**דו"ח תהליך העבודה בפרויקט**

**שלב העיבוד המקדים וניתוח הנתונים**

**גיא חן**

**תוכן עניינים**

[עיבוד מקדים 3](#_Toc170734887)

[ניתוח נתונים – תיאוריה ובחירת המבחנים 4](#_Toc170734888)

[בחירת המבחנים המתאימים לניתוח 4](#_Toc170734889)

[תיקון אלפא 10](#_Toc170734890)

[מדד גודל אפקט 11](#_Toc170734891)

[אופטימיזציה של מבחני ההמשך לפני ניתוח 12](#_Toc170734892)

[Welch's t-test – Test Formulas 14](#_Toc170734893)

[Analysis Of Variance – Test Formulas 15](#_Toc170734894)

[ניתוח נתונים – תוצאות 17](#_Toc170734895)

[מוחצנות - Extraversion 17](#_Toc170734896)

[יציבות רגשית - Emotional Stability 18](#_Toc170734897)

[נעימות - Agreeableness 19](#_Toc170734898)

[מצפוניות - Conscientiousness 20](#_Toc170734899)

[פתיחות לחוויות – Openness to experience 21](#_Toc170734900)

[מקיאווליזם - Machiavellism 22](#_Toc170734901)

[נרקיסיזם - Narcissism 23](#_Toc170734902)

[פסיכופתיה - Psychopathy 24](#_Toc170734903)

[הערכה עצמית – Self Esteem 25](#_Toc170734904)

[הסקת מסקנות 26](#_Toc170734905)

[שאלת המחקר 26](#_Toc170734906)

[ממצאים נוספים 29](#_Toc170734907)

[מגבלות המחקר 31](#_Toc170734908)

[נספחים 34](#_Toc170734909)

[סולם ליקרט 34](#_Toc170734910)

[ציינון השאלונים 35](#_Toc170734911)

[בילביוגרפיה 37](#_Toc170734912)

**שלב שני ושלישי: עיבוד מקדים וניתוח הנתונים (עבודה ב-R)**

# עיבוד מקדים

בשלב הזה מרבית העבודה מתועדת דרך קובץ RMARKDOWN ולכן נשתמש בדו"ח בשביל להסביר ולתעד את רציונל העבודה שלנו ולצרף נספחים שחשבנו שמקומם אינו בקובץ. אם כן, נסביר את אופן העבודה בשלב העיבוד המקדים (הנספחים יצורפו בסוף המסמך).

למעשה, בשלב העיבוד המקדים הבנו כי אי אפשר לעבוד על העיבוד הראשוני באופן שהוא בלתי תלוי בניתוח הנתונים המאוחר יותר, ולכן יש אלמנט מקבילי לעבודה. הסבר:

בתחילה, רצינו להתחיל את העיבוד בכך שנסנן את המדינות שלא בחרנו בהן למחקר, מתוך מחשבה על יעילות – לא נרצה לבצע מניפולציות בדאטה (כמו ציינון שאלונים) עבור מדינות שלא נשתמש בהן. לאחר עוד מחשבה על הנושא, הגענו למסקנה כי בכל זאת כדאי להשאיר את המדינות הללו, מתוך מחשבה על שלב ניתוח הנתונים שעתיד להגיע מאוחר יותר. לאחר מחשבה על הניתוח ובדיקה במקורות ידע, הגענו למסקנה כי נרצה לבצע ניתוחי שונות חד גורמיים על הדאטה. ניתוחי שונות חד גורמיים מניחים כי מתקיים שיוויון שונויות בין האוכלוסיות השונות במחקר. במחקר שלנו המשמעות היא שאם ניקח למשל את המשתנה "הערכה עצמית", אז השונות של המשתנה בישראל צריכה להיות שווה לשונות של המשתנה בבריטניה, וכן הלאה. אם כן, יכולה להיות הפרה של ההנחה הזאת.  
הרציונאל שלנו היה שאם תתקיים הפרה בוטה של הנחת שיוויון השונויות במדינות שבחרנו לפרויקט, ועדיין נרצה לבצע ניתוח שונות סטנדרטי, יהיה לנו קל לבחון האם ההנחה מתקיימת עבור מדינות אחרות אם נבצע את עיבוד הנתונים (בפרט ציינון שאלונים) על כל המדינות שעומדות לרשותינו מלכתחילה. לכן, בחרנו שלא לסנן את המדינות ה"לא רלוונטיות" בשלב הראשוני מתוך הנחה שאולי הן כן יהיו רלוונטיות במקרה של הפרה של הנחות המבחנים שאנחנו רוצים לקיים – ואז אולי נרצה לשנות את המדינות בניתוח כדי שיעמדו בהנחות.

החלטנו להתחיל לעבד שאלון פשוט – שאלון ההערכה העצמית של רוזנברג (ראו אופן ציינון [כאן](#ציינון)). כשסיימנו את העיבוד, החלטנו לבדוק האם דאטה ההערכה העצמית עומד בהנחת שיוויון השונויות כבר בשלב הזה. התלבטנו בין שני מבחנים שבודקים את ההנחה:

1. Levene's Test – המבחן הנפוץ ביותר לבחינת שיוויון שונויות. היתרון שלו הוא שהוא פחות רגיש לסטיות מנורמליות של המשתנה התלוי בקבוצות המחקר לעומת מבחנים אחרים.
2. Bartlett's Test – עוד מבחן לבדיקת שיוויון שונויות. הוא מניח שהמשתנה התלוי בכל אחת מהקבוצות מתפלג נורמלי, והוא רגיש להפרות של ההנחה הזאת.

ההסתמכות שלנו במחקר היא על משפט הגבול המרכזי, והמשתנים עצמם במקור כן סוטים מנורמליות על סמך ויזואליזציות שעשינו. בגלל שאנחנו לא מתיימרים לכך שהמשתנים עצמם נורמליים, הלכנו על מבחן לווין שהוא גם הנפוץ ביותר.

לצערנו, דחינו את השערת האפס במבחן לווין – מה שהוביל אותנו למסקנה כי השונויות שונות בין אוכלוסיות הניסוי. כמה הערות:

1. אמנם דחינו את השערת האפס, אבל השימוש שלנו במדגמים גדולים במחקר תורם מאוד כדי לדחות את השערת האפס, גם אם ההבדלים בין השונויות אינם גדולים (כלומר גודל אפקט קטן). [2]
2. מתוך ההבנה כי ייתכן וההבדל בשונויות אינו משמעותי מספיק בשביל לפסול שימוש במבחן ניתוח שונות, החלטנו לשמור את כל סטיות התקן של המדינות בשלב הזה. בתור התחלה, לא נראה שיש הבדלים דרסטיים בסטיות התקן, אבל נצטרך להסתכל על כך בצורה מעמיקה יותר בהמשך כדי להגיע להכרעה.

אם כן - אחרי שנבצע את העיבוד עבור כל משתני האישיות שלנו, ונבדוק את הנחת שיוויון השונויות על כולם, נחליט איך להתקדם בשלב ניתוח הנתונים.  
אם נכריע כי ההבדלים בשונויות אינם משמעותיים, נבצע מבחני ניתוח שונות. אם נזהה מדינות בעייתיות ספציפיות בשונות שלהן, נשקול להחליף אותן במדינות אחרות ובכך לפתור את הבעיה. לבסוף, אם לא יהיה מנוס, נשתמש במבחן סטטיסטטי אחר שלא מניח שיוויון שונויות.

עיבוד השאלונים האחרים אינו הצריך שימוש נוסף במסמך זה מעבר לאופן ציינון השאלונים – ראו אופני ציינון לשאלונים השונים [כאן](#ציינון).

הערה לגבי השימוש בפונקציות צפיפות למשתנים התלויים – השתמשנו בפונקציות צפיפות עבור משתנים שהם אינם רציפים (אבל כן עם הרבה ערכים). הרציונל לכך היה שחשבנו שהדאטה הרבה יותר נוח להצגה והבנה באמצעות פונקציות צפיפות - והמטרה שלנו לא הייתה להשיג דיוק מושלם באמצעות הויזואליזציות, אלא רק לקבל איזשהו רושם ראשוני בסיסי על הדאטה. בנוסף, לא הסתפקנו רק בפונקציות הצפיפות והצגנו גם boxplots שאפשרו לקבל תובנות נוספות על הדאטה. האלטרנטיבה המרכזית הייתה שימוש בהיסטוגרמות, ובחרנו שלא להשתמש בהן כי חשבנו שהן פחות "נקיות" ואסתטיות, ומסורבלות לעבודה לעומת פונקציות הצפיפות. הסרבול בא לידי ביטוי בכך שיש צורך לשנות את מספר הbins עבור כל התפלגות כדי שהויזואליציה תהיה ברורה להבנה – חשבנו שעדיף לא לרדת לרזולוציות האלה למטרה שלנו, ולכן בחרנו את הרדוקציה של פונקציית צפיפות.

# ניתוח נתונים – תיאוריה ובחירת המבחנים

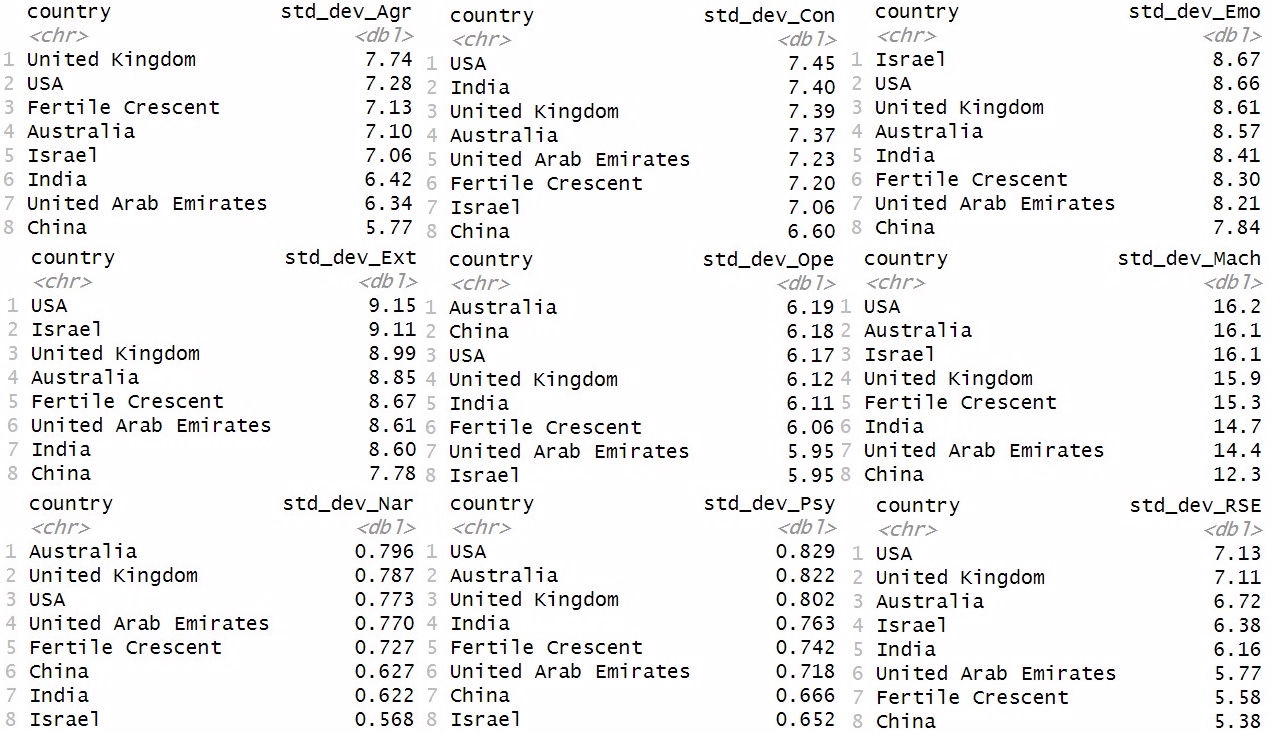
## בחירת המבחנים המתאימים לניתוח

ראשית, נציג את המשתנים שאספנו לניתוח:

