

סיכום פסיכולוגיה ניסויית - שחף קולן

מבוא לשיטות מחקר

מהו מחקר מדעי?

השיטה המדעית מציגה דרך פורמלית לחקור את העולם סביבנו. מחקר מדעי משפיע על הרבה מאוד היבטים של החיים שלנו.

ההבדל בין מדע לפסאודו מדע

מדע	פסאודו מדע
השערות ניתנות להפרכה (אפשר למצוא ראיות שמראות על כך שההשערה אינה נכונה)	השערות אד-הוק, תחומי ידע שיכולים להסביר כל נתון בעזרת תיאוריה נתונה. זהו לא תחום שיש בו השערה שניתנת להפרכה
מסתמך על נתונים אובייקטיביים שאפשר לאמת, על תוצאות שנתונות לביקורת עמיתים, ועל מחקרים ברי שחזור.	נשען על אנקדוטות סובייקטיביות, התוצאות של הפסאודו מדע לרוב מתחמקות מביקורת עמיתים והמחקרים אינם ברי שחזור.

מחקר מדעי – צבירת ידע

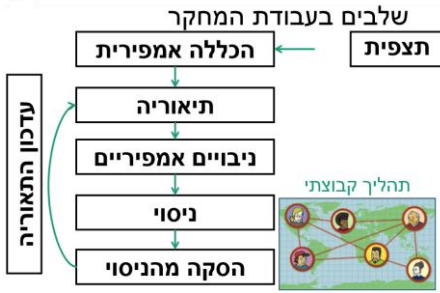
המטרה העיקרית של המחקר המדעי היא צבירת ידע. יש כמה שיטות עיקריות שהמדע משתמש בהן כדי לצבור ידע:

- ניתוח תיאורטי – חשיבה על נושאי המחקר, ניסיון לנתח אותו מכוח העליון בלבד, או לבנות מודלים חישוביים שיכולים לנתח את המציאות. הכל נשאר בעולם התיאוריה.
- מחקר תיאורי – למדוד את המציאות, לתעד את מה שקיים, אבל בלי לחקור את הקשר בין תופעות. אפשר לכלול פה שיטות כגון תצפיות, ראיונות, ומחקרים איכותניים
- מחקר מתאמי – חוקרים את הקשרים בין משתנים עבור תופעות מדידות. המבחן הסטטיסטי הבסיסי שמשמש את המחקר המתאמי הוא מתאם (קורלציה). יש גם מערכי מחקר מתאמיים שמאפשרים גם הסקה סיבתית, ולא רק מתאמית ברמת ביטחון מסוימת.
- מחקר ניסויי – חוקר גם את הקשר בין משתנים, וגם מאפשר מחקר של הקשר הסיבתי בין משתנים. הסיבה שמחקר ניסויי הוא היחיד שמאפשר להסיק על סיבתיות בצורה מוחלטת, היא בגלל שזה המחקר היחיד שבו לחוקרת יש שליטה על סביבת המחקר.

מחקר יישומי לעומת מחקר בסיסי

- בעולם נהוג לחלק את המחקר לשני סוגים – מחקר יישומי, ומחקר בסיסי. בפועל לא חייבת להיות הפרדה בין השניים ביום יום.
- מחקר יישומי – ממוקד בפתרון בעיות אקולוגיות ופרקטיות. לדוגמה – מחקר הבודק השפעתו של טיפול התנהגותי על ילדים עם ADHD.
- מחקר בסיסי – מטרתו להעצים את ההבנה המדעית, ללא מטרת פרקטיות. ממצאיו לרוב משולבים במחקר יישומי. למשל, מחקר שבודק איך חזרה בע"פ על ההחלטה להיות מרוכז משפיעה על יכולת הריכוז במטלת סטרופ.

