיית תקן גדולה יחסית (למשל 0.8) מראה כי הממוצעים קרובים מאוד, יחסית לפיזור בכל קבוצה. הבדל כזה אינו משמעותי, ודי בטוח שלא מובהק סטטיסטית (אלא אם המדגם גדול מאוד). הבדל גדול בין הממוצעים (למשל, מעל 1 סטיית תקן) יהיה משמעותי ולרוב מובהק גם במדגם קטן.
2. **המובהקות הסטטיסטית של ההבדל:** רק אחרי שבדקנו מהו גודל וכיוון ההבדל בין הממוצעים בשתי הקבוצות, נוכל לבדוק באופן פורמאלי באיזו מידה ההבדל בין הקבוצות הוא מובהק. נפעל בשני שלבים:

א2. בדיקת מבחן שוויון השונויות (Levene's test for equality of variances): תחילה נתייחס לעיבוד מקדים זה, אותו SPSS מבצע כדי לקבוע לעצמו באיזו נוסחה חישובית של t יש להשתמש, זאת כי יש מספר ואריאציות למבחן t המתאימות לשימוש במצבים שונים. "מבחן שוויון השונויות" נדרש כדי לדעת אם שתי הקבוצות שמושוות הן בעלות שונות (פיזור) דומה מספיק, או שהן שונות בפיזור – כי בכל מצב, התיאוריה הסטטיסטית מחייבת (את SPSS) להשתמש בנוסחה חישובית שונה.

מה בודקים? מבחן שוויון השונויות מופיע לפני תוצאות מבחן t והוא מבוסס על מבחן F (אשר מאפשר בדיקת מובהקות במצבים מגוונים – במקרה הזה הבדל בין מדדי פיזור). למעשה, התוכנה מחשבת לנו מראש שתי תוצאות אפשריות, ואנו נבחר מי המתאימה:

אם מבחן F לא מובהק ($sig > .05$) זה אומר שמותר להניח שהשונויות שוות, ואז נסתכל על תוצאות מבחן t בשורה הראשונה - Equal variances assumed. אם מבחן F מובהק ($sig < .05$), זה אומר כי יש הבדל מובהק בין השונויות בשני המדגמים (הפיזור בהם שונה), ואז נסתכל על תוצאות מבחן t בשורה השנייה - Equal variances not assumed.

2. בדיקת תוצאות מבחן t: אחרי שקבענו באיזו שורה להסתכל על בסיס תוצאות מבחן שוויון השונויות, נוכל לבדוק את תוצאות מבחן ה-t. שתי עמודות חשובות לנו: בעמודת "t" מופיע ערך t, ובעמודת "sig" המובהקות של t. אם ערך ה"sig" יהיה 0.05 או נמוך מכך (תלוי ברמת המובהקות בה אנו מעוניינים), זה אומר שבין הממוצעים יש הבדל מובהק (כי ערך ה"t" מספיק גבוה). את ערכי "t" ו"sig" תדווחו בפרק התוצאות, יחד עם דרגות החופש (df) – אך תמיד הכרחי גם לדווח את הממוצעים וסטיות התקן בשתי הקבוצות, ואת הגודל שלהם, כי רק כך אפשר למעשה להבין מה אופיו של ההבדל בין הקבוצות.

3. הערות נוספות לגבי פירוש פלט מבחן t והבנת משמעות הממצאים:

- בגלל נוסחת מבחן t, ערך t יכול להיות שלילי או חיובי, תלוי איזה ממוצע גבוה יותר. כאשר ממוצע קב' 1 גבוה יותר, t יהיה חיובי; כאשר ממוצע קב' 2 גדול יותר, t יהיה שלילי. היות והבחירה איזו קבוצה היא 1 או 2 היא שרירותית ותלויה רק באופן בו קודדו הנתונים, הסימן של t (שלילי או חיובי) אינו חשוב לכשעצמו. **מה שחשוב הוא זהות הקבוצה בה הממוצע גבוה יותר**, והאם t מובהק או לא (ללא קשר לסימן שלו).
- יש לזכור כי מבחני המובהקות מבוצעים כדו-צדדיים (מחמירים יותר) ולא חד-צדדיים. אם הנושא חשוב לכם, ניתן לשנות את צורת הבדיקה בתוך התפריט.
- יש לזכור כי ביצוע הפעולות המתוארות במדריך זה אינו מבטיח כי תבחרו בעיבוד הנכון או תפרשו את תוצאותיו בצורה נכונה. תמיד יש להפעיל שיקול דעת וחשיבה ביקורתית, לחזור על חומר הלימוד מהקורס בסטטיסטיקה, ולהתייעץ עם מורה הסמינר או מנחה המחקר בהתאם לנדרש.

ד. דוגמה לפירוש פלט של t-test:

להלן הסברים קצרים לסיוע בפירוש פלט של פרודצורת t-test ב SPSS. כאמור, פרודצורת t-test (independent samples) מאפשרת להעריך אם ההבדל בין ממוצעי שתי קבוצות במשתנה תלוי כלשהו הוא מובהק. הדוגמה כאן מתייחסת לשלושה שלבים בהשוואת גברים (קב' 1) ונשים (קב' 2) במשתנה התלוי "Efficacy". שימו לב כי כאן שתי הקבוצות הן לא באותו גודל, אך בכל זאת שתיהן בגודל סביר להערכת ההבדלים ביניהן. אך לעתים, אחת הקבוצות קטנה מדי, לכן רצוי תמיד לשים לב למספר הנבדקים בכל קבוצה (N). (הערה – הפלט עצמו בעמוד הבא "חתוך" בצד ימין וגולש בגלל פגם (bug) בתוכנת SPSS. מה שמוצג הוא החלק החשוב).