1. משתנה בלתי תלוי – מדינה. בעל 8 רמות: ישראל, ארצות הברית, אנגליה, אוסטרליה, איחוד האמירויות הערביות, הסהר הפורה (אזור), סין והודו.
2. 9 משתנים תלויים – מוחצנות, יציבות רגשית, נעימות, מצפוניות, פתיחות לחוויה, הערכה עצמית, מקיאווליזם, נרקיסיזם ופסיכופתיה.

השאלה הראשונה שאולי מתבקשת לשאול: האם יש אפשרות לבצע מבחן יחיד שמסוגל לתמוך במשתנים תלויים רבים ומשתנה בלתי תלוי יחיד? התשובה היא שקיים מבחן כזה – הוא נקרא Manova (Multivariate analysis of variance).   
המבחן מהווה טכניקה אנליטית שמרחיבה את העקרונות של ANOVA על מנת לבחון הבדלים על פני מספר משתנים תלויים במקביל. בניגוד לאנובה שמתמקדת בהשתנות של משתנה תלוי אחד על פני משתנים בלתי תלויים, מאנובה לוקחת כמה משתנים תלויים יחדיו. מאנובה מבצעת אינטגרציה של המשתנים התלויים באמצעות קומבינציה ליניארית, והופכת אותם למשתנה תרכובת יחיד. לאחר מכן היא בודקת את ההשתנות של אותו משתנה על פני המשתנים הבלתי תלויים. [8]  
יש מספר סיבות ללמה השיטה הזו לא מתאימה עבור מחקרנו. הראשונה שבהן היא הפשוטה והבסיסית ביותר - אופן איסוף הדאטה שלנו לא מתאים לשיטה הזו. אספנו מידע על משתנים תלויים שונים דרך מדגמים שונים, לכן אין ביכולתנו לבנות וקטור רב מימדי לכל נבדק שמכיל בתוכו את 9 המשתנים התלויים במחקר בשביל הניתוח. אם היינו רוצים להשתמש במבחן הזה בצורה חלקית, למשל בשביל חמש התכונות הגדולות (שכן נמדדו על אותם הנבדקים), אז יכולנו טכנית לעשות זאת, אבל אז לא יכולנו להסתפק רק במבחן יחיד בכל מקרה (כי היו נשארות לנו עוד 4 תכונות). מעבר לסיבה הטכנית, יש צורך לקחת בחשבון את הרדוקציה הכלולה במבחן הזה – אם נהפוך את המשתנים התלויים למשתנה אחד מאוחד, לא נוכל להבין את ההשתנות של כל משתנה תלוי בנפרד על פני המשתנה הבלתי תלוי שלנו (מדינה), וזה מידע חשוב ומעניין שאנחנו לא רוצים לאבד. סיבה נוספת וחשובה היא הנחות המבחן. המבחן דורש הנחות משמעותיות, וכבר בשלב העיבוד המקדים ראינו כי הדאטה שלנו מפר אפילו את הנחת שיוויון השונויות של ANOVA בסיסית. לסיכום, לאור הסיבות שציינו, פסלנו את השימוש במבחן מאנובה.

אם כן, לאחר שפסלנו שימוש במאנובה, הגענו למסקנה כי נצטרך לבצע 9 מבחנים – אחד עבור כל משתנה תלוי במחקר. מלכתחילה, השאיפה שלנו הייתה לבצע מבחנים פרמטריים, ולכן דאגנו לבחור שאלונים שיתאימו להנחות המבחנים הפרמטריים עוד בשלב הראשון של הפרויקט (למשל מדגמים מספיק גדולים על מנת להניח נורמליות של התפלגות הדגימה). הסיבה לכך הייתה שמבחנים א-פרמטריים מבצעים רדוקציה של הדאטה – הדאטה מאבד את ערכו הכמותי,

והעוצמה (היכולת לגלות אפקטים קיימים) של המבחן נפגעת לעומת המבחנים הפרמטריים. [9]  
כשלוקחים בחשבון את מספר הרמות של המשתנה הבלתי תלוי במחקר (גדול מ2), הניתוח הראשון שניתן לחשוב עליו הוא ניתוח שונות (ANOVA). זהו אכן הניתוח ששאפנו לבצע במחקר. למרות זאת, בשלב הניתוח המקדים ראינו כי הנחת שיוויון השונויות הדרושה לביצוע אנובה, נדחית עבור כל משתני המחקר. למרות שגדלי המדגם הגדולים השפיעו מאוד על מבחני הלווין שביצענו, הבנו כי אין להוזיל מהמשמעות של ההפרה הזו, ושמרנו את סטיות התקן המדגמיות כדי לבחון את גודל ההפרות. נציג אותן כאן:

תזכורת לנוסחת סטיית התקן המדגמית:

*\*יש לשים לב כי סטיית התקן המדגמית שונה במקצת מסטיית התקן באוכלוסייה – מחלקים ב ולא ב הסיבה לכך היא שאם מחשבים שונות מדגמית על בסיס חלוקה ב – מקבלים אומד מוטה כלפי מטה ביחס לפרמטר (במקרה הזה שונות באוכלוסייה). החלוקה ב מגדילה את האומד ויוצרת אומד לא מוטה לשונות באוכלוסייה. למרות זאת, חשוב לציין כי סטיית התקן המדגמית עצמה היא עדיין מוטה, אבל פחות.*

ראשית, חשוב לציין כי חמש התכונות הגדולות הן כולן בסקאלה של 10-50, ולכן גדלי סטיות התקן הם בערך באותו סדר גודל עבורן. נשים לב כי סין באופן עקבי (חוץ מבפתיחות) מפגינה שונות קטנה יותר משאר המדינות, ולכן היא אמורה להשפיע מאוד על ההפרה של הנחת שיוויון השונויות. הבעייתיות של סין באה לידי ביטוי גם בשאר התכונות, והיא בולטת במיוחד למשל במשתנה מקיאווליזם. לאור בעייתיות זו, שקלנו להחליף את סין במדינה אחרת כדי שיהיה סביר יותר להניח שיוויון שונויות. למרות זאת, החלטנו שהפתרון הנכון במקרה הזה הוא כנראה לא לבצע חילוף שכזה. ההחלטה נבעה מכמה סיבות:

1. חשבנו כי סין היא מדינה משמעותית בניתוח שלנו. סין היא אולי המדינה הבולטת ביותר כשמדברים על העיצוב התרבותי של המזרח הרחוק (ביחד עם הודו), ולכן חשבנו שהנוכחות שלה בתור מייצגת של התרבות היא חשובה.
2. הבנו כי גם אם נבצע את החילוף, יש סיכוי קלוש שנוכל להניח שיוויון שונויות מבלי להתעלם מתוצאות מבחן לווין. המדגמים שלנו גדולים, וקיימים פערים בשונויות גם כאשר לא לוקחים את סין בחשבון.

בתחילה, ניסינו לחשוב על דרכים יצירתיות להראות כי ההבדלים בשונויות אינם משמעותיים למרות המובהקות של מבחן לווין – דרך אחת שחשבנו עליה היא למשל לבצע ממוצע של סטיות התקן ולבחון כמה כל סטיית תקן סוטה מאותו ממוצע ברמת האחוזים, באמצעות חישוב של std\_i/Std\_mean \* 100 עבור כל סטיית תקן. חשבנו שלאחר מכן נוכל לקחת את "סטיות התקן" המתוקננות הללו, ולחשב להן סטיית תקן על מנת לקבל את הסטייה הממוצעת באחוזים, ולהגדיר איזשהו threshold למתי הסטייה משמעותית ומתי היא לא.

ככל שחקרנו את הנושא יותר, הגענו למסקנה כי אולי אנחנו מתעקשים להניח הנחה שאין בסיס אמיתי להניח אותה. ברמה התיאורתית – לא כל כך סביר להניח שאין הבדלים בשונויות של מדינות שונות כאשר מודדים את תכונות האופי של אוכלוסייתן. אם מדינה מסוימת היא יותר הטרוגנית ממדינה אחרת ומאגדת בתוכה הרבה סוגים של אוכלוסיות שונות ומגוונות (גיוון אתני/ גיוון דתי וכו') אז גם הגיוני להניח שהיא תפגין טווח רחב יותר של מדידות בתכונות האופי מאשר מדינה עם אוכלוסייה שהיא הרבה יותר הומוגנית. באופן כללי, המערב ידוע בכך שהוא שואף לקלוט למדינותיו אוכלוסיות מגוונות על פי עקרונות של צדק, חופש וחירות ללא הבדלי דת, גזע או מין. אם כן, גם ברמה התיאורתית, יש הצדקה לחשוב שמדינות כאלו יפגינו יותר שונות מאשר מדינות אחרות.

בנוסף לבעייתיות של הנחת שיוויון השונויות, קיימת בעיה נוספת בבחירה של מבחן אנובה. מקורות ידע רבים שמים את האצבע על כך שהפרה של הנחת שיוויון השונויות אינה הורסת את האמינות של אנובה בפני עצמה. התסריט הגרוע ביותר להרצת אנובה קורה כאשר ההנחה מופרת במקביל לכך שגדלי המדגמים במחקר אינם שווים. [10] [11]  
במחקר שלנו יש הבדלים משמעותיים מאוד בגדלי המדגמים, ולכן הבנו כי בשלב הזה עדיף לחפש מבחן אחר, ולא לבצע אנובה סטנדרטית (עלתה גם האפשרות לצמצם מדגמים מסוימים על מנת שהמדגמים יהיו שווים, אבל הפערים בגדלי המדגמים גדולים מדי ו"לזרוק" הרבה דאטה זה פתרון בעייתי).