שלבים בעבודת המחקר



1. תצפית – בד"כ התיאוריה נולדת מתצפיות. למשל, בתיכון התלמידה יכולה להביט סביבה ולנסות להבין את ההתנהגות של הסובבים אותה.
2. הכללות אמפיריות - מהתצפיות של התלמידה ניתן להכליל מספר הכללות אמפיריות שמתאימות לתצפיות. למשל – כשהן מתעניינות בבנים, בנות מציגות עצמן כמתקשות בהבנת החומר בכיתה.
3. הגיית תיאוריה שתסביר את ההכללות האמפיריות – פסוקים תיאורטיים. ניסוח הפסוקים התיאורטיים כרוך בהמשגה של הכללים האמפיריים שמצאתי לכדי כללים תיאורטיים, לפעמים גם בהוספה של משתנים שבכלל לא נמצאו בהכללה האמפירית. למשל – כשהם מעוניינים למצוא חן בעיני בני המין השני, בני נוער מנסים להראות שהם מתאימים לאידיאל הנחשק.
4. גזירת ניבויים אמפיריים – אחרי שבנינו תיאוריה אנחנו צריכים לבחון אותה, אז אנחנו רוצים לגזור ניבויים שאפשר למצוא במחקר. למשל – בנות יטעו יותר בתרגילי חשבון כשבן נחשק נמצא בכיתה.
5. ביצוע ניסוי – בחינת התיאוריה על ידי ניסויים. למשל – שליטה בהימצאות בן נחשק בכיתה, ומדידת הצלחה בתרגילי חשבון על הלוח.
6. עדכון התיאוריה עקב תוצאות הניסוי – לפעמים התוצאות לא מאוששות את ההשערה אבל הן עוזרות לעדכן את התיאוריה. בעקבות עדכון התיאוריה ניתן לגזור פסוקים תיאורטיים חדשים, מהם לגזור ניבויים אמפיריים חדשים ולהמשיך לבחון את התיאוריה. כך הניסוי עוזר לצבור ידע להבנת הקשרים הסיבתיים בעולמנו.

שאלת המחקר

בניסוי בד"כ יש שאלת מחקר, שמתארת את סוג ודפוס הקשר (או סוג ההשפעה) שהחוקרת מצפה למצוא.

- לרוב בשאלת המחקר משערים השערה כיוונית. לדוגמא: השאלה "כיצד משפיעים חלומות בהקיץ על דיכאון?" מעידה על השפעה כיוונית של "חלומות בהקיץ מגבירים דיכאון" או "חלומות בהקיץ משפיעים על דיכאון".
- ניתן גם לנסח שאלות מחקר שבהן אין השערה כיוונית. לדוגמא: "האם חלומות בהקיץ משפיעים על דיכאון?"

בהשערת מחקר כיוונית קל יותר להפריך את H_0 . במקרה כזה, אם הניסוי כשל בלהפריך אותה אז האמון שלנו בנכונות השערת ה-0 יהיה חזק יותר.

חשוב לזכור - לעולם לא ניתן להוכיח את השערת ה-0. בסטטיסטיקה בייסיאנית אפשר לאמוד את עוצמת הראיות לטובת השערת האפס, אבל זה הכי קרוב שאפשר.

מקובל להניח שהשערת מחקר טובה מבוססת על תיאוריה :

לדוגמא: אם ישנה תיאוריה לפיה לאנשים יש נטייה טבעית לחשוב מחשבות מדכאות, ועל כן עליהם להימנע מחלימה בהקיץ, אז השערה סבירה היא שהגברת חלומות בהקיץ (למשל בניסוי), תגביר דיכאון.

אם ישנן שתי תיאוריות מתחרות שמייצרות השערות מנוגדות, הרי שזה הגיוני לחלוטין שלא תהיה השערת מחקר אחת לחוקרים. מה שחשוב הוא שלפני הניסוי החוקרים יהיו יכולים להסביר מה תהיה השערת המחקר לפי התיאוריות הרלוונטיות לניסוי. ניסוי יכול להיות מועיל גם אם החוקרים לא ביססו את שאלת המחקר על התיאוריה. אבל, תיאוריות עוזרות לחוקרים להשקיע את המשאבים שלהם בשאלת מחקר שיש הגיון רב יותר בבדיקה שלהם. לפעמים שאלת המחקר חשובה רק כי תשובה עליה תעזור להבחין בין תיאוריות שונות, כלומר הקשר הסיבתי לא בהכרח חשוב, אלא חשובה ההשלכה של הממצא על נכונות התיאוריות, שייצרו ניבויים לגבי שאלת המחקר.

שאלת המחקר מכתובה את ההתנהלות של המחקר:

- קובעת מה הם המשתנים שייבחנו במחקר
- מהי האוכלוסייה שתיבחן ואיך יבחר המדגם
- מה תהיה שיטת המחקר
- אילו ניתוחים סטטיסטיים יתבצעו

לדוגמא: שאל המחקר "האם חלומות בהקיץ מגבירים דיכאון"? קובעת שהמשתנים שנבחן הן חלימה בהקיץ ודיכאון. האוכלוסייה היא כנראה האוכלוסייה הכללית, שיטת המחקר צריכה להיות כנראה מחקר ניסויי (כי ההשערה היא על קשר סיבתי), והניתוח הסטטיסטי ייקבע לפי המשתנים והרמות שלהם, והשאלה הספציפית שמעניינת אותנו.

משתנה

משתנה הוא כל סט של אירועים שיכולים להיות לו ערכים שונים.