שלב 1: בידקו כאן למי משתי הקבוצות ממוצע גבוה יותר, והאם ההבדל בין הקבוצות הוא בכיוון המשוער או בכיוון מנוגד. התייחסו לגודל ההבדל ביחס לגודל סטיות התקן. (במקרה הנוכחי, ממוצע קב' 2 (נשים) גדול במקצת מהממוצע של קב' 1 – אך ההבדל שולי יחסית לגודל סטיות התקן (והוא הבסיס לממצא בשלב 3).

T-Test

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
EFFICACY	1	113	6.3319	.6343	5.967E-02
	2	253	6.3834	.5818	3.658E-02

תזכורת: רק אחרי שבדקנו מהו גודל וכיוון ההבדל בין הממוצעים בשתי הקבוצות, נוכל לבדוק באופן פורמלי באיזו מידה ההבדל שנמצא בין הקבוצות הוא מובהק, ועל כך בשלבים 2 ו-3 למטה.

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
EFFICACY	Equal variances assumed	1.725	.190	-.761	364	.447	-5.154E-02	6.771E-0
	Equal variances not assumed			-.736	199.491	.462	-5.154E-02	6.999E-0

שלב 2: שלב מקדים זה מאפשר לנו להחליט איזו נוסחה חישובית של t היא הנכונה לשימוש. אם F לא מובהק ($sig > .05$), כמו במקרה הנוכחי, זה אומר שמותר להניח שהשונויות שוות, ואז נסתכל על השורה הראשונה - Equal variances assumed. אם מבחן F היה מובהק ($sig < .05$), זה היה אומר כי השונויות בשני המדגמים שונות באופן מובהק ואז צריך נוסחה חישובית שונה של t; במצב כזה היה צריך להסתכל בשורה השנייה.

שלב 3: לאחר בדיקת תוצאות מבחן שוויון השונויות, נבדוק את תוצאות מבחן t. ההבדל בין הממוצעים יחשב כמובהק אם ערך t מספיק גבוה - נדע זאת אם ערך ה"sig" נמוך מרמת המובהקות בה אנו מעוניינים (למשל $sig < .05$). כאן, ערך t הוא נמוך (בין 0 - 1), ולא מובהק כי הוא .447 ($sig > .05$).

3. בדיקת הבדלים בין שלוש או יותר קבוצות: ניתוחי שונות (ANOVA)

כפי שתואר בקצרה בתחילת פרק י', בדיקת ההבדלים בין שלוש קבוצות ומעלה דורשת שיטת ניתוח שונה מאשר בדיקת ההבדלים בין שתי קבוצות בלבד (הנערכת באמצעות t-test). השיטה המוצגת כאן הינה "ניתוח שונות", ונקראת גם מבחן F או ANOVA. בפרק זה נתייחס לשני סוגים בסיסיים של ניתוח שונות:

- א. ניתוח שונות [מבחן F] חד-כיווני (one-way analysis of variance)
- ב. ניתוח שונות [מבחן F] דו-כיווני (two-way analysis of variance)

הערה: אם אינכם מכירים מבחנים מתקדמים כאלו, אל תחששו לנסות דברים חדשים! אם אתם מכירים את מבחן t (המוסבר בסעיף קודם) אך לא את מבחן F, תוכלו לבצע **מספר מבחני t-test על זוגות של תת-קבוצות** (למשל להשוות את ממוצע השחיקה של בעלי שביעות רצון "נמוכה" ו"בינונית" ובנפרד להשוות את הקבוצות "בינונית ו"גבוהה"). חשוב לציין כי שימוש במספר מבחני t באופן כזה לא מומלץ בספרות, כי הוא מגדיל את הסיכוי לטעות בהסקה סטטיסטית. אולם, בסמינר מחקר לימודי הדבר עשוי להיות קביל (התייעצו עם המורה!), כל זמן שהממצאים מפורשים בזהירות הראויה ואתם מודעים לכך שיש דרך עדיפה אותה תוכלו ללמוד בקורס מתקדם. עדיף שתשתמשו בטכניקה המוכרת לכם (גם אם אינה אלגנטית מבחינה סטטיסטית), מאשר בטכניקה מתקדמת שאינכם מבינים לעומק.

לחילופין, תוכלו לנסות להריץ מבחני שונות לפי ההסבר בהמשך, והתייעצו עם המורה כדי לוודא כי אתם מפרשים נכון את התוצאות.

א. ניתוח שונות חד-כיווני (One-Way ANOVA)

1. רקע: נשתמש בניתוח שונות חד-כיווני כאשר נרצה לבחון איזו השפעה יש למשתנה בלתי-תלוי אחד שבו שלוש קבוצות ומעלה, על משתנה תלוי. ניתוח השונות מאפשר לנו להשוות ולבחון את ההבדלים בין הממוצעים של הקבוצות הללו.

דוגמה (נשתמש בדוגמה זו במהלך כל הפרק): חוקר רוצה לבחון את הקשר בין גיל לבין שביעות רצון בעבודה. שאלת המחקר היא: האם יש הבדל בין עובדים בשלוש הקבוצות גיל ("צעירים", "גיל ביניים" ו"בוגרים") מבחינת שביעות רצון בעבודה? שימו לב כי בדוגמה זו, החוקר מודד משתנה רציף, גיל בשנים, אך לא משתמש בו בצורתו הרציפה, אלא צריך לבצע הקבצה של המשתנה (כלומר של הנבדקים) לשלוש קבוצות גיל.