לאחר מחקר נוסף, הגענו למסקנה כי המבחן Welch's ANOVA הוא כנראה המבחן הראוי ביותר לבצע על הנתונים שלנו. המבחן הוא למעשה אלטרנטיבה ישירה למבחן האנובה הרגיל, והוא לא מניח שוויון שונויות. [12]  
מחקרים הראו שמבחן Welch הוא המבחן הטוב ביותר במקרה שבו הדאטה הטרוגני (שונויות שונות), נורמלי, ומאוזן/לא מאוזן (גם כאשר המדגמים שווים וגם כאשר הם לא). [13]

כעת כשבחרנו את ניתוח האנובה של וולץ', נרצה לבחור את סוג ניתוח ההמשך שנבצע עבור כל מבחן מובהק. המובהקות של מבחן האנובה מוסרת לנו כי יש הבדל מובהק בין קבוצות הניסוי שלנו – אבל היא לא אומרת לנו מאיפה ההבדלים הללו נובעים. חשוב להבין כי במחקר שבו יש 8 רמות במשתנה הבלתי תלוי, והמדגמים גדולים מאוד, המובהקות של האנובה בפני עצמה היא לא מאוד אינפורמטיבית.   
על מנת לבחור באופן הטוב ביותר מבחן המשך לניתוח השונות, אנו צריכים להבין קודם כל אילו השוואות מעניינות אותנו. יש כמות עצומה של השוואות שניתן לבחון – ואנחנו החלטנו להתמקד בהשוואות זוגיות בלבד (רמה אל מול רמה, בלי לאגד רמות ל"גושים"). גם כאשר מתמקדים רק בהשוואות הזוגיות, קיימות בעצם השוואות זוגיות שונות – כלומר 28 השוואות. אנחנו עורכים 9 מבחני אנובה, מה שאומר שאם היינו בוחנים את כל ההשוואות הזוגיות, היינו מגיעים ל-252 (!) השוואות זוגיות. אנחנו לא צריכים את כל ההשוואות הללו, ואם היינו מתעקשים לבצע את כולן, היינו צריכים לבצע התאמה משמעותית לאלפא שלנו, שהייתה פוגעת בעוצמת המבחנים.   
במחקר שלנו, אנחנו מתעניינים בישראל לעומת שאר תרבויות העולם – ולכן נתמקד רק בהשוואות של ישראל. אם כן, אנחנו נרצה לבצע **7 השוואות זוגיות** עבור כל מבחן אנובה שנבצע.   
ניתן לחשוב כי הגיוני גם לבצע השוואות שהן לא זוגיות – כל עוד ישראל נמצאת בהן. למשל, ישראל אל מול כל מדינות המערב ביחד. הסיבה שאנחנו לא עושים זאת היא כי אנחנו רוצים לבצע כמה שפחות השוואות. אנו חושבים כי ההשוואות המשמעותיות ביותר הן ההשוואות של ישראל מול מדינה אחרת, והן מאפשרות הסקנת מסקנות גם על הרמה הגבוהה יותר של התרבות. שאלה אפשרית שניתן להעלות היא למה לא לבצע השוואות בין התרבויות לישראל, אם הרצון שלנו הוא לעשות מינימום השוואות? התשובה לכך היא שזה יגרום לאיבוד הספציפיות של הדאטה ורדוקציה של המסקנות. מעבר לתרבות העל שהכי מתאימה לפרופיל הישראלי, מעניין אותנו לדעת אילו מדינות מייצגות את הפרופיל הישראלי באופן הטוב ביותר. דרך נוספת לחשוב על זה היא שכל תרבות על מחולקת לתתי תרבויות, והשימוש במדינות מאפשר לסווג את הפרופיל הישראלי לתת התרבות המתאימה לו ביותר. חשוב לציין, אם המטרה שלנו הייתה לבחון תרבויות על בלבד – החלוקה למדינות היא מלכתחילה הייתה מיותרת למחקר.

אם כן, מבחן המשך שכשיר למטרתנו הוא Games-Howell Test. מדובר במבחן שאינו מניח שיוויון שונויות, והוא נועד לביצוע השוואות זוגיות בין הרמות השונות של המשתנה הבלתי תלוי. [14]   
בנוסף, המבחן לוקח בחשבון את מספר ההשוואות שנעשות ומקשיח את הדרישות לקבלת מובהקות בהתאם – כך שנוכל להימנע מהאיום של ניפוח אלפא.

הערה- אנחנו לא יכולים להשתמש בהשוואות מתוכננות (planned comparisons) עבור הזוגות שמעניינים אותנו, מכיוון שהן מניחות שוויון שונויות והן מבוצעות אחרי אנובה סטנדרטי (שגם הוא מניח שוויון שונויות).

## תיקון אלפא

המחקר שלנו מבצע מבחנים סטטיסטיים רבים, ולכן יש צורך לדון באיום של ניפוח אלפא. באופן ספציפי, אנחנו מבצעים 9 מבחני אנובה במינימום (אם כולם אינם מובהקים ואז אין מבחני המשך). המשמעות היא שהסיכוי שלנו לבצע טעות מסוג ראשון (קביעה כי יש אפקט גם כאשר אין אפקט) עולה. הסטנדרט המחקרי לאלפא הוא 0.05 [15], מה שאומר שבבדיקת השערות סטנדרטית יש סיכוי של חמישה אחוזים לבצע טעות מסוג ראשון – חמישה אחוזים שהתוצאות שקיבלנו הינן מקריות ולמעשה לא קיים אפקט. אם נבצע 9 מבחנים עם אלפא של חמישה אחוזים, זה אומר שבכל מבחן יש סיכוי של חמישה אחוזים שנמצא אפקט "בטעות", ונחזור על כך 9 פעמים - זוהי סיטואציה שאנחנו רוצים להימנע ממנה. באופן ספציפי, אפשר לחשב את הסיכוי לטעות מסוג ראשון עם הנוסחא הבאה:  


כאשר x הוא מספר המבחנים. נסביר – בכל מבחן יש לנו סיכוי של 1-a לא לבצע טעות מסוג ראשון, ואנחנו מבצעים x מבחנים. כאשר אנחנו מעלים את 1-a בחזקת x, אנחנו מקבלים את הסיכוי לא לבצע טעות מסוג ראשון על פני x מבחנים. כאשר נחסיר מ1 את הסיכוי הזה, נקבל את הסיכוי לבצע טעות מסוג ראשון לפחות במבחן אחד.

עבור a=0.05, x=9: הסיכוי לבצע לפחות טעות מסוג ראשון אחת הינו 0.37 אם לא נטפל בניפוח האלפא.

יש לציין עוד נקודה חשובה אחת – רשמנו לעיל כי נבצע במינימום 9 מבחני ניתוח שונות, זה מכיוון שאנחנו לא יודעים מראש כמה מהם יצאו מובהקים. בכל פעם שמבחן שונות יצא מובהק, אנחנו נעשה מבחן גיימס-הוול. המבחן אמנם מבצע תיקונים להשוואות מרובות, אבל גם אם נתייחס אליו בתור מבחן שעונה על השערה סטטיסטית אחת בכל פעם שנשתמש בו, הוא עדיין מנפח את האלפא (פשוט משמעותית פחות). אם כל האנובות יצאו מובהקות – נבצע אותו 9 פעמים, שזה לא מעט.

לאור האמור לעיל, בחרנו את תיקון בונפרוני על מנת לפתור את איום ניפוח האלפא. תיקון בונפרוני הוא תיקון נפוץ ופשוט להבנה: אם אני רוצה לשמור על אלפא מסוימת על פני x מבחנים, אני אחלק את האלפא ב-x, וזוהי תהיה האלפא החדשה שלי. [16]  
תיקון בונפרוני נחשב לתיקון מאוד שמרני ומחמיר, ולכן חשבנו שאם נשתמש בו נוכל להסתפק "רק" בחלוקה של האלפא שלנו ב-9, שזה בעצם מספר המבחנים המינימאלי שנעשה במחקר (אם כי זה מאוד לא סביר). הרעיון הוא שגם אם כמעט בוודאות נבצע יותר מבחנים, תיקון בונפרוני הוא מספיק מחמיר מלכתחילה כדי שנוכל לחמוק מהאיום של ניפוח אלפא כאשר נחלק את האלפא "רק" ב-9.

**לסיכום: האלפא החדשה במחקר שלנו היא a=0.05/9=0.005555555.**

## מדד גודל אפקט

הדבר האחרון שנשאר לנו לקבוע הוא מדד גודל אפקט. המטרה של מדד גודל אפקט הוא לומר לנו עד כמה ההבדלים בין קבוצות משמעותיים - מעבר לכך שהם קיימים (חשוב במיוחד כאשר בודקים השערות על מדגמים גדולים).

אנחנו נשאף להשתמש במדד גודל אפקט נפוץ שנקרא Hedge's g. המדד משווה בין 2 קבוצות, והוא מאוד דומה למדד הידוע Cohen's d – אבל הוא נחשב כטוב יותר.

*נביט במדד g - המונה מציין את ההבדלים בין הממוצעים של שתי הקבוצות, והמכנה מציין את סטיית התקן המשוכללת, הנוצרת מממוצע משוכלל של סטיות התקן של שתי הקבוצות. באופן כללי, יש מעט הבדלים בספרות בנוגע לנוסחא המדויקת, אבל העיקרון זהה – המדד מנסה לתת ציון לגודל האפקט בכך שהוא מתייחס להפרש בין הממוצעים ביחידות של סטיות תקן*.  
דרך אינטואיטיבית להסתכל על הנוסחא: מה ההפרש בין האדם הממוצע בקבוצה 1 לבין האדם הממוצע בקבוצה 2 ביחידות של סטיות תקן?

כלל אצבע לפרשנות המדד: small effect = 0.2, medium effect = 0.5, large effect = 0.8.  
הערה: יש צורך להתייחס לקונטקסט ולתחום המחקר ולא להסתמך רק על כלל האצבע לפרשנות התוצאות. [18]

במקרה שבו השונויות בין הקבוצות שונות מאוד, לא מומלץ להשתמש במדד - אלא דווקא במדד Glass's Delta. [17][18]   
חשוב לציין – אמנם אנחנו לא מניחים שוויון שונויות עבור "מבחני העל" שלנו, אבל ייתכן כי השונויות יהיו דומות בקרב ההשוואות הזוגיות שלנו שיצאו מובהקות, ואז נוכל להשתמש במדד של הדג', שנחשב הטוב ביותר.  
  