לדוגמא: במחקר שבוחן את השפעת מינון התרופה על ציון במבחן ישנם שני משתנים:

1. מינון התרופה – משתנה בעל ערכים מוגדרים (למשל 10, 20, 30 מ"ג)
2. ציון במבחן הזיכרון – משתנה בעל ערכים מ-1 עד 100 (למשל)

במחקר מהסוג הזה המשתנה הנמדד הוא משתנה מקרי, שאנחנו לא יודעים מראש מה הערך שנקבל בו.

סוגי משתנים: במערך מחקר אנחנו בוחנים את הקשר בין משתנים, כשאנחנו מניחים הרבה פעמים השפעה של משתנים מסוימים על אחרים.

- מקובל לכנות את המשתנים החשודים כמשפיעים משתנים בלתי תלויים – **Variables Independent**
- המשתנים החשודים כמושפעים משתנים תלויים **Variables Dependent**

יש המון הגיון בשימוש בכינויים הללו, כאשר מערך המחקר הוא מערך ניסויי, אשר בו אנחנו בוחנים את ההשפעה של משתנים על משתנים אחרים, ואז קל להבחין בין משתנים תלויים לבין בלתי תלויים. אבל זה לא אומר שבמחקר מתאמי אנחנו לא יכולים להשתמש בכינויים האלו.

לדוגמא: אם התיאוריה שלנו מניחה ש IQ גבוה גורם להנאה פחותה מהחיים, אפשר להפריך את התיאוריה גם במחקר מתאמי, אם מוצאים קשר חיובי בין שני המשתנים ולא שלילי כמו שהתיאוריה ניבאה, במקרה כזה החוקרת עדיין יכולה לכנות את IQ כמשתנה בלתי תלוי ואת ההנאה כמשתנה תלוי, זאת למרות שאי אפשר להסיק סיבתיות במערך המתאמי.

הבחנה נוספת בין סוגי משתנים:

- כאשר משתנה נמדד אנחנו נרצה לקרוא לו משתנה ייחוס (**attribute**) זה משתנה שאין לנו שליטה ישירה עליו בניסוי, אנחנו רק מודדים אותו. כל משתנה שלא נערכת עליו מניפולציה ישירה הוא משתנה ייחוס.
- משתנה שאנחנו שולטים ברמה שלו נקרא משתנה מתופעל/ניסויי/מופעל (**active**) למשל אם אנחנו משווים שיטת הוראה מקוונת לשיטת הוראה לא מקוונת - שיטת ההוראה היא משתנה בלתי תלוי מתופעל. המשתנה התלוי יכול להיות כל מיני משתנים (למשל- כסף).

יש משתנים שקשה מאוד לתפעל (למשל- מין, גיל, מגדר, מעמד סוציו אקונומי) לחוקרת יהיה קשה מאוד להשפיע על הדברים האלו, ולכן הם בד"כ משתני ייחוס. יש מקרים שבהם אותו משתנה יכול להופיע במקרים מסוימים כמשתנה ייחוס ובמקרים אחרים כמשתנה מתופעל (למשל- אפשר למדוד חרדה כמשתנה ייחוס, אבל אפשר גם לעשות מניפולציה בחרדה וליצור לו דרגות שונות ואז זה יהיה משתנה מתופעל).

הבחנה נוספת בין משתנים היא ההבחנה בין משתנה תיאורטי לבין משתנה אופרציונלי.

- משתנה תיאורטי – הוא המשתנה הכללי שמעניין אותנו
- משתנה אופרציונלי – הוא המשתנה שמש אותו בניסוי כדי ללמוד על המשתנה התיאורטי.

במקור השיח עוסק בהגדרה תיאורטית מול הגדרה אופרציונלית -

- הגדרה תיאורטית – (נומינלית) הגדרת המשתנה באמצעות מילים, בהתאם לתיאוריה (הגדרה מילונית).
- הגדרה אופרציונלית – הגדרה שמסבירה כיצד למדוד או לתפעל את המשתנה, איך לקבוע את הרמה של המשתנה, מהי האופרציה שאנחנו עושים כדי למדוד את המשתנה.

חשוב להבחין בין משתנים תיאורטיים לאופרציונליים כי שאלת המחקר היא לרוב תיאורטית (מעניין אותנו איך חרדה משפיעה על ריכוז, ולא איך תפעול מסוים של חרדה משפיע על תפעול מסוים של ריכוז). אבל, כאשר יש תוצאות ואנחנו לומדות כחוקרות על הקשרים בין המשתנים כדי שבין מה התוצאות, הקהילה המדעית הרבה פעמים מתחילה להתעסק במדדים והמניפולציות עצמן, לעיתים על חשבון קישור בין המשתנים התיאורטיים.