הערה טכנית לגבי הצורך בהקבצה: יתכן ובמחקר שלכם, אחד המשתנים הבלתי-תלויים שלכם הוא אינו רציף כמו בדוגמה, אלא משתנה בדיד עם 3 או יותר קבוצות (כגון מצב משפחתי). במצב כזה, לא נדרשת פעולת ההקבצה המוסברת בשלב 1 של הסעיף הבא.

2. כיצד נפעיל פרוצדורת ניתוח שונות חד-כיווני?

נעבוד בשני שלבים: הקבצת המשתנים הרציפים, ולאחר מכן הרצה של ניתוח השונות:

שלב 1: הקבצת משתנים רציפים: אם שאלת המחקר מתייחסת למשתנה רציף (למשל גיל, ותק, או כל אינדקס שמחושב כממוצע של מספר פריטים בסולם מדידה כלשהו), עליכם לבצע הקבצה במשתנה זה, כדי לחלק את הנבדקים לכמה רמות (למשל נמוך, בינוני, גבוה). **ביצוע הקבצה של משתנים רציפים מוסבר במדריך זה בפרק ז' – עיינו שם כעת!** בדוגמא בהמשך, על בסיס בדיקת פלט התפלגות הגיל, המשתנה "גיל" עבר הקבצה (recode) לשלוש רמות (1=צעירים עד 24, 2=גיל ביניים 25-40, 3=בוגרים +41).

שלב 2: הרצת ניתוח שונות חד-כיווני:

בתפריט Analyze נלחץ על `One-Way ANOVA ← Compare means` בתיבת הדיאלוג שתופיע, יש לבצע את הצעדים הבאים:

- לבחור ברשימה משמאל את המשתנה התלוי ולהעביר למשבצת העליונה (Dependent List).
- לבחור ולהעביר את המשתנה הבלתי-תלוי למשבצת התחתונה (Factor).
- ללחוץ על Options ולסמן Descriptive כדי לקבל טבלת ממוצעים וסטיות תקן בפלט – הדבר הכרחי כדי שתהיו מסוגלים לפרש את התוצאות!
- ללחוץ Continue להרצה. התוצאות יופיעו בחלון הפלט הנפרד.

3. פירוש התוצאות: כדי לפרש את הפלט, יש לעבוד בשלושה שלבים המוסברים להלן.

שלב 1: הסתכלות בטבלת הממוצעים. להלן דוגמה לטבלה כזו, שנוצרה לגבי שאלת המחקר בדוגמא למעלה בה המשתנה "גיל" עבר הקבצה לשלוש רמות. בטבלה זו נסתכל על הממוצעים של שלושת הקבוצות, על סטיות התקן (הפיזור), וגדלי הקבוצות (לוודא שיש מספיק נבדקים בכל קבוצה כדי שהממצאים לגביה יהיו אמינים?).

שביעות רצון

	N	Mean	Std. Deviation
1= 16-24	119	4.9202	1.36308
2=25-40	201	4.7575	1.14174
3=41-70	41	5.4280	.98088
Total	361	4.8873	1.21786

מה נחפש בטבלה ?

לאיזו קבוצה יש את הממוצע הגבוה ביותר? הנמוך ביותר? (בדוגמה: האם שביעות הרצון הממוצעת עולה ככל שהגיל עולה? שימו לב לסדר הממוצעים- כאן מתקבלת עקומת U!)

שלב 2: בדיקת טבלת התוצאות של ניתוח שונות חד-כיווני :

נבדוק האם ערך F הוא מובהק, בעזרת שתי עמודות: עמודת "F" ועמודת "sig". אם ערך ה"sig" הוא 0.05 או נמוך מכך (תלוי ברמת המובהקות בה אנו מעוניינים), זה אומר לנו שבין הממוצעים של הקבוצות יש הבדל מובהק (כי ערך ה"F" מספיק גבוה). (את ערכי "F" ו"sig" נדווח בפרק התוצאות, יחד עם דרגות החופש (df). בדוגמה שלהלן, המשתנה התלוי הוא "שביעות רצון בעבודה" (שמו רשום בראש הטבלה), ובדקנו את השפעת המשתנה "גיל", כפי שהוסבר למעלה. ניתן לראות בטבלה שערך F מובהק (sig נמוך מ-0.05). משמעות הדבר כי יש הבדל מובהק בין עובדים בשלוש קבוצות הגיל מבחינת תפישת "שביעות רצון בעבודה".

ANOVA

שביעות רצון

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15.506	2	7.753	5.354	.005
Within Groups	518.441	358	1.448		
Total	533.946	360			

שימו לב: בדיווח הממצאים, לא ניתן לכתוב רק "נמצא הבדל מובהק בין הקבוצות...". אלא **חובה גם לציין גם את הממוצעים וסטיות התקן בקבוצות השונות** – רק כך תוכלו להבין ולהסביר מי גבוה ממי, והאם מה שמצאתם דומה או שונה למה שמופיע בהשערות. הכרחי לבדוק את הממוצעים וסטיות התקן (של המשתנה התלוי) בכל הקבוצות ולהציגם בדיווח הממצאים, כדי שקורא דו"ח המחקר שלכם יוכל להבין מה מצאתם. (ראו דוגמאות לדיווח בהנחיות נפרדות – הן לא חלק ממדריך זה).