לכן, אנחנו נשאף להשתמש במדד של הדג' באופן בלעדי - זה גם יהיה לנו נוח יותר להשוואות כאשר יש סקאלה אחידה. אם השונויות יהיו הטרוגניות מדי, נשתמש גם בגלאס (ונסביר עליו יותר במקרה הזה).

## אופטימיזציה של מבחני ההמשך לפני ניתוח

לפני ביצוע הניתוח הסופי – חשבנו שראוי להתעמק יותר במבחן ההמשך גיימס-הוול ולהבין האם הוא משרת את מטרתנו באופן מיטבי לעומת אלטרנטיבות אחרות. לשם כך נעזרנו במחקרם של שינגאלה וראג'יאגורו [14] ובמקורות ידע נוספים. כשהתעמקנו באופן ביצוע המבחן, שמנו לב לדבר משמעותי וחשוב – דרך ההתמודדות של המבחן עם השוואות מרובות אינה קשורה למספר ההשוואות עצמן, אלא למספר הקבוצות במשתנה הבלתי תלוי. ככל שמספר הקבוצות גדול יותר, ככה הקושי לדחות את H0 גדול יותר. ההבחנה הזו היא קריטית - מכיוון שמשמעותה היא שאם יש לנו במחקר 8 קבוצות, זה לא משנה כמה השוואות זוגיות אנחנו נרצה לעשות, אנחנו עדיין "נשלם" את אותו המחיר שהיינו משלמים אם היינו רוצים לבצע את **כל ההשוואות הזוגיות הקיימות**. כלומר – ניתן להפיק את המקסימום ממבחן ההמשך הזה כאשר אתה מעוניין בכל ההשוואות הזוגיות הקיימות. באופן ספציפי, במחקר שלנו ניצן לבצע 28 השוואות זוגיות לכל ניתוח שונות (8 רמות במשתנה הבלתי תלוי), אבל אנחנו מעוניינים רק ב7 (ישראל מול יתר המדינות). אם כן, ניתוח ההמשך הזה מהווה overkill משמעותי למטרת הניסוי. הדרך הראשונה שחשבנו עליה על מנת להתמודד עם הבעיה הזו היא לבצע כל השוואה זוגית בנפרד – למעשה לבצע 7 מבחני גיימס-הוול ו"להעמיד פנים" שיש לנו רק 2 קבוצות בניסוי בכל הרצה כזאת.   
הגישה הזו מעלה שתי בעיות:

1. גיימס-הוול הוא מבחן שנועד לבדוק את כל ההשוואות הזוגיות כאשר יש 3 קבוצות או יותר, אם מריצים אותו על 2 קבוצות, יש פחות היגיון בשימוש בו.
2. אנחנו מנפחים את האלפא, מכיוון שאנחנו עושים 7 מבחנים שונים בנפרד כעת אנחנו נצטרך לתקן את האלפא בעצמנו, זה לא יהיה מובנה כמו שזה היה במבחן גיימס-הוול.

אנחנו ניתן מענה לשתי הבעיות הללו:

1. אנחנו נשתמש במבחן welch's t-test – ששקול לגיימס-הוול במקרה הפרטי שמספר הקבוצות שווה ל-2.
2. נתקן את האלפא. לשם כך, אנחנו נשתמש במבחן הולמ-בונפרוני סדרתי [19]. התיקון נחשב פחות שמרני לעומת תיקון בונפרוני והוא יחסית פשוט.  
     
   פורמליזציה של המבחן:

* יהיו p-values ממוינים מהנמוך לגבוה *, וההיפותזות התואמות להם . תהי רמת המובהקות שיש לשמר.*
* *האם מתקיים? אם כן, דחה את והמשך, אחרת תפסיק.*
* *האם מתקיים? אם כן, דחה את והמשך, אחרת תפסיק.*
* *ובאופן כללי: יהי k אשר מקיים בחן האם מתקיים. אם כן, דחה את והמשך, אחרת תפסיק.*
* *כאשר , בחן את ההשערה ותסיים.*

הסבר מילולי:

השלב הראשון הוא דירוג הp-values, מהמובהק ביותר למובהק פחות. עכשיו ניקח את האלפא שלנו ונחלק אותה ב-m, זו תהיה האלפא הכי נוקשה. את הp-value הכי מובהק נעמיד למבחן מול האלפא הכי נוקשה. אם הp-value נמוך יותר מהאלפא -נעבור לp-value הבא שהוא הכי מובהק. כעת, החלוקה של אלפא תהיה בm-1. נמשיך ככה עד שהתוצאה תהיה לא מובהקת – ואז נעצור מבלי להסתכל בp-values הבאים. אם התוצאות מובהקות לאורך כל הדרך, ההשוואה האחרונה תהיה בין הp-value הכי פחות מובהק לבין האלפא המקורית.

הערה – הסיבה שאנחנו לא משתמשים שוב בתיקון בונפרוני היא שאנחנו לא רוצים להיות נוקשים מדי, כבר השתמשנו בתיקון הזה ואנחנו לא רוצים לאבד את העוצמה הסטטיסטית שלנו לאור שימוש מוגזם בתיקונים שנחשבים שמרניים.

## Welch's t-test – Test Formulas

*נסביר:*

1. *הסטטיסטי t – המונה הוא ההפרש בין הממוצעים והמכנה הוא סטיית התקן (נקרא גם טעות תקן) של התפלגות הדגימה. למעשה מדובר בחיבור של סטיות התקן של התפלגויות הדגימה של כל אחת משתי הקבוצות שנותן לנו את סטיית התקן של התפלגות הדגימה של החיסור בין שתי קבוצות. אם כן, הסטטיסטי t מביע את ההפרש בין הממוצעים ביחידות של סטיות תקן, והוא למעשה תצפית מסוימת בהתפלגות t.*
2. *df – דרגות החופש. הן חייבות להיות מספר שלם ולכן אנחנו מעגלים את החישוב. העיגול הוא כלפי מטה מכיוון שזה מקשיח את המבחן, דרגות חופש נמוכות מקשות יותר על דחיית השערת האפס.*

*הלוגיקה של ביצוע המבחן: התפלגות הדגימה היא החסרה של שני משתנים מקריים שמתפלגים נורמליים, ולכן היא נורמלית בעצמה. על ידי חיסור ההתפלגות בתוחלת שלה וחילוק באומד לסטיית התקן שלה אנחנו מבצעים טרנספורמציה שעוברת להתפלגות t.*

*הערה: בנוסחה לעיל אין התייחסות לתוחלת של התפלגות הדגימה מכיוון שההנחה היא שהיא 0, כלומר אין הבדל בין הקבוצות.*

*כעת צריך להבין – אין התפלגות t יחידה. לכל מספר של דרגות חופש יש התפלגות t משלו, וככל שעולים בדרגות החופש ההתפלגות נהיית צרה יותר – באינסוף דרגות חופש ההתפלגות היא בעצם התפלגות z. אם כן, על סמך דרגות החופש נדע באיזו התפלגות מדובר, ואז נוכל לזהות כמה קיצוני הסטטיסטי t שלנו ביחס להתפלגות. הרעיון הוא שאם הוא מספיק קיצוני (מעל אחוזון 95 באלפא של 0.05 בהשערה חד צדדית), אז נוכל לדחות את השערת האפס ולומר כי ברמת ביטחון (תלוי באלפא) מסוימת קיים הבדל בין הקבוצות.*

## Analysis Of Variance – Test Formulas

נבהיר – נציג את הנוסחאות לניתוח שונות סטנדרטי שמניח שוויון שונויות. הסיבה לכך היא שהנוסחאות של welch's anova הן הרבה יותר מסורבלות וארוכות אבל הלוגיקה זהה – ההבדל הוא שwelch צריך לבצע התאמות כדי להתמודד עם ההפרה של הנחת שוויון השונויות בין הקבוצות.

*הלוגיקה של ניתוח שונות היא להשתמש בשונויות על מנת להסיק מסקנות על התוחלות.*   
MSW (Mean Square Within) *הוא למעשה השונות הקיימת בתוך קבוצות הניסוי. בגלל ההנחה של שיוויון השונויות בין הקבוצות, מבצעים ממוצע משוכלל של השונויות של קבוצות הניסוי השונות על מנת לאמוד את השונות התוך קבוצתית הזו באופן הטוב ביותר.*

MSB (Mean Square Between) *הוא למעשה השונות בין ממוצעי קבוצות הניסוי. לא ניתן להשוות בין שונות של ממוצעים לשונות של תצפיות "רגילות" מבלי ליצור הטיה, ולכן כל תצפית (ממוצע)* בMSB *מוכפלת במספר הנבדקים שממנו נוצר הממוצע – זה יוצר מעין סקאלה שהיא אחידה.*

*אם כן, סטטיסטי* F *שלנו בעצם משווה את השונות בין הקבוצות לשונות בתוך הקבוצות. תחת השערת האפס, מכיוון שאין הבדל בין הקבוצות, מדובר ב2 אומדים שונים ובלתי תלויים לאותה השונות, ולכן נצפה שהמנה הזאת תהיה קרובה לאחד. מנגד, אם יש אפקט חזק, השונות בין ממוצעי הקבוצות אמורה להיות משמעותית יותר, והשונות הפנימית בתוך הקבוצות צריכה להיות זניחה יותר.   
המנה של שני אומדים (בלתי תלויים) לשונות תצפיות מתפלגת* F *עם דרגות החופש שהצגנו לעיל, ובאמצעות התפלגות* F *עם דרגות החופש המתאימות נדע עד כמה קיצוני סטטיסטי* F *שיצא לנו. אם נחליט שלא סביר לקבל את התוצאה שקיבלנו (הרף נקבע על ידי אלפא) נדחה את השערת האפס ברמת וודאות מסוימת (שגם תלויה באלפא) ונאמר כי לא סביר כי השונויות הגיעו מאותה אוכלוסייה – השונות הבין קבוצתית גדולה מספיק כדי לומר שהיא הגיעה מקבוצות שונות שמובחנות בתוחלתן.*

# ניתוח נתונים – תוצאות

## מוחצנות - Extraversion

המבחנים נעשים בR. במסמך זה נשתמש כדי להציג באופן נוח את הנתונים והתוצאות.  *מייצגת את סטיית התקן המדגמית.*

הערה: נזכיר כי סקאלת חמש התכונות הגדולות נעה בין 10-50.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 29.64 | 29.81 | 29.87 | 28.45 | 29.51 | 29.23 | 29.34 | 29.61 | (Mean) |
| 8.60 | 7.78 | 8.61 | 8.67 | 8.85 | 8.99 | 9.15 | 9.11 | (SD) |
| 8,128 | 722 | 1,353 | 1,321 | 20,196 | 28,188 | 196,897 | 875 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | NS | NS | NS | NS | NS | Significant   P-Value |
| X | X | X | X | X | X | X | Hedge's g Effect Size |

## יציבות רגשית - Emotional Stability

הערה: נזכיר כי סקאלת חמש התכונות הגדולות נעה בין 10-50.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 28.76 | 30.21 | 28.23 | 27.64 | 29.81 | 28.49 | 29.97 | 30.11 | (Mean) |
| 8.41 | 7.84 | 8.21 | 8.30 | 8.57 | 8.61 | 8.66 | 8.67 | (SD) |
| 8,128 | 722 | 1,353 | 1,321 | 20,196 | 28,188 | 196,897 | 875 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| V | NS | V | V | NS | V | NS | Significant   P-Value |
| 0.16  (negligible) | X | **0.22 (small)** | **0.29**  **(small)** | X | 0.19  (negligible) | X | Hedge's g Effect Size |

הערה: החלטנו כי ראוי להשתמש במדד גודל האפקט של הדג', מכיוון שסטיות התקן של המדינות המובהקות כמעט שוות לישראל. אם נראה כי ההבדלים בסטיות התקן משמעותיים נתייחס לזה בהמשך, אחרת נמשיך לחשב את המדד של הדג'.