מערכי מחקר

באופן כללי המטרה של מדע היא למצוא פונקציות או גלגות שונות שיטתית בנתונים. אנחנו רוצים לדעת איך השתנות במשתנה תלוי קשורה לשינויים במשתנה בלתי תלוי מסוים. אנחנו מחפשים קשרים בין משתנים. שתי משפחות מערכי המחקר הללו שונות בכמה דברים, ובראש ובראשונה בדרך בה מושגת השונות במשתנה הבלתי תלוי.

· מערך מחקר מתאמי – כל המשתנים נמדדים, אין שום שליטה בשונות המשתנים. אין אפשרות להציב נבדקים בקבוצות באופן אקראי.

· מערך מחקר ניסויי – לפחות אחד מהמשתנים הבלתי תלויים מתופעל. יש שליטה בשונות המשתנה, אנחנו יוצרים שונות על ידי יצירת מניפולציה במשתנה הבלתי תלוי. במערך הניסויי החוקרים מציבה את הנבדקים בקבוצות הניסוי ויכולה להחליט לאיזה קבוצה ילך כל נבדק.

ההבדל הזה מוביל לכך שאין לנו אפשרות לתת הסבר סיבתי לתוצאות במערך מתאמי. מה שמאפשר לנו לעסוק בסיבתיות בניסויים הוא השליטה הישירה שלנו במשתנה הבלתי תלוי. אנחנו יודעים שסביר להניח שההבדל היחיד בין הנבדקים בשתי הקבוצות הוא במשתנה הבלתי תלוי. השיטה המקובלת להצבה של נבדקים בתנאי ניסוי שונים במערך הניסויי הוא הקצאה מקרית, הפעולה הזו ששקולה להטלת מטבע נקראת רנדומיזציה. ההקצאה המקרית נועדה לסלק השפעות שיטתיות של משתנים נוספים שאנחנו לא מעוניינים בהם. בד"כ הקצאה מקרית מאפשרת היסקים סיבתיים. כלומר, אם נמצא הבדל בין הקבוצות בניסוי בערך המשתנה התלוי – נדע בביטחון גבוה יחסית, שהמשתנה הבלתי תלוי אחראי לכך, כי הוא המבדיל בין הקבוצות.

דוגמה להבדל בין מחקר ניסויי למחקר מתאמי -

השערת המחקר – אשמה על התנהגות לא מוסרית מעלה את הרצון לשטוף ידיים

מחקר מתאמי – מדידת שני המשתנים: החוקרת יכולה לשאול כמה אשמה הרגשנו היום וכמה פעמים שטפנו ידיים היום. נוכל להסיק מקשר בין המשתנים שאולי באמת קיים ביניהם קשר סיבתי, אבל לא נדע איזה סוג של קשר סיבתי.

אם יש לנו מתאם בין שני משתנים יש לנו לפחות ארבעה הסברים אפשריים למתאם כזה:

1. משתנה א' גורם למשתנה ב' – תחושת אשמה גורמת לשטיפת ידיים מוגברת
2. משתנה ב' גורם למשתנה א' – שטיפת ידיים מוגברת גורמת לתחושת אשמה
3. משתנה א' ומשתנה ב' משפיעים זה על זה במעגל הדדי – שטיפת ידיים ואשמה גורמות זו לזו במעגל חיזוק הדדי
4. משתנה ג' משפיע על משתנה א' וב' – משתנה שלישי מגביר שטיפת ידיים ותחושות אשמה

במערך מתאמי אין לנו דרך להבדיל בין ההסברים האפשריים, אין לנו יכולת להסיק סיבתיות.

מחקר ניסויי – החוקרת צריכה לתפעל את המשתנה שע"פ ההשערה משפיע על המשתנה האחר (תפעול תחושת האשמה), ולגרום לכך שרמת המשתנה הבלתי תלוי תהיה שונה בכל קבוצה בניסוי, זאת תוך שמירה על הקצאה אקראית לקבוצת האשמה מול קבוצת הביקורת.

במערך כזה אנחנו מקצים באופן אקראי נבדקים לאחד משני התנאים (באמצעות הטלת מטבע), ואם נמצא הבדל בין שתי הקבוצות (למשל מה שרואים בתמונה) אנחנו יכולים להסיק מה הייתה הסיבה שגרמה להבדל הזה- מה היה ההבדל בין שתי הקבוצות שהשפיע על השוני בין הקבוצות. למעשה, היתרון המרכזי במערך ניסויי הוא ששלטנו בסדר הזמנים ובכך שללנו את האפשרות שמשנתנה ב' (התלוי) יכול להוביל למשתנה א' (הבלתי תלוי), כיוון שאנחנו שולטים על מה שבא קודם. חשוב לציין שיש אקסיומה שנכונה להרבה מהתיאוריות ברוב המדעים, והיא שאירוע יכול להשפיע רק על אירוע שקרה אחריו, ולא על אירוע שהתרחש לפניו.