שלב 3: ניתוחי המשך (Post Hoc) - מומלץ, אך לא חובה (תלוי בתוצאות שלב 2)

גם אם נמצא ערך F מובהק בשלב 2, איננו יודעים **בין אלו קבוצות קיים ההבדל**. (אולי שתי קבוצות בעצם זהות זו לזו והשלישית שונה מהן במובהק? או אולי כל קבוצה שונה במובהק מהקבוצות האחרות?). כדי להבין מה בעצם קורה בתוך הנתונים, צריכים לרוב בצע ניתוח המשך פשוט, שמטרתו אינה לאושש את קיומו של ההבדל (את זה כבר ביססנו בשלב 2 ע"י ערך F מובהק), אלא לאתר את מיקומו של ההבדל.

ביצוע ניתוחי המשך הוא פשוט: נלחץ על כפתור Post Hoc... כאשר מפעילים את פרוצדורת ניתוח שונות חד-כיווני (מתחת ל Factor). מהרשימה העליונה (של Equal Variances Assumed) יש לבחור את המבחן אותו תרצו לבצע. [שימו לב: המבחנים נבדלים זה מזה במספר ההשוואות שהם עורכים בין הקבוצות, ובסיכוי לטעות מסוג I. להבנת נושא זה לעומק, יש לחזור על החומר שלמדתם בקורס סטטיסטיקה!]

בדוגמה למטה, בחרנו בניתוח מסוג בונפרוני (Bonferroni). בבלט SPSS נוצרת טבלה שנקראת Multiple Comparisons, בה ניתן לראות את ההבדלים בין הקבוצות:

Multiple Comparisons

שביעות רצון: Dependent Variable: Bonferroni

(I) Age group	(J) Age group	Mean Difference (I-J)	Std. Error	Sig.
1= 16-24	2=25-40	.16271	.13919	.730
	3=41-70	-.50788	.21792	.061
2=25-40	1= 16-24	-.16271	.13919	.730
	3=41-70	-.67059*	.20622	.004
3=41-70	1= 16-24	.50788	.21792	.061
	2=25-40	.67059*	.20622	.004

*. The mean difference is significant at the .05 level.

בניתוח זה, כל קבוצה במשתנה הבלתי-תלוי (עמודה I) משווה לכל קבוצה אחרת (עמודה J), וגודל ההבדל בין הממוצעים מוצג בטור הראשון. **יש להסתכל בעמודת Sig. כדי לבחון איזו השוואה מובהקת.** בדוגמא זו, המתייחסת למצב בו יש 3 קבוצות במשתנה הבלתי-תלוי, רואים כי יש הבדל מובהק בין קבוצות 1-3 ו 2-3, אך לא בין 1-2. כלומר למעשה קבוצה 3 היא זו הגורמת ל F המובהק הכללי שנמצא קודם בשלב 2 (ראו למעלה)!

ב. ניתוח שונות דו-כיווני (Two-Way ANOVA)

1. רקע: לעתים, חוקרים מעלים השערות המניחות קיום קשר בין כמה משתנים בלתי-תלויים לבין משתנה שלישי (משתנה תלוי). לדוגמה:

השערה 1: "עובדים בעלי תחושה גבוהה של משמעות בעבודה, ושהם ותיקים יותר בעבודה, יהיו עם שביעות רצון גבוהה יותר מעובדים עם משמעות וותק נמוכים".

השערה 2: "ימצא קשר חיובי בין אמון הלקוח בספק השירות ותפיסת מסוגלות עצמית של הלקוח, לבין נאמנות הלקוח לספק השירות"

שימו לב שבהשערות אלו, כל המשתנים רציפים (כל אחד נמדד על ידי סולם מדידה המבוסס על כמה פריטי דירוג מסוג ליקרט). כל השערה מניחה כי לשני משתנים בלתי-תלויים (למשל אמון ומסוגלות, בהשערה 2) יש השפעה על משתנה רציף שלישי (למשל נאמנות, בהשערה 2).

איך בודקים השערות כאלו מבחינה סטטיסטית? גישה מקובלת היא להשתמש ברגרסיות. אולם, למרות שתלמידים רבים אכן לומדים ניתוחי רגרסיה, לרבים קשה להבין איך להריץ אותם, או איך לפרש את משמעות התוצאות. לכן, אני מעדיף להסביר כאן איך לבדוק קשר בין שלושה משתנים על ידי ניתוח שונות דו-כיווני. דרך זו פשוטה להבנה (מנקודת הראות של התלמיד) וקלה להצגה והסבר (בדו"ח המחקר), ונשענת על ההסברים שניתנו בפרקים קודמים במדריך זה.

שלב מקדים לביצוע ניתוח שונות דו-כיווני: יש צורך לבצע הקבצה (grouping, categorization) של המשתנים הבלתי-תלויים (הרציפים) שלנו לשתי רמות (למשל נמוך-גבוה), או שלוש רמות (נמוך-בינוני-גבוה). **ביצוע הקבצה של משתנים רציפים מוסבר במדריך זה בפרק ז' – עיינו שם כעת!**

הסבר כללי - מה שאנו מחפשים בניתוח שונות דו-כיווני: נרצה לבדוק איך ההצלבה של רמת [ציוני] הנבדקים במשתנה בלתי-תלוי אחד (בדוגמה: ותק) ובמשתנה בלתי-תלוי שני (בדוגמה: משמעות בעבודה) משפיעה על הנתונים [ציונים רציפים] שלהם במשתנה התלוי (שביעות רצון). למשל, את השערה 1 למעלה נוכל לבדוק במודל 2X2 הבא (שימו לב להסבר מתחת):

		ותק			
		נמוך	גבוה		
משמעות בעבודה	נמוכה	1	2	5	
	גבוהה	3	4	6	
	סה"כ	7	8		

בניתוח שונות דו-כיווני, נרצה לבחון קודם כל מה מה קורה ב"שוליים" של הטבלה, בתאים 5-6 (בשורות) וגם 7-8 (בטורים), ובהמשך מה קורה בהצלבות של שני המשתנים, בתאים 1-2-3-4.