## נעימות - Agreeableness

הערה: נזכיר כי סקאלת חמש התכונות הגדולות נעה בין 10-50.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 37.34 | 37.56 | 38.05 | 36.77 | 38.04 | 37.00 | 38.37 | 36.86 | (Mean) |
| 6.42 | 5.77 | 6.34 | 7.13 | 7.10 | 7.74 | 7.28 | 7.06 | (SD) |
| 8,128 | 722 | 1,353 | 1,321 | 20,196 | 28,188 | 196,897 | 875 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | V | NS | V | NS | V | Significant   P-Value |
| X | X | -0.18  (negligible) | X | -0.17  (negligible) | X | **-0.21 (small)** | Hedge's g Effect Size |

הערה: המשמעות של גודל האפקט השלילי היא שישראל הפגינה נעימות נמוכה לעומת המדינות האחרות. בטבלה הקודמת המספר היה חיובי כי ישראל הפגינה יציבות רגשית גבוהה יותר.

## מצפוניות - Conscientiousness

הערה: נזכיר כי סקאלת חמש התכונות הגדולות נעה בין 10-50.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 33.61 | 33.40 | 34.00 | 33.03 | 33.77 | 32.38 | 34.13 | 33.14 | (Mean) |
| 7.40 | 6.60 | 7.23 | 7.20 | 7.37 | 7.39 | 7.45 | 7.06 | (SD) |
| 8,128 | 722 | 1,353 | 1,321 | 20,196 | 28,188 | 196,897 | 875 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | NS | NS | NS | NS | V | Significant   P-Value |
| X | X | X | X | X | X | -0.13  (negligible) | Hedge's g Effect Size |

## פתיחות לחוויות – Openness to experience

הערה: נזכיר כי סקאלת חמש התכונות הגדולות נעה בין 10-50.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 37.76 | 36.60 | 37.61 | 38.75 | 38.67 | 39.27 | 39.21 | 40.90 | (Mean) |
| 6.11 | 6.18 | 5.95 | 6.06 | 6.19 | 6.12 | 6.17 | 5.95 | (SD) |
| 8,128 | 722 | 1,353 | 1,321 | 20,196 | 28,188 | 196,897 | 875 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| V | V | V | V | V | V | V | Significant   P-Value |
| **0.52**  **(medium)** | **0.71**  **(medium)** | **0.55**  **(medium)** | 0.36  (small) | 0.36  (small) | 0.27  (small) | 0.27  (small) | Hedge's g Effect Size |

## מקיאווליזם - Machiavellism

הערה: נזכיר כי סקאלת המקיאווליזם נעה בין 20-100.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 69.55 | 67.32 | 68.51 | 73.56 | 66.09 | 68.76 | 65.81 | 68.86 | (Mean) |
| 14.73 | 12.33 | 14.44 | 15.34 | 16.12 | 15.89 | 16.18 | 16.11 | (SD) |
| 1,507 | 885 | 200 | 189 | 2,555 | 5,113 | 31,675 | 140 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | NS | NS \* | NS | NS | NS | Significant   P-Value |
| X | X | X | X | X | X | X | Hedge's g Effect Size |

\*הסהר הפורה היה קרוב מאוד למובהקות – תקנון ליברלי יותר של האלפא היה משיג מובהקות בהשוואה הזו. באופן ספציפי, היה הבדל של 0.00005 בין האלפא המתוקננת לבין הp-value.

## נרקיסיזם - Narcissism

הערה: נזכיר כי סקאלת נרקיסיזם נעה בין 1-5.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 3.217 | 3.120 | 3.616 | 3.391 | 3.054 | 3.055 | 3.096 | 3.036 | (Mean) |
| 0.622 | 0.627 | 0.770 | 0.727 | 0.796 | 0.787 | 0.773 | 0.568 | (SD) |
| 214 | 49 | 24 | 25 | 751 | 2,753 | 9,082 | 25 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | NS | NS | NS | NS | NS | Significant   P-Value |
| X | X | X | X | X | X | X | Hedge's g Effect Size |

## פסיכופתיה - Psychopathy

הערה: נזכיר כי סקאלת פסיכופתיה נעה בין 1-5.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 2.898 | 2.796 | 2.981 | 3.124 | 2.847 | 2.880 | 2.800 | 2.640 | (Mean) |
| 0.763 | 0.666 | 0.718 | 0.742 | 0.822 | 0.802 | 0.829 | 0.652 | (SD) |
| 214 | 49 | 24 | 25 | 751 | 2,753 | 9,082 | 25 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | NS | NS | NS | NS | NS | Significant   P-Value |
| X | X | X | X | X | X | X | Hedge's g Effect Size |

## הערכה עצמית – Self Esteem

הערה: נזכיר כי סקאלת ההערכה העצמית נעה בין 0-30.

***סטטיסטיקה תיאורית***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | Israel | - |
| 18.10 | 17.74 | 19.00 | 18.72 | 15.68 | 14.90 | 16.35 | 18.10 | (Mean) |
| 6.16 | 5.38 | 5.77 | 5.58 | 6.72 | 7.11 | 7.13 | 6.38 | (SD) |
| 1,253 | 69 | 112 | 309 | 2,313 | 6,459 | 21,810 | 98 | N |

***תוצאות (Israel vs)***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Far East Culture | | Arab Culture | | West Culture | | | - |
| India | China | United Arab Emirates | Fertile Crescent | Australia | United Kingdom | USA | - |
| NS | NS | NS | NS | V | V | NS | Significant   P-Value |
| X | X | X | X | **0.36**  **(small)** | **0.45**  **(small)** | X | Hedge's g Effect Size |

# הסקת מסקנות

## שאלת המחקר

נרצה לנסות לסווג את הפרופיל האישיותי הישראלי לאחת מתרבויות העל של המחקר על סמך הניתוחים שעשינו – לשם כך ניעזר בגדלי האפקט. ככל שתרבות מסוימת תציג הבדלים משמעותיים יותר בתכונה לעומת ישראל, נוכל להגיד ביותר ביטחון כי האוכלוסייה הישראלית שונה מאותה תרבות בתכונה הזו. כאשר לא יהיו הבדלים, נוכל להגיד כי האוכלוסייה הישראלית דומה לאותה תרבות באותה תכונה. נרצה לקחת בחשבון את כל התכונות כדי להגיע למסקנה.

נסכם את כל הממצאים בטבלה:

Hedge's g effect sizes (Israel – World Countries)

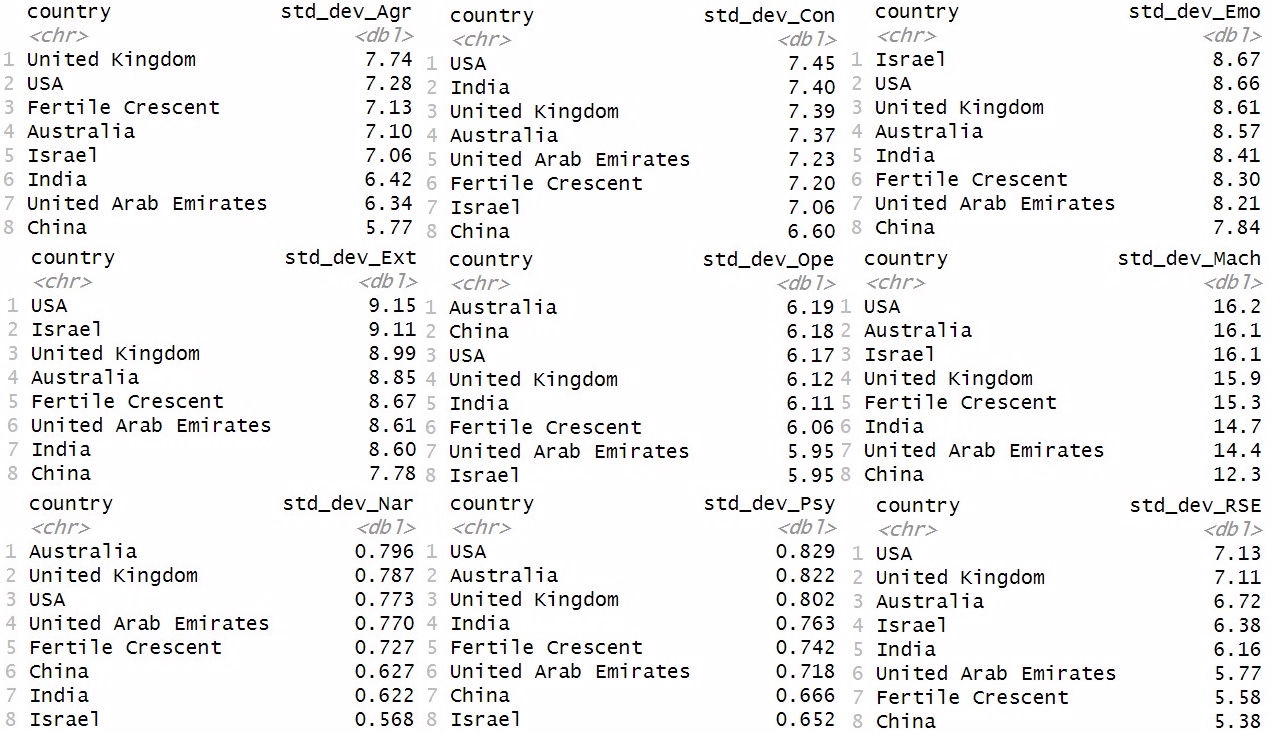
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Culture | West Culture | | | Arab Culture | | Far East Culture | |
| Country  Trait | USA | United Kingdom | Australia | Fertile Crescent | United Arab Emirates | China | India |
| Extraversion | X | X | X | X | X | X | X |
| Emotional Stability | X | 0.19 (negligible) | X | **0.29 (small)** | **0.22 (small)** | X | 0.16 (negligible) |
| Agreeableness | **-0.21 (small)** | X | -0.17 (negligible) | X | -0.18 (negligible) | X | X |
| Conscientiousness | -0.13 (negligible) | X | X | X | X | X | X |
| Openness To Experience | **0.27 (small)** | **0.27 (small)** | **0.36 (small)** | **0.36 (small)** | **0.55 (medium)** | **0.71 (medium)** | **0.52 (medium)** |
| Machiavellism | X | X | X | \*X | X | X | X |
| Narcissism | X | X | X | X | X | X | X |
| Psychopathy | X | X | X | X | X | X | X |
| Self Esteem | X | **0.45**  **(small)** | **0.36**  **(small)** | X | X | X | X |

\*היה קרוב מאוד למובהקות.