נשארנו עם אפשרות שמשנתנה א' (בלתי תלוי) השפיע על משנתנה ב' (תלוי), או שיש משנתנה שלישי שהשפיע על שניהם.

האם יתכן שמשנתנה חיצוני השפיע על הטלת המטבע? זה כנראה לא סביר. אבל אפשר לחשוב על מצב בו אדם שיחד את מטיל המטבע ודאג לכך שמשנתנים ששוטפים ידיים יוקצו יותר לקבוצת הניסוי, וההסבר החלופי הזה גורם לכך שהשוחד הוא המשנתנה החיצוני שגרם לתוצאות שקיבלנו. אבל זה לא תרחיש מאוד סביר. לכן, לא סביר שמשנתנה חיצוני השפיע על הטלת המטבע וזו הסיבה שהקצאה מקרית לקבוצות ותפעול המשנתנה הב' ת בכל קבוצה, זו דרך בטוחה למדי לדעת שההבדל בין שתי קבוצות נבע מההבדל במשתנה הבלתי תלוי.

אם אנחנו יודעים שההבדל היחיד בין הקבוצות הוא המשנתנה הב' וחילקנו את הקבוצות בהקצאה אקראית, אז אנחנו יודעים להסיק שהמשנתנה הב' גרם להבדלים בין הקבוצות במשתנה התלוי.

חשוב להזכיר שיש סיכוי שהדבר קרה במקרה, כלומר הטלת המטבע במקרה הקצתה אנשים ששוטפים ידיים במשך זמן ארוך יותר לקבוצת הניסוי. במקרה כזה אנחנו נמצא הבדל בין שתי הקבוצות אבל הוא לא ייגרם מהמשנתנה הבלתי תלוי, אלא מההקצאה לקבוצות. הסיכוי שדבר כזה יקרה הוא ערך ה P שחישבנו בחישוב הסטטיסטי, במקרה הזה החישוב המתאים הוא מבחן T כי אנחנו משווים בין שתי קבוצות במשתנה מקרי רציף, אנחנו מניחים התפלגות נורמלית במשתנה "משך שטיפת ידיים". בד"כ בפסיכולוגיה אנחנו מתייחסים להבדל כהבדל לא מקרי (שלא נבע מההקצאה לקבוצות), אם המבחן הסטטיסטי שבו השתמשנו מראה שהסיכוי לקבל הבדל כזה באקראיות קטן מ 5% ($P < 0.05$), נעריך שההבדל לא נבע משונות מקרית בין נטיות שטיפת ידיים של נבדקים, אלא משונות שיטתית שנגרמה מהמשנתנה הבלתי תלוי.

חשוב להזכיר - ערך ה P הוא הסיכוי לקבל את התוצאות שקיבלנו בהנחה שהשערת ה 0- נכונה. ערך ה P הוא לא הסיכוי שהשערת ה 0 נכונה. חוקרים בתחומים אחרים משתמשים בחישוב הסיכוי הזה כדי לנסות להפריך את השערת ה 0 - אם ההשערה היחידה שנותרה לאחר הפרכת השערת ה 0 - היא השערת המחקר, אפשר להסיק שהיא ההשערה הנכונה.

לסיכום, בניסוי מבוקר היטב, ההבדל בין הקבוצות נוסע מאחת משתי סיבות:

1. מקריות – כשהסיכוי קטן מ 5% לא נקבל את השערת המקריות, ונניח שהסיבה להבדל הינה -
2. הבדל בערכי המשנתנה הבלתי תלוי של כל קבוצה – כלומר, נסיק שיש השפעה של המשנתנה הב' על המשנתנה התלוי ושיש קשר סיבתי ביניהם.

המבחן הסטטיסטי מנסה לספק הערכה מה ההסתברות שנקבל תוצאה כזו בהשערת מקריות.