בדיקת "אפקטים עיקריים": תחילה, נרצה לבדוק האם לכל משתנה בלתי-תלוי לחוד יש השפעה על המשתנה התלוי, ללא קשר (או מעבר) להשפעה של המשתנה הבלתי-תלוי השני. (במינוח סטטיסטי, זה קרוי "אפקט עיקרי" main effect). למשל: האם יש השפעה ל"משמעות בעבודה" על "שביעות רצון"? כדי לתת תשובה לשאלה זו, נבדוק את תאים 5 ו-6, בהם יופיעו ממוצעי שביעות רצון בשתי הרמות הקיימות במשתנה "משמעות בעבודה" [יש שתי שורות, כי קיבצנו את הנבדקים לכאלו שתופסים את העבודה כבעלת משמעות נמוכה, או גבוהה]. אם יש הבדל מובהק בין שני התאים, נוכל לטעון כי יש אפקט עיקרי ל"משמעות בעבודה" על שביעות הרצון, כלומר ציוני משמעות בעבודה משפיעים על שביעות הרצון, מעבר לרמות של "ותק" [בטורים]. באופן דומה, נרצה לבדוק את האפקט העיקרי של המשתנה הבלתי-תלוי השני, כדי לתת תשובה לשאלה: האם יש השפעה לותק על שביעות רצון? נבדוק את תאים 7 ו-8, בהם יופיעו ממוצעי הנבדקים בשתי הרמות של ותק [טור ותק נמוך, טור ותק גבוה]. אם היינו מוצאים הבדל מובהק בין תאים 7 ל-8, היינו יכולים לטעון כי "ותק" משפיע על שביעות הרצון, מעבר לרמות של משמעות בעבודה.

בדיקת אינטראקציות: בהמשך, נרצה לבדוק האם יש אינטראקציה (interaction) בין המשתנים הבלתי-תלויים, כלומר האם הממוצעים בארבעת התאים 1-2-3-4 שונים זה מזה, באופן שמראה שהשילוב של שתי רמות כלשהן (למשל משמעות גבוהה וותק גבוה) מעלה או מוריד את הממוצע בתא הזה באופן חריג ביחס למה שקורה בתאים האחרים (למשל: משמעות גבוהה אך ותק נמוך).

2. כיצד נפעיל ניתוח שונות דו-כיוונית? נעבוד בשני שלבים: הקבצת המשתנים הבלתי-תלויים (אם הדבר נדרש), ולאחר מכן הרצה של ניתוח השונות:

שלב 1: הקבצת המשתנים הרציפים: אם המשתנים הבלתי-תלויים הם רציפים, נקבץ אותם וניצור משתנים חדשים בעלי מספר קטן של רמות, לרוב שתיים עד ארבע רמות (למשל: נמוך, בינוני, גבוה). דבר זה נדרש כדי להריץ את העיבוד – ויותר חשוב מכך, כדי לפרש את תוצאותיו! **ביצוע הקבצה של משתנים רציפים מוסבר במדריך זה בפרק ז'.** (רצוי תחילה להריץ התפלגות פשוטה (frequency) של כל משתנה רציף, ואז לקבוע את הגבולות של כל קבוצה).

שלב 2: הרצת ניתוח שונות דו-כיוונית (2-way Analysis of Variance / ANOVA).

היכנסו לתפריט Analyze, בחרו General Linear Model, ובתפריט שנפתח: Univariate. (כלומר, נבחר לבדוק מודל המניח קשר ליניארי בין משתנים, ומטרתנו לנבא את הקשר שלהם למשתנה תלוי בודד). בתיבת הדיאלוג שנפתחה, בצעו את הפעולות הבאות:

1. הכניסו את המשתנה התלוי (למשל: שביעות רצון) לחלון למעלה: "Dependent variable"
2. הכניסו את שני המשתנים הבלתי-תלויים (למשל: משמעות, ותק) לחלון: "Fixed factors"
3. לחצו על כפתור Options, ובקשו Descriptive statistics. סטטיסטיקה תיאורית (ממוצעים, סטיות תקן) – הדבר הכרחי כדי לראות את הממוצעים בתאים שנוצרו וכבסיס להבנת התוצאות.
4. לחצו OK להרצה.

3. פירוש התוצאות: כדי לפרש את הפלט, יש לעבוד בשלושה שלבים המוסברים להלן.

שלב 1: הסתכלות בטבלת הממוצעים. למטה משמאל מוצגת דוגמה לטבלה כזו, שנוצרה לגבי השערה 1: המשתנה "ותק" עבר הקבצה לשלוש רמות (לפי מספר חודשי ותק בעבודה), והמשתנה "יתפישת משמעות בעבודה" עבר הקבצה לשתי רמות (נמוך = מתחת לממוצע, גבוה = מעל לממוצע). בטבלה זו נסתכל על הממוצעים שנוצרים מההצלבה 2X3. אך כשתנסו לקרוא את הטבלה, תראו שקשה להבין מה קורה בה - כי SPSS מסדר את כל הממצאים בזה אחר זה!