* אין בסיס לחשוב שקיים הבדל ברמת המוחצנות בין ישראל לבין אף אחת מהתרבויות ברמה ההיסקית. גם ברמה התיאורית לא נראים הבדלים משמעותיים.
* יש בסיס לחשוב שקיים הבדל ברמת היציבות הרגשית בין ישראל לבין כל התרבויות ברמה ההיסקית. ההבדל המשמעותי ביותר הוא ללא ספק בין ישראל לעולם הערבי – שם אנחנו רואים 2 גדלי אפקט קטנים. ניכר כי ישראל יציבה יותר רגשית מהעולם הערבי. ישראל גם יציבה רגשית יותר מהמזרח והמערב, אבל ההבדלים זניחים יותר. אם כן, האוכלוסייה הישראלית יותר דומה למזרח ולמערב כשזה נוגע ליציבות רגשית.
* יש בסיס לחשוב שקיים הבדל ברמת הנעימות בין ישראל לבין המערב והעולם הערבי ברמה ההיסקית. באופן אולי מפתיע, האפקט החזק ביותר דווקא נצפה בין ישראל לארצות הברית. ישראל פחות "נעימה" מהאוכלוסייה המערבית, והיא גם פחות "נעימה" מהאוכלוסייה הערבית. לא נמצאו הבדלים מובהקים בין ישראל למזרח, ולכן ניתן לומר כי האוכלוסייה הישראלית הכי דומה למזרח כשזה נוגע לנעימות.
* יש בסיס לחשוב שקיים הבדל ברמת המצפוניות בין ישראל לבין המערב ברמה ההיסקית. ישראל מראה מצפוניות נמוכה יותר ביחס למערב, ובאופן ספציפי ביחס לארצות הברית. יש לציין כי גודל האפקט קטן במיוחד – אך עדיין קיים. אם כן, האוכלוסייה הישראלית יותר דומה למזרח ולעולם הערבי כשזה נוגע למצפוניות.
* יש בסיס לחשוב שקיים הבדל ברמת הפתיחות לחוויות בין ישראל לבין כל התרבויות ברמה ההיסקית. ישראל הראתה את הפתיחות לחוויות הגבוהה ביותר – לא משנה מול איזו תרבות. באופן שהוא אולי אינטואיטיבי, גדלי האפקט הקטנים ביותר הם מול המערב. מול העולם הערבי גדלי האפקט קטנים-בינוניים, ומול המזרח גדלי האפקט בינוניים בלבד. במקרה הזה, קל לראות כי האוכלוסייה הישראלית הכי דומה למערב כשזה נוגע לפתיחות לחוויות.
* אין בסיס לחשוב שקיים הבדל ברמת המקיאווליזם בין ישראל לבין אף אחת מהתרבויות ברמה ההיסקית. למרות זאת, בין ישראל והסהר הפורה היה הבדל תיאורי משמעותי שהיה על סף מובהקות. לכן, ניתן לומר בעירבון מוגבל כי האוכלוסייה הישראלית מזכירה יותר את המערב ואת המזרח הרחוק מאשר את העולם הערבי כשזה נוגע לתכונת האופי מקיאווליזם.
* אין בסיס לחשוב שקיים הבדל ברמת הנרקיסיזם בין ישראל לבין אף אחת מהתרבויות ברמה ההיסקית. למרות זאת, ברמה התיאורית, ישראל נראית ממש כמו אחת המדינות מהעולם המערבי. הפער התיאורי הכי גדול היה בין ישראל לבין העולם הערבי. קיים חשש כי לא השגנו הבדל מובהק מכיוון שהעוצמה שלנו לא הייתה גבוהה מספיק – תכונה זו נמדדה על סמך מספר תצפיות נמוך יחסית.
* אין בסיס לחשוב שקיים הבדל ברמת הפסיכופתיה בין ישראל לבין אף אחת מהתרבויות ברמה ההיסקית. למרות זאת, ברמה התיאורית, נראה כי ישראל שונה מהעולם הערבי. קיים חשש כי לא השגנו הבדל מובהק מכיוון שהעוצמה שלנו לא הייתה גבוהה מספיק – תכונה זו נמדדה על סמך מספר תצפיות נמוך יחסית.
* יש בסיס לחשוב שקיים הבדל בהערכה העצמית בין ישראל לבין המערב ברמה ההיסקית. ישראל מראה הערכה עצמית גבוהה באופן מובהק ביחס למערב. אם כן, האוכלוסייה הישראלית דומה יותר למזרח ולעולם הערבי כשזה נוגע להערכה עצמית.

**מסקנה**: כאשר מסתכלים רק על ההבדלים המובהקים (להם חישבנו גדלי אפקט) – נראה כי הסיווג ההולם ביותר לישראל הוא דווקא סיווג לעולם המזרח הרחוק. ישראל הכי דומה למזרח ב4 תכונות: יציבות רגשית, נעימות, מצפוניות והערכה עצמית. לעומת זאת, היא הכי דומה למערב ב2 תכונות: יציבות רגשית ופתיחות לחוויות, והכי דומה לעולם הערבי ב2 תכונות: מצפוניות והערכה עצמית.  
חשוב לציין – הדרך שבה בחרנו את הסיווג הנכון ביותר היא דרך אחת מיני רבות, והמסקנה שהתקבלה אינה מתיימרת להיות חד משמעית. יש לשים לב כי רוב האפקטים שהתקבלו הם אינם אפקטים חזקים, והדפוסים לא תמיד היו ברורים - ולכן רמת הביטחון בנוגע לסיווג הכי הולם היא נמוכה יותר. התכונה היחידה שהראתה חוקיות ברורה לעין ואפקטים משמעותיים היא פתיחות לחוויות, ובאשר אליה אנחנו מרגישים הרבה יותר ביטחון כשאנו מביטים בממצאים. בנוסף, למחקר זה היו הרבה מגבלות (עליהן נפרט בהמשך) ולכן אנחנו חושבים כי חשוב להמשיך ולחקור את הנושא במחקרי המשך שלא יסבלו ממגבלות אלו, על מנת לתת ביסוס נוסף לממצאים שהתגלו במחקר ואף להגיע לממצאים חדשים ומעניינים בתחום.

## ממצאים נוספים

1. ישראל היא הפתוחה ביותר לחוויות מבין כל מדינות העולם שהשתתפו במחקר ומראה אפקטים מובהקים ומשמעותיים לעומת שלושת תרבויות העל שלנו. מעניין לראות את ההדרגתיות – ככל ש"הולכים מזרחה" הפתיחות לחוויות יורדת. מתבקש לנסות ולהבין למה ישראל כל כך מובחנת משאר העולם כשזה נוגע לפתיחות לחוויות. נרצה להציע כמה אפשרויות למחשבה לממצא הזה כשאנחנו מתבססים על אופייה המיוחד של האוכלוסייה הישראלית:
2. מדינת ישראל היא ישות מאוד הטרוגנית, ואפשר לומר שהיא נולדה כזו. מדינת ישראל הוקמה קודם כל כמקום מקלט עבור אוכלוסייה בסיכון – מיעוט דתי שעבר רדיפות ופוגרומים במשך דורות רבים בכל העולם. מה שאיחד את העם היהודי יותר מכל דבר אחר (אפילו יותר מהדת עצמה) ואפשר לנו להקים ולשמר את מדינת ישראל הוא הרצון הבסיסי שלנו לשרוד. הרצון הזה לאפשר קיום עבור היהודים היה קריטי יותר מכל דבר אחר - לכן אנשים ממדינות שונות, רקע תרבותי שונה, חינוך שונה, גזע שונה ואפילו רמת אמונה שונה בדת שמו את שאר ההבדלים ביניהם בצד לצורך מטרה חשובה יותר. אם כן, שיתוף הפעולה הזה בין בני אדם שיכלו להיות מאוד שונים באספקטים רבים חייב לפתח גמישות מחשבתית ופתיחות – והנסיבות אילצו אותנו לכך. היהודים תמיד נדרשו לגמישות ויכולות הסתגלות גבוהות במהלך השנים בגלל המציאות הלא צפויה והלא יציבה שתמיד נדרשנו לעמוד בה – סנקציות, רדיפה, ואפילו גירוש בגלות, ומלחמות ומציאות ביטחונית בלתי יציבה בישראל. מעבר לכך, האיום הקיומי תמיד דחף את הישראלים לחדשנות ויצירתיות על מנת להשיג יתרונות על אויבנו ולהבטיח את העתיד שלנו כאן.  
   פתיחות לחוויות משמעותה גמישות, חשיבה על רעיונות יצירתיים, הערכה לגיוון חוויתי, רצון ללמוד ועוד – כל אלו הן יכולות שהישראלים נדרשו לפתח בגלל הנסיבות הייחודיות שלנו לעומת מדינות אחרות.  
   חשוב לציין – קבלת האחר וגיוון תרבותי ואתני הם ערכים כלליים של המערב ושל הגישה הדמוקרטית, ולכן זה מסתדר שהמערב היא התרבות שהכי קרובה אלינו בפתיחות לחוויות (אבל עדיין בצורה מובהקת נמוכה מאיתנו).
3. מדינת ישראל היא "אומת ההייטק", ויש עידוד משמעותי בחברה הישראלית לרדוף אחר החדשנות. מעבר לרצון לחולל שינוי ותחושת שליחות, המציאות הכלכלית ויוקר המחיה בישראל נותנות תמריץ משמעותי ללכת לעולם ההייטק ולנסות "להצליח בגדול", על מנת לאפשר חיים בכבוד בישראל. אם כן, מופעל יותר לחץ על האוכלוסייה הישראלית לרדוף אחר החדשנות מאשר אוכלוסיות אחרות.
4. סין הראתה את השונות הנמוכה ביותר כמעט בכל המדידות שערכנו בניסוי:



ניתן לחשוב כי הסיבה לכך היא שהמדגמים של סין היו קטנים יחסית, ולכן לא באה לידי ביטוי השונות שקיימת באוכלוסייה. למרות זאת, אין עקביות בנתונים מהבחינה הזו – סין הראתה שונות נמוכה מאוד גם כאשר המדגם שלה היה גדול מאוד ביחס לאחרים. דוגמא בולטת לזה הייתה בשאלון [המקיאווליזם](#סטייתתקןמקיוואליזם). לפחות כשמדובר בשילוש האפל, לא נראה שגודל המדגם קשור להומוגניות שאנחנו רואים בסין. כאשר מדובר בתכונות האחרות, זה נראה שיש יותר מקום לדיבייט (לפחות לפי ההתרשמות שלנו).   
ייתכן שאפשר להסביר את ההומוגניות הזו דרך הדמוגרפיה הפחות מגוונת יחסית למערב (שבו מדיניות ההגירה יותר משוחררת), ועם המשטר המדכא של סין – שאולי "מהנדס" את האוכלוסייה שלו להתנהגות מאוד מסוימת, ואינו מוכן להכיל התנהגויות אחרות. כלומר, ייתכן כי החינוך בסין מגביל את כיווני התפתחות האדם בבגרותו, וייתכן גם כי יש חשש לדווח על התנהגויות או דעות שהן פחות קונפורמיסטיות תחת משטר כזה.