הסקה בייסיאנית לא משפיעה על הצורך בהקצאה מקרית להסקת סיבתיות. בהסקה בייסיאנית שבד"כ משתמשים בה בפסיכולוגיה, אנחנו מחשבים הסתברות לקבל את התוצאות שקיבלנו, בהינתן השערת המחקר, ואת הסיכוי לקבל את התוצאות שקיבלנו בהינתן כל שאר התוצאות למרות השערת המחקר, ואז משווים בין שני הסיכויים האלו. הבייס פקטור הוא היחס בין שני הסיכויים האלו. אבל, אם לא הייתה לנו הקצאה מקרית, הסיבה לתוצאות עשויה להיות הגורם שהשפיע על ההקצאה, וזה יסתור הנחות יסוד שהשתמשנו בהם לחישוב שתי ההתפלגויות (הסיכויים). ואז הבדלים במשתנה התלוי ברמות השונות של המשנתנה הבלתי תלוי יכולים לנבוע רק ממקריות או מהמניפולציה. בלי הקצאה מקרית הבייס פקטור שחישבנו עדיין יהיה נכון, אבל הסיבה להבדלים בין הקבוצות עשויה להיות ההקצאה, ולא רמות המשנתנה הבלתי תלוי במשתנה התלוי. לכן, גם אם אנחנו משתמשים בסטטיסטיקה בייסיאנית הקצאה מקרית עדיין חשובה.

לסיכום, מחקר ניסויי הוא הכלי הכי נפוץ בפסיכולוגיה. הסקה סיבתית נותנת לנו אפשרות להבין, לנבא ולהשפיע על המציאות. אי אפשר להשפיע על משהו בלי לדעת מה הקשר הסיבתי. לכן, כל אדם שרוצה לשנות משהו בחייהם חייב להניח קשר סיבתי. והדרך הטובה ביותר להסיק קשר סיבתי הוא ניסוי. לכולם כדאי לדעת להעריך את תוקף המסקנות הסיבתיות של עצמם.

תוקף = איכות הניסוי, מידת האמון שאנחנו יכולים לתת במסקנותיו.

תוקף המחקר

נשאל את השאלה מהו ניסוי טוב - ניסוי שהמסקנות ממנו תקפות. תקפות היא תכונה רציפה ורב היבטית - ניסויים הם תקפים יותר או פחות, לא בצורה דיכוטומית. התוקף לא קשור למובהקות או החוזק הסטטיסטית וכן העקרונות נכונות לכל מחקר מדעי ניסויי ולא רק פסיכולוגיה.

נמדוד תקפות באמצעות לפחות 4 מובנים שונים:

- תוקף המסקנה הסטטיסטית – עד כמה ודאית המסקנה הסטטיסטית שהתקבלה לגבי הקשר בין המשתנים?
- תוקף פנימי – עד כמה ודאי שבהשוואה בין תנאי הניסוי, המשתנה התלוי נתון אך ורק להשפעת המשתנה הבלתי תלוי?
- תוקף מבנה – עד כמה משקפים המשתנים האופרציונליים את המשתנים התיאורטיים?
- תוקף חיצוני – עד כמה ניתן להכליל ממסקנות המחקר לגבי כלל האוכלוסייה ולגבי מצבים רבים במציאות?

תוקף המסקנה הסטטיסטית:

עד כמה ודאית המסקנה הסטטיסטית שהתקבלה לגבי הקשר בין המשתנים. ניסוי בעל תוקף מושלם מסוג זה, זהו ניסוי שבו לא התקבלו טעויות החלטה בניתוח הסטטיסטי - אפעם לא נדע אם זה קרה במאה אחוז. לכן ניסוי בעל תוקף מסקנה סטטיסטית גבוה הוא ניסוי בו אנחנו בוטחים במסקנה הסטטיסטית אם יש או אין קשר בין המשתנים. אנחנו בודקים את הקשר בין המשתנים, או במילים אחרות - כמה מהשונות במשתנה התלוי ניתן להסביר באמצעות המשתנה הבלתי תלוי. אם יש להם שונות משותפת רבה, ניתן לומר כי יש קשר בין המשתנים. החוקרים עורכים מבחנים סטטיסטים כדי לבדוק תחת השערת האפס מהי ההסתברות למצוא את הקשר שהתקבל.

למשל - במבחן T אנחנו בודקים אם כאשר המשתנה הבלתי תלוי משתנה ברמתו, אז גם הערך של המשתנה התלוי משתנה. אם שניהם משתנים ביחד אז יש שונות משותפת, ואם הסבירות שנקבל קשר כזה למרות שאין קשר כזה באוכלוסייה היא נמוכה למדי, נסיק שאכן קיים קשר כזה במציאות.

כשאנחנו בוחנים את תוקף המסקנה הסטטיסטית אנחנו בוחנים אם מסקנת החוקרים תקפה בעינינו, למעשה אנחנו בוחנים עד כמה המסקנה שלהם - יש קשר/אין קשר/אי אפשר לדעת סבירה בעינינו.