Descriptive Statistics

Dependent Variable: Sviut

		Mean	Std. Deviation	N
Vetek	Mashma2			
	1=1-6 mo.	4.43	1.212	48
	2=7-36 mo.	5.76	.987	43
	Total	5.06	1.290	91
2=7-36 mo.	1=low	4.14	1.013	99
	2=high	5.53	.763	74
	Total	4.74	1.143	173
3= 37+ months	1=low	3.94	1.303	30
	2=high	5.59	.892	61
	Total	5.04	1.297	91
Total	1=low	4.19	1.128	177
	2=high	5.60	.865	178
	Total	4.90	1.229	355

לכן אני ממליץ לכם להעתיק את הממוצעים לטבלה נפרדת שתיצרו לבד (תבנו אותה בעצמכם באופן ידני ב'וורד') ובה שישה תאים, והיא מוצגת בהמשך. טבלה זו מאורגנת כמו הטבלה שהוצגה בסעיף "רקע" בעמוד הקודם, ושימוש בה יסייע לכם מאוד ויפשט את פירוש הממצאים. **העבירו אליה בהתחלה רק את הממוצעים** – כדי להדגים, שלושת הערכים הרלוונטיים למשמעות "נמוכה" (4.43, 4.14, וכו') סומנו בחיצים בטבלה.

כעת חיזרו להסבר הכללי שהוצג בעמוד הקודם, לגבי מה אנו מחפשים בניתוח דו-כיווני. שאלו עצמכם: האם יש בנתונים שלנו אפקט עיקרי לגבי המשתנה בטורים? בשורות? האם יש אינטראקציות?

סה"כ	משמעות בעבודה		ותק
	נמוכה	גבוהה	
נמוך	4.43	5.76	5.06
בינוני	4.14	5.53	4.74
גבוה	3.94	5.59	5.04
סה"כ	4.19	5.60	

למשל. בהתייחס לטבלה עם הממוצעים מצד שמאל (שנוצרה מהעברת נתונים בפלט SPSS לתוך טבלה שיצרתם לבד): האם אתם יכולים להגיע למסקנות לגבי ההבדל בין שתי הרמות של משמעות? (בידקו את שורת הסה"כ בטורים "נמוכה", "גבוהה")? בין שלוש הרמות של ותק (סה"כ בשורות)? בנוסף, האם יש אינטראקציה (שינוי בגודל היחס של הממוצעים), בשילובים שונים של משמעות+ותק?

שלב 2: בדיקת טבלת התוצאות של ניתוח שונות:

להלן טבלה כזו, שנוצרה לגבי השערה 1.

כזכור, בדקנו את

ההשפעות של שני

משתנים בלתי-תלויים

– "ותק" ו"משמעות",

אשר מאפייניהם כבר

הוסברו למעלה, על

המשתנה התלוי

"שביעות רצון" (Sviut)

אשר שמו רשום בראש

הטבלה.

Tests of Between-Subjects Effects

Dependent Variable: Sviut

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	184.256 ^a	5	36.851	36.735	.000
Intercept	7355.697	1	7355.697	7332.481	.000
Vetek	5.529	2	2.765	2.756	.065
Mashma2	161.463	1	161.463	160.954	.000
Vetek * Mashma2	1.262	2	.631	.629	.534
Error	350.105	349	1.003		
Total	9050.563	355			
Corrected Total	534.361	354			

a. R Squared = .345 (Adjusted R Squared = .335)

בטבלה נבדוק אפקטים עיקריים, ואינטראקציות. להלן הערות לגבי כמה מהם:

- השפעת ותק (Vetek): רואים בטבלה כי נמצא ערך F לא גבוה – ובטור sig. רואים כי רמת המובהקות שלו היא 0.065 – ערך קרוב למדי אך קצת מעל רמת המובהקות הבסיסית 0.05 אליה מתייחסים הרבה מחקרים במדעי החברה וההתנהגות. כדאי כעת לעיין בטבלת הממוצעים בעמוד הקודם – והנתונים שם יבהירו לנו מדוע ערך F אינו גבוה (שימו לב שאין למעשה הבדל בין שתיים מקבוצות הותק, ברמה הנמוכה והגבוהה). בנוסף, הנתונים גם מראים שההשפעה של ותק על שביעות רצון אינה ליניארית (שימו לב לסדר הממוצעים בשלושת השורות - שביעות הרצון אינה עולה בהתמדה מנמוך לגבוה עם העלייה בותק).
- השפעת משמעות בעבודה (Mashma2): למשתנה יש אפקט עיקרי מובהק על שביעות רצון (הסתכלו בשורה של Mashma2 בפלט בראש עמוד זה – רמת המובהקות היא מתחת 0.001. אך בגלל ש-SPSS מציג בטבלאות רק 3 ספרות אחרי הנקודה העשרונית, כתוב בטבלה שהמובהקות היא 0.000]. עיון בממוצעים בעמוד הקודם תומך בממצא זה - אכן יש הבדל די גדול בממוצעים בשני הטורים.
- אינטראקציה (Vetek * Mashma2): ערך F בשורה זו נמוך מ-1, לא מובהק.