3. העולם הערבי הראה את מדידות השילוש האפל ואת ההערכה העצמית הגבוהות ביותר ברמה התיאורית באופן עקבי. אנחנו חוששים שבגלל המדגמים הקטנים יחסית של העולם הערבי במדידות הללו, והמבחנים הרבים שערכנו (שגרמו לנו להקשיח את הדרישות למובהקות סטטיסטית), לא הייתה לנו מספיק עוצמה סטטיסטית על מנת לגלות את האפקטים הללו ברמה ההיסקית. למרות זאת, יש לנו סיבה לחשוב כי הם קיימים וכי שווה לשים עליהם את הזרקור במחקרים עתידיים על סמך התוצאות התיאוריות. אם וכאשר הבדלים מובהקים כאלה יימצאו, יהיה מעניין לדון בנוגע לסיבות להבדלים הללו.

## מגבלות המחקר

1. המחקר לא התבסס על דגימה מקרית של נבדקים. האוכלוסייה שנבדקה במחקר זה הינה בעצם "אנשים שנכנסו לאתר OpenPsychometrics ובחרו למלא שאלונים באנגלית באופן התנדבותי". ניתן לחשוב כי האוכלוסייה הזו אינה מייצגת את האוכלוסייה הכללית במדינות השונות במחקר ומייצרת הטיות, ניתן דוגמאות להטיות אפשריות:
2. תרבות המזרח והתרבות הערבית (וגם ישראל) אינן מדברות אנגלית כשפה ראשונה – משמעות הדבר היא שייתכן כי הנבדקים לא הבינו את השאלון באותה רמת דיוק של המדינות המערביות שמדברות אנגלית כשפה ראשונה, מדובר פה בפער שיוצר הטיה מסוימת ויכול לפגוע באמינות המידע.
3. הצורך להבין אנגלית הוא למעשה חסם לנתחים מהאוכלוסיות עבור המדינות שאינן מדברות אנגלית כשפת אם. המשמעות היא שדגמנו נתחים מסוימים מאותן אוכלוסיות ולא הגענו לכל הקשת הקיימת בהן. ככל שהאוכלוסייה פחות יודעת אנגלית, ההטיה הזו למעשה גדלה – כי אנחנו דוגמים נתח קטן יותר מהאוכלוסייה הקיימת באותה מדינה. האיום הזה כנראה משמעותי יותר עבור מדינות פחות מפותחות כמו למשל לבנון או ירדן.
4. ניתן לחשוב כי באופן כללי האוכלוסייה שנכנסת לאינטרנט על מנת לבצע מבחני אישיות בהתנדבות היא אוכלוסייה יותר פתוחה, ליברלית ואולי אפילו אינטליגנטית מהאדם הממוצע (האינטליגנציה כנראה קשורה יותר כאשר אנגלית היא אינה השפה הראשונה של המדינה).
5. ישנה בעיה של מהימנות בהקשר לשאלוני דיווח עצמי. באופן ספציפי, ישנה בעייתיות בדיווח עצמי בהקשר לתכונות חיוביות או שליליות. רוב בני האדם רוצים לראות את עצמם כאינדיבידואליים בעלי תכונות חיוביות, והם פחות רוצים לראות את עצמם כבעלי תכונות שליליות. ניתן לחשוב כי ההנחה שהאדם כנה במילוי השאלון וגם מודע לעצמו מספיק כדי לדווח תשובות מדויקות הינה בעייתית. לדוגמא – ייתכן כי אדם ידווח כי הוא אדם נעים וחברותי, אבל עמיתיו לעבודה ידווחו באופן אחיד וחד משמעי כי לא כך הדבר. לכן, כשמדובר בתכונות שאינן ניטרליות, קיימות אלטרנטיבות טובות יותר למדידה מאשר דיווח עצמי – כמו למשל דיווח מהסביבה, או מדידה של ההתנהגות עצמה במצבי מבחן. כמובן שקשה יותר להשיג מדידות מהסוג הזה, ולכן מסתפקים הרבה פעמים במדידות טובות פחות כמו הדיווח העצמי.
6. היו פערים עצומים בגדלי המדגמים במחקר שלנו, וזה גרם למספר קשיים:
7. אנובה הוא מבחן עמיד בפני הפרה של שיוויון שונויות, אבל רק כאשר גדלי המדגמים הינם שווים. כאשר גדלי המדגמים שונים משמעותית, האנובה מאבד את האמינות שלו וכבר מומלץ שלא להשתמש בו. נאלצנו לנתח את הנתונים בדרכים פחות שגרתיות ואולי גם עם עוצמה נמוכה יותר לאור המצב הזה.
8. פערי עוצמה. הפערים הגדולים בגדלי המדגמים גרמו לכך שיהיה קל יותר לגלות אפקטים במערב, מה שיכול לגרום לפרשנות לא נכונה של התוצאות. למשל, מי שמסתכל על התוצאות בעין שאינה ביקורתית, יכול לחשוב כי קיימים הבדלים בין המערב לישראל, אבל לא קיימים הבדלים בין המערב לעולם הערבי. בפועל, המערב "נהנה" מכמות אדירה של דגימות, והעולם הערבי "סבל" מכמות קטנה של דגימות. לכן, קשה היה להשיג מובהקות סטטיסטית בעולם הערבי לעומת העולם המערבי. המצב הזה יכול להוביל למסקנה הבעייתית שאין הבדלים בין ישראל לעולם הערבי, אבל כן יש הבדלים בין ישראל למערב. בפועל, אנחנו מאמינים כי לא הצלחנו לגלות אפקטים קיימים בעולם הערבי בגלל מספר הדגימות המצומצם שנאלצנו לעבוד איתו.
9. המשך לסעיף ב' – בישראל ובעולם הערבי היה מספר קטן של דגימות באופן כללי בחלק מהמבחנים, ואנו מאמינים כי היכולת שלנו לגלות אפקטים נפגעה לאור זאת.
10. עבור חלק מהמשתנים שנמדדו והופיעו בדאטה הגולמי, היה חסר מידע חשוב בנוגע לאיך שהם נמדדו, גם בcodebook המצורף לדאטה. המידע היה חשוב על מנת להכריע איך להתנהל עם הדאטה במקרים מסוימים. למשל – בחלק מהשאלונים נמדדו גם הזמנים שלקח לנבדקים לענות על השאלות. מדובר במידע חשוב ומהותי, כדי שהחוקר יוכל לבחון האם הנבדק ענה על השאלות בצורה רצינית ומהימנה. הבעיה הייתה שאופן החישוב של הזמנים לא הועבר במספיק בהירות, ולא היה ברור כיצד מדידת הזמנים מתבצעת כאשר נבדק חוזר אחורה בשאלון (במקרה שהוא למשל מדלג בטעות על שאלה). הידע הזה הוא מהותי על מנת להכריע האם הנבדק ענה על השאלה בצורה רצינית, וכאשר הוא לא נמצא – קיים קושי בהכרעה בנוגע לאמינות של תשובות מסוימות. מנגד, היו גם שאלונים שכמעט לא סיפקו מידע שאפשר להשתמש בו על מנת לזהות משיבים בעייתיים – כמו למשל זמני מילוי לשאלות.
11. ייתכן כי המחקר היה שאפתני מדי. רצינו לבצע מבחנים סטטיסטיים על מנת לבחון הבדלים מובהקים בין ישראל לבין כל תרבויות העל בעולם. על אף שניסינו לייצג כל תרבות כזו באופן ממצה ככל הניתן, עדיין היו לנו 8 רמות במשתנה הבלתי תלוי על פני 9 תכונות – והמשמעות הייתה ביצוע מבחנים רבים. רצינו לגלות את כל האפקטים, ורצינו גם לשמור על הרזולוציה המדינית (כי חשבנו שקטלוג שמבוסס רק על תרבות העל עצמה מבצע רדוקציה גדולה מדי ונותן מידע לא מספיק ספציפי) – ייתכן כי היינו צריכים להיות צנועים יותר במטרתנו. העלות של ביצוע כל המבחנים הללו הייתה שנאלצנו לתקן את האלפא שלנו באופן יחסית אגרסיבי אם אנחנו רוצים לשמור על אמינות המחקר. בעצם, איבדנו עוצמה בכל ההשערות על מנת שתהיה לנו הזדמנות לבדוק את כל ההשערות. אם היינו משערים פחות, תיקון האלפא היה אגרסיבי פחות, והעוצמה לגלות את כל האפקטים הייתה גדלה.  
    חשוב לציין בכל זאת - אם גדלי האפקט במחקר היו גדולים מספיק, היינו מגלים את האפקטים למרות התיקונים שביצענו לאלפא.

# נספחים

## סולם ליקרט

סולם ליקרט הוא הכלי הנפוץ ביותר למדידת תשובות במחקרים שהם מבוססי סקרים. בגדול, משתמשים במונח בשני מצבים:

1. בשביל לתאר פורמט שבו תשובות מקבלות ציון מספרי לאורך טווח.   
   למשל: 1 – לא מסכים, 2 – מסכים מעט, 3 – ניטרלי וכו'.
2. בשביל לתאר סולם שנוצר מאינטגרציה של היגדי ליקרט (לרוב סכימה שלהם או ממוצע שלהם).   
   למשל: סכימה של היגדי ליקרט שמודדים הערכה עצמית על מנת ליצור טווח מספרי של ציוני הערכה עצמית.