ישנן 2 טעויות החלטה:

1. טעות מסוג ראשון – הוסק שיש קשר כשאין.
2. טעות מסוג שני – הוסק שאין קשר כשיש.

תוקף זה מתייחס לסבירות המסקנה אך ורק מתוצאות המבחנים הסטטיסטים שנערכו (וגם ביסיאנים).

4 גורמים אפשריים שיכולים לגרום לטעויות אלה - איומים על תוקף המסקנה הסטטיסטית:

אלו הם כלים שמראים לנו כי תוקף המסקנה הסטטיסטית מפוקפק, הקשר לא מייצג את המציאות.

1. טעויות בקביעת מקדם הטעות (מקדם הטעות הוא אלפא):

(טעות מסוג 1) - ניפוח אלפא

קובעים אלפא גדול מידי - להגדיל את הסיכוי להגיע למסקנה שקיים קשר למרות שהוא לא קיים. אין ביטחון שלא נעשתה

טעות מסוג ראשון.

לדוגמא - אם נבצע ניסוי על הרבה משתנים שונים שלכל אחד מהם הסיכוי למצוא קשר הוא 5% ונמצא קשר מובהק אחד, הסיכוי למצוא קשר היה גדול מ 5%.

סיבות לניפוח האלפא:

1. כתוצאה משימוש במספר רב של מדדים - ולא רק משתנה תלוי ובלתי תלוי. הסיכוי שבלפחות אחד מהמדדים שנבדקים תהיה תוצאה מובהקת גבוהה מ 5%.
לדוגמא - נניח שאנחנו בודקים את השפעת החשיפה אל פרופוסורית למדעים על תפיסות סטריאוטיפיות לגבי נשים ומדע. קבוצת הניסוי למדה אצל הפרופוסורית, קבוצת הביקורת למדה אצל פרופוסור גבר. בניסוי נשתמש בכמה מדדים תלויים: דיווח של נבדקים עד כמה בממוצע הם מאמינים שנשים טובות במדעים כמו גברים, עד כמה הם מאמינים שחלק מהנשים יכולות להיות טובות במדעים כמו גברים, עד כמה הם מאמינים שנשים וגברים נולדים עם פוטנציאל דומה במדעים, עד כמה הם יעודדו את ביתם ללמוד מדעים, ולתת לנבדקים לבחור בין מועמד גבר למועמדת אישה במשרה מדעית. נניח שמצאנו הבדל מובהק בין הקבוצות רק במדד אחד שנבדק. במצב כזה לא נוכל להגיד שיש הבדל מובהק, הסיכוי למצוא את ההבדל היחיד הזה הוא לא 5% כי היו 5 מדדים שונים שמדדנו. הסיכוי שבלפחות אחד מהמדדים שבדקנו תהיה תוצאה מובהקת גבוהה מ 5% (כי זה לא כמו לבדוק שכל המדדים מקבלים תוצאה מובהקת בו זמנית). הסיכוי לא למצוא הבדל בהנחה שאכן אין הבדל הוא $4/5$ בחזקת 5= כלומר. 67% לכן הסיכוי למצוא לפחות הבדל מובהק 1 הוא 33% אם יש לנו 5 מדדים שונים והאלפא שלנו היא 5% בכלום.
2. כתוצאה מריבוי השוואות - כשמבצעים יותר מהשוואה אחת הסיכוי לטעות מסוג ראשון גדל.

פתרונות לניפוח אלפא:

1. רפליקציה של הממצא - הדרך הכי טובה לוודא שהתוצאה היא לא מקרית היא פשוט לערוך את הניסוי שוב. כשיש לנו השערה ברורה לכיוון התוצאות, אם נמצא תוצאה מובהקת נדע שהסיכוי שתוצאה זו מקרית קטן מ 5%. בשחזור הניסוי אפילו לא צריך שהתוצאה תהיה $p < 0.05$ ומספיק ערך P יותר גבוה כי אנחנו לוקחים בחשבון את החישוב לתוצאה מובהקת בשני הניסויים, אבל לרוב אנחנו לא מדקדקים בפרטים האלה, פשוט מחפשים $P = 0.05$ בשם הפשטות. לעיתים קשה מאוד לשחזר את הניסוי (מפאת משאבים או כסף). במקרה זה נפרסם את הותצאות ונדווח על סכנת ניפוח אלפא.
2. רישום מראש pre-registration - פרסום מראש של ההשערות, איזה קשר אנחנו מצפים שיוביל לתוצאה מובהקת. נבדוק קשר אחד בלבד, אם שיערנו שהשוואה כלשהי תוביל לתוצאה מובהקת ואכן קיבלנו תוצאה מובהקת בהשערה הספציפית הזאת, אנחנו לא מנפחים את האלפא כי שיערנו ספציפית על ההשוואה הזאת.
3. תיקון אלפא/P - קביעת מקדם טעות אחר לכל אחד מהמבחנים כדי שהסכום שלהם יהיה באמת 5%.