שלב 3: הרחבות לגבי פירוש הממצאים ולגבי הצגה וכתובה:

- פירוש הנתונים - תמיד צריך להתחיל מהסתכלות על הממוצעים (ועל מספרי הנבדקים בכל תא).** כדי להבין מה משמעות הממצאים בטבלת ניתוח השונות עצמה (בה מוצגים ערכי F) ולפרש את המידע המוצג ביחס למשתנים הנכללים בבדיקה, תמיד חשוב לבדוק תחילה באילו תאים הממוצעים גבוהים יותר ובאילו נמוכים יותר. הכרחי לשים לב למספרי הנבדקים בכל תא, כי אנו חייבים לדעת על כמה נבדקים מבוסס חישוב הממוצע והפיזור. רצוי שלא יהיו הבדלים גדולים מאוד במספר הנבדקים בתאים השונים, ושכל תא יהיו לפחות 5 נבדקים, רצוי יותר. אם מספר הנבדקים בתאים מסוימים הוא קטן מאוד, העיבוד הסטטיסטי עשוי לייצר תוצאות משובשות והפירוש צריך להיעשות בזהירות.
- הצגת הממצאים וכתובת ההסבר לגבי מה שנמצא בעבודה הסמינריונית:** את הממצאים לגבי ערכי F ורמות המובהקות מציגים יחד עם מידע לגבי הממוצעים, סטיות התקן, ומספרי הנבדקים. לעתים משלבים רק בטקסט, כחלק מהמלל (ראו דוגמא בעמוד הבא), או שמשלבים בין הצגת ממוצעים וערכי F בטבלה, והסבר במלל מלווה. **בכל מקרה, לא מכניסים פלטים של SPSS כמות שהם לתוך הטקסט!** את המידע הרלוונטי לנו מציגים רק בטבלאות שהחוקרים בונים באופן ידני – בהן כוללים רק את הנתונים הנדרשים, ובמבנה שמסייע לנו להסביר לקורא העבודה מה מצאנו. הטבלה הרצויה היא במבנה דומה לדוגמה הבסיסית שמוצגת בתחתית העמוד הקודם – אך תמיד נוסיף בכל תא שלושה מרכיבים: ממוצעים, סטיות תקן ומספרי נבדקים (N).

כדי לבדוק את ההשערה הרביעית, המנבאת הברלים ברמת הסמכותנות הטיפולית בין תומכי מפלגות הימין לתומכי מפלגות השמאל, נערך ניתוח שונות חד-כיווני של ציוני הסמכותנות הטיפולית לפי דפוסי ההצבעה של הנבדקים. בלוח 2 מוצגות תוצאות ניתוח זה.

לוח 2: ממוצעים וסטיית-חקן של הסמכותנות הטיפולית, לפי דפוס הצבעה

דפוס הצבעה	n	ממוצע	סטיית-חקן
מפלגות השמאל ¹	370	3.21	0.65
מפלגת העבודה	116	3.48	0.61
מפלגות הימין ²	64	3.70	0.67
דתיים ³	35	3.38	0.79
סה"כ ⁴	585	3.33	0.67

ניתוח השונות הצביע על הכדלים מובהקים סטטיסטית ברמת הסמכותנות הטיפולית בין תומכי המפלגות השונות, $F(3, 581)=12.98, p<.0001$. רמת הסמכותנות הטיפולית הגבוהה ביותר נמצאה בקרב תומכי מפלגות הימין, והנמוכה ביותר נמצאה בקרב תומכי מפלגות השמאל. תוצאות מבחן Scheffé, המתאים להשוואת קבוצות שונות בגודלן (רי (Keppel, 1973)), מצביעות על כך שההכדלים המובהקים סטטיסטית נובעים מכך שממוצע הסמכותנות הטיפולית של תומכי הימין גבוה באופן מובהק סטטיסטית מזה של תומכי מפלגות השמאל ומה של תומכי המפלגות הדתיות (אך לא מזה של תומכי מפלגת

דוגמה להצגה וכתובה של תוצאות ניתוח שונות דו-כיווני.

שימו לב כי בדו"ח המחקר תמיד יוצגו כמה מרכיבים: תחילה מה רצינו לבדוק ואיך הדבר נבדק (בדוגמה זו, בטקסט לפני הטבלה), בהמשך יוצגו הנתונים בכל התאים (ממוצעים, סטיות תקן, מספר נבדקים – אלו מופיעים כאן בלוח 2). אחרי כן יוצגו תוצאות ניתוח השונות (בדוגמה זו, בטקסט שמלווה את הטבלה).

מחשבות לסיום

מדריך זה נכתב בעיקר כדי לסייע בהתלבטויות אופייניות לשלבים טיפוסיים בעיבוד בסיסי של נתונים האופייניים לסמינר מחקר. לפיכך, המדריך אינו מנסה לכסות את כל סוגי העיבודים הקיימים, אלא מניח שתלמידים יבחרו עיבודים המתוארים במדריך או יוסיפו ניתוחים מתקדמים (שאינם כלולים במדריך) בהתאם למאפייני הנתונים ושאלות והשערות המחקר שלהם, תוך העזרות בחומר שלמדו בקורס לסטטיסטיקה, ובייעוץ מהמרצה. לפיכך, יש מקום נרחב לשיקול דעת והפעלת חשיבה יצירתית!