טכנית, ליקרט (יוצר הסולם) התכוון לאפשרות השניה.

היצירה של סולם על ידי סכימה או ממוצע של ההיגדים מניחה למעשה את הליניאריות של היגדי הליקרט – במילים אחרות, היא מניחה שהמרחקים בין אפשרויות התשובות השונות הם למעשה שווים – ההבדל בין 1 ל-2 הוא זהה להבדל בין 2 ל3 וכן הלאה. הנחה נוספת היא (אלא אם כן משתמשים במשקלים לממוצע) היא שכל היגד שווה בחשיבותו על מנת לקבל ציון סופי.

הנחת הליניאריות מאפשרת לנו להתייחס לסולם ליקרט בתור משתנה מסולם רווח\*\*, שניתן לניתוח על ידי מבחנים פרמטרים ולכן יותר גמיש לעבודה. אם ההנחות אינן מתקיימות, אז סולם ליקרט הוא למעשה משתנה מסולם סדר\*, ומשתנה מסוג כזה אינו יכול להיות משומש במבחנים פרמטריים.

\*סולם סדר - הערכים המספריים בסולם נועדו לסמל היררכיה מסוימת, אבל אין משמעות כמותית להבדלים המספריים. לדוגמא – 1 = מנהל, 2 = ראש מחלקה, 3 = פועל.

\*\*סולם רווח – בין הערכים של סולם רווח קיימים מרווחים שווים המציינים הבדלים שווים בתכונה הנמדדת. לסולם רווח אין נקודת אפס אבסולוטית. לדוגמא – טמפרטורה.

על מנת שסולם הליקרט יהיה איכותי ויאפשר הנחת ליניאריות, חשוב לבנות את היגדי הליקרט באופן מחושב וזהיר. נקודה אחת שיש לקחת בחשבון היא הסימטריה – יש צורך לתת מספר שווה של אפשרויות "חיוביות" ו"שליליות" סביב נקודת הניטרליות (בין אם היא נמצאת באפשרויות ובין אם לא). נקודה נוספת משמעותית ועדינה יותר – יש לדאוג להשתמש בנוסחים שניתן להסיק מהם מרחק שווה בין הערכים (קשור ישירות להנחת הליניאריות).

דוגמא לסולם ליקרט טוב (כמו במחקר שלנו):

1-לא מסכים, 2 -לא מסכים במקצת, 3 – ניטרלי, 4 – מסכים במקצת, 5 – מסכים.

דוגמא לסולם ליקרט לא טוב:

1 – לא מסכים, 2 – ניטרלי, 3 – מסכים, 4 – מסכים מאוד.

בספרות קיימת תמיכה לכך שסולם ליקרט יכול לשמש למבחנים פרמטריים. דוגמא לתמיכה כזו ניתן למצוא במחקרו של ג'וף נורמן שהתבצע בשנת 2010. [7]

## ציינון השאלונים

הערה: כל הורדה של דאטה סט מהאתר OpenPsychometrics כוללת בתוכה codebook שמסביר את המשמעות של כל עמודה. בפרט, המסמך מכיל את הנוסחים המדויקים של כל השאלות.   
לדוגמא – אם ישנה עמודה בשם 1Q בדאטה סט "הערכה עצמית", אז קיים בcodebook של דאטה סט "הערכה עצמית" הנוסח המדויק של השאלה הזו.

1. Rosenberg Self-Esteem Scale - ציינון השאלון על פי W.W Norton [1]. סדר השאלות זהה לסדר השאלות בשאלון שלנו והניסוח זהה – ניתן לאמת זאת באמצעות מסמך ה-codebook.

הערה: ההבדל היחיד הוא שטווח המספרים בשאלון שלנו מתחיל מ1 עד 4, בעוד הטווח של W.W Norton מתחיל ב-0 ונגמר ב-3, מה שאומר שהציונים אצלנו ינועו בין 10-40 לעומת 0-30 אצל נורטון (10 היגדים בסך הכל). זה לא באמת משנה, ולאחר מתן הציונים נוכל להזיז את הסקאלה ל0-30 על ידי החסרה של 10 נקודות באופן אחיד לכל ציון הערכה עצמית בטבלת המידע (חוקרים רבים עובדים עם הסקאלה 0-30 והיא יותר אינטואיטיבית).

1. Machiavellianism Test – מבחן המקיאוואליזם שלנו הוא גרסת MACH -IV, והיא הנפוצה ביותר. יש כמה דרכים לציינן את השאלון, ובמאגר שלנו משתמשים בסקאלת ליקרט של 5 נקודות. ציון מקיאווליזם נקבע על ידי סכימה של 20 ההיגדים, מה שאומר שהסקאלה נעה בין 20-100. אופן הציינון הזה הוא פחות נפוץ למבחני MACH-IV אבל הוא קיים, ונעשה בו שימוש למשל במחקרם של אלטאן ואונסאר ב-2013 [3]. יש לקחת בחשבון שיש להפוך את התשובה ב-10 היגדים בשאלון שמשמעותם הפוכה, לשם כך נעזרנו בספר Studies in Machiavellianism שמנחה כיצד עושים זאת. [4]
2. Short Dark Triad – הציינון של כל אחת מהתכונות במאגר התבצע כפי שהתבצע במחקר המקורי שהציג את שאלון השילוש האפל הקצר (ציטטנו אותו בשלב המחקרי, אבל נוסיף רפרנס גם במסמך זה). אם כן, במחקרם של דניאל ג'ונס ודלרוי פולהאוס שהתבצע בשנת 2014 [5], הוצג שאלון זה לראשונה, והציינון התבצע באופן הבא:  
   המשתתפים השיבו לכל ההיגדים באמצעות סולם ליקרט הנע בין 1-5. כל תכונה אפלה נמדדה באמצעות 9 היגדים, וציון התכונה האפלה היה ממוצע התשובות לאותם 9 היגדים. חשוב לציין – קיימים היגדים הפוכים בשאלון, ולכן יש צורך להפוך אותם לפני חישוב הממוצע. בנוסף, באותו מחקר יש נספח שמציג את ההיגדים עצמם, ומציג אילו היגדים יש צורך להפוך. בהתבסס על מידע זה, ידענו כיצד לציינן את התכונות במחקר שלנו.
3. Big Five Factor Markers – OpenPsychometrics משתמש ב-50 ההיגדים שנמצאים במאגר IPIP (International Personality Item Pool) על מנת למדוד את חמש התכונות הגדולות. באתר הרשמי שנוצר למאגר [6], קיימת רשימה מפורטת של כל ההיגדים של השאלון, ועבור כל היגד מפורט איזו תכונה הוא נועד למדוד, והאם הוא מהופך. בנוסף, האתר מדריך כיצד מצייננים את חמש התכונות. כל תכונה מקושרת ל10 היגדים, וכל היגד מקושר למספר בטווח 1-5 שמסמל את מידת ההסכמה של המשיב עם ההיגד, כאשר 1 מביע אי הסכמה מלאה, ו5 מביע הסכמה מלאה. את התשובות להיגדים המהופכים יש צורך להפוך, ולבסוף מבצעים סכימה של כל 10 היגדים לציון סופי עבור התכונה. אם כן, כל תכונה תקבל ציון שנע בין 10-50.

# בילביוגרפיה

1. *Rosenberg’s Self-Esteem Scale*. (n.d.). Retrieved April 7, 2024, from <https://wwnorton.com/college/psych/psychsci/media/rosenberg.htm>
2. Park, H. M. (2015). *Hypothesis testing and statistical power of a test*. <https://scholarworks.iu.edu/dspace/handle/2022/19738>
3. Altan, A., Ünsar, S., & Kahraman, G. (2013). A research on the determination of Machiavellian personality tendencies. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, *14*(1), 103–121. <https://dergipark.org.tr/en/pub/ogusbd/issue/11003/131657>
4. Christie, R., & Geis, F. L. (2013). *Studies in machiavellianism*. Academic Press. <https://books.google.com/books?hl=iw&lr=&id=d5tGBQAAQBAJ&oi=fnd&pg=PP1&dq=Machiavellianism+scale+scoring+instructions&ots=rkiJCEQEF1&sig=-6kwfyfdMNANMkQZpIi5tihgzSY>
5. Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. *Assessment*, *21*(1), 28–41. <https://doi.org/10.1177/1073191113514105>
6. *Administering IPIP Measures, with a 50-item Sample Questionnaire*. (n.d.). Retrieved April 16, 2024, from <https://ipip.ori.org/New_IPIP-50-item-scale.htm>
7. Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
8. *MANOVA*. (n.d.). Statistics Solutions. Retrieved April 19, 2024, from <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/manova/>
9. *Parametric and Non-parametric tests for comparing two or more groups | Health Knowledge*. (n.d.). Retrieved April 19, 2024, from <https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1b-statistical-methods/parametric-nonparametric-tests>
10. Zach. (2019, December 22). How to Check ANOVA Assumptions. *Statology*. <https://www.statology.org/anova-assumptions/>
11. Grace-Martin, K. (2020, December 18). When Unequal Sample Sizes Are and Are NOT a Problem in ANOVA. *The Analysis Factor*. <https://www.theanalysisfactor.com/when-unequal-sample-sizes-are-and-are-not-a-problem-in-anova/>
12. Tim. (2016, October 20). *Welch’s ANOVA: Definition, Assumptions*. Statistics How To. <https://www.statisticshowto.com/welchs-anova/>
13. Liu, H. (2015). *Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance*. Virginia Commonwealth University. <https://search.proquest.com/openview/8265640b89baba766064b16099df3991/1?pq-origsite=gscholar&cbl=18750>
14. Shingala, M. C., & Rajyaguru, A. (2015). Comparison of post hoc tests for unequal variance. *International Journal of New Technologies in Science and Engineering*, *2*(5), 22–33. <https://www.ijntse.com/upload/1447070311130.pdf>
15. *Which alpha value should I use?* (n.d.). Scribbr. Retrieved April 20, 2024, from <https://www.scribbr.com/frequently-asked-questions/which-alpha-value-to-use/>
16. Weisstein, E. W. (2004). Bonferroni correction. *Https://Mathworld. Wolfram. Com/*. <http://mathworld.wolfram.com/BonferroniCorrection.html>
17. Turner, H. M., Iii, & Bernard, R. M. (2006). Calculating and Synthesizing Effect Sizes. *Contemporary Issues in Communication Science and Disorders*, *33*(Spring), 42–55. <https://doi.org/10.1044/cicsd_33_S_42>
18. Tim. (2016, October 16). *Hedges’ g: Definition, Formula*. Statistics How To. <https://www.statisticshowto.com/hedges-g/>
19. Tim. (2016, September 18). *Holm-Bonferroni Method: Step by Step*. Statistics How To. <https://www.statisticshowto.com/holm-bonferroni-method/>