בין השאר, אדגיש שוב את החשיבות שבביצוע העיבודים המתוארים במדריך זה לא רק במדגם המחקר כולו אלא גם בתת קבוצות, גם אם אין לכם השערות ספציפיות בנושא. זה הכרחי כדי לבדוק אם ממצאים שקיימים במחקר כולו אכן יציבים ומופיעים גם בתוך תת-קבוצות בהן יש לכם עניין או שידוע מהספרות כי הם עשויים להשפיע על ממצאי המחקר (למשל בקרב גברים ונשים, או אנשים בגילים שונים, וכו'). לעתים מגלים דפוסים מעניינים ובלתי צפויים בדרך זו, והדבר מאפשר להעמיק את תרומת המחקר לידע ולהבנה, ומסייע להבנת המידה בה ממצאי המחקר שלכם ניתנים להכללה לאוכלוסיות נוספות.

סמינר מחקר הוא תהליך רב-שלבי שמטרתו לפתח ולחזק הבנה ומיומנויות בתחומים רבים. לכן, חשוב לא רק לבצע עיבודים סטטיסטיים מתאימים, אלא גם לפרש נכון את התוצאות, ולכתוב בצורה תמציתית וברורה את הממצאים ואת המסקנות הנובעות מהם. כאשר תכתבו את פרקי התוצאות והדיון, שאלו את עצמכם אם הכתיבה שלכם היא תמציתית, מאורגנת וברורה לקורא העבודה. הסתכלות ביקורתית על מה שכתבתם תסייע לכם לשפר את בהירות דו"ח המחקר שתפיקו. בכל מקרה, היעזרו בדוגמאות שמספק המרצה שלכם לגבי תכנון, ניסוח וארגון פרקי עבודת המחקר, ואמצו ניסוחים שנראים לכם מתאימים לדו"ח שלכם.

אני מקווה כי מדריך זה יסייע לכם בעיבודים ובעבודת המחקר. אנא העבירו הערות והצעות לשינויים בדוא"ל <iddo@research.haifa.ac.il>, כדי שאוכל להמשיך ולעדכן את המדריך.

16 138 9"9

נספח: טיפול בקבצים, העתקת נתונים, ויצירת טבלאות ב- Word

1. ביצוע שינויים בקובץ הנתונים:

לאחר קבלת הקובץ המאוחד ממורה הסמינר או יצירת קובץ צוותי, תוכלו לשנות את שמות המשתנים בהתאם לצרכים שלכם, או לנפות משתנים שאינם נדרשים.

ניתן במצבים מסוימים גם לנפות (למחוק) נבדקים (כלומר להקטין את המדגם), למשל אם נבדקים בעלי מאפיינים מסוימים אינם מתאימים למסגרת המחקר שלכם. אולם, אם החלטתם לעשות זאת, תצטרכו להסביר בפרק ה"תוצאות" מה עשיתם, ומדוע, ולדון במשמעויות של פעולותיכם מבחינת יכולת ההכללה מהמחקר או השלכותיהן על המסקנות.

2. איך שומרים (מאחסנים) פלטי SPSS?

חיוני לשמור את הפלטים העיקריים שיצרתם, כדי שתוכלו להכין את דו"ח המחקר ולא תצטרכו להריץ שוב דברים בהמשך. בחרו באחת מהחלופות הבאות:

1. שמירת כל הפלט: ניתן לבצע "save as" של חלון הפלט. SPSS שומר אותו בקובץ עם סיומת .spo. תוכלו לפתוח קובץ כזה בהזדמנות אחרת כדי להסתכל עליו שוב ולשלוף נתונים מתוכו - אולם תוכלו לעשות זאת מתוך SPSS בלבד, לא מתוך תוכנה אחרת.
2. העתקת טקסט לתוך WORD: ניתן להעתיק חלקים נבחרים של פלט SPSS למסמך WORD ע"י סימון (בחירה והשחרה) בחלון הפלט, העתקה ע"י פקודות copy או copy objects ב-SPSS, והדבקה בתוך WORD עם paste או paste special. בדרך כלל, העתקה עם "copy objects" והדבקה עם "paste" יעתיקו את החלק שסימנתם בחלון הפלט כ"תמונה" אשר לא ניתן לערוך אך אפשר להגדיל ולהקטין כמו כל אובייקט גרפי. שימוש ב"paste special" יעתיק את הפלט כטקסט (לרוב כטבלה בפורמט RTF) שנראה פחות אלגנטי אך ניתן לערוך ולשכתב אותו (וגם למחוק מתוכו חלקים). עשו ניסיונות למצוא את הדרך הטובה ביותר כאשר אתם משלבים את הפקודות השונות. (יתכן ויהיו הבדלים בהתאם למהדורות התוכנות שלכם).

3. על שילוב פלטי SPSS בעבודת מחקר:

לידיעתכם: אין להעתיק חלקים של פלטי SPSS לתוך עבודת מחקר, זאת כי פלט גולמי כולל מרכיבים מיותרים ופרטים שלעולם לא מוצגים בטבלאות הסופיות!! אם יש בתוך טבלאות או פלטי SPSS חומר שאתם מעוניינים לשלב בעבודת המחקר שלכם, הדרך הנכונה היא ליצור בתוך מסמך Word טבלאות ריקות אשר ערוכות בדיוק בצורה הרצויה לכם, ואליהן הקלידו את הערכים הרלוונטיים מתוך הפלט. אם ניתן, רצוי לרכז נתונים בעלי אופי דומה בתוך אותה טבלה במקום ליצור טבלאות חסכוניות וקטנות, למשל, לרכז בטבלה אחת ממוצעים וסטיות תקן לגבי מספר משתנים, או מספר מתאמים הקשורים לכמה השערות, או תוצאות של כמה מבחני t.