(טעות מסוג 2) - קביעת ערך אלפא מחמיר מידי

לקבוע אלפא נמוך מדי - עשוי להוביל לפספוס ההבדלים בין הקבוצות, כלומר פספוס קשר קיים. קביעת ערך האלפא צריכה לאזן בין שתי סוגי הטעויות.

2. עוצמה סטטיסטית נמוכה: יכולה להוביל לטעות מסוג 2

עוצמה סטטיסטית היא ההסתברות שהמבחן ידחה את השערת האפס כשההשערה האלטרנטיבית נכונה, כלומר לא למצוא אפקט כאשר יש אפקט.

גורמים לעוצמה סטטיסטית נמוכה מידי:

1. המדגם קטן מידי - אם נבחן את ההשערה על מספר קטן מדי של אנשים ייתכן ולא נמצא אפקט שכן קיים באוכלוסייה, לא תמיד ניתן לדעת בוודאות מהו גודל מדגם מספק אבל יש תוכנות שעוזרות בכך.
2. איכות המדדים התלויים נמוכה - אם המדדים עצמם לא רגישים למשתנה התלוי, אם המדדים לא מוצלחים כל כך. לדוגמא אם נרצה לבחון את ההשפעה של הרצאה מקוונת לעומת הרצאה פרונטלית על קורס מסוים, אם המבחן בסוף הקורס לא מוצלח הוא לא יצליח אפקט לטובת אף אחד מהמשתנים הבלתי תלויים. קשור לתוקף ומהימנות המדידה.
3. איכות המניפולציה נמוכה/אינה קבועה - ניתן לבדוק את איכות המניפולציה, אם היא לא הייתה מדויקת ולא תפעלנו את המשתנה התלוי כמו שהיינו רוצים לתפעל אנו עלולים לפספס אפקט שקיים.

פתרון: תכנון הניסוי בצורה טובה יותר, נשתמש ביכולת שלנו בתור חוקרים לשלוט בשונות בצורה ישירה.

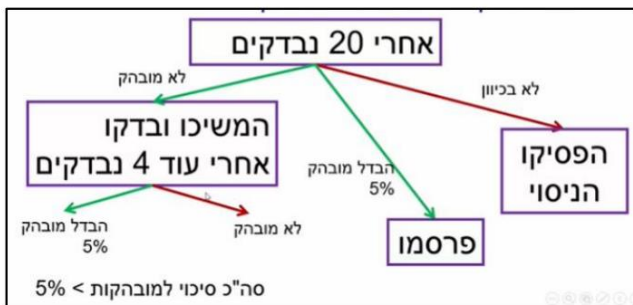
3. שימוש במבחנים סטטיסטיים לא מתאימים: יכול להוביל לטעויות מסוג 1 ו 2

אם נשתמש במשתנה איכותי בתור כמותי, או אם נבצע מבחן שאנו לא עומדים בהנחות שלו. ההתפלגות עליה אנו מסתמכים בביצוע המבחן הסטטיסטי יכולה להיות שגויה. כשנרצה לדעת מהי ההסתברות שתוצאה כלשהי תתקבל במקרה תחת בדיקת מובהקות אנחנו בהכרח תלויים בצורת ההתפלגות שהמבחן הסטטיסטי מניח. לדוגמה - אם נשתמש במבחן T למדגמים תלויים כאשר המדגמים שלנו הם בלתי תלויים, אנחנו עשויים לבצע יותר טעויות מסוג 1 כי ההתפלגות שמבחן T מניח היא אחרת.

4. הטיות בניתוח הנתונים:

יש החלטות שונות שיש לקבל במהלך ניתוח הנתונים (למשל האם נבדק מסוים ביצע את המטלה שלו כמו שצריך), אבל פעם כזו יש חשש להטיה בניתוח - חשש שהחוקר יקבל החלטה שמושפעת מהמוטיבציה שלה והטיות פנימיות לקבל תוצאה מסוימת. הטיות אלה יכולות להתרחש באופן לא מכוון. דרגות החופש של החוקר = החלטות שהחוקר מקבל בשלב ניתוח הנתונים שמגבירים את הסיכוי שיתקבלו התוצאות המבוקשות. כל החלטה כזו שנעשית זו הזדמנות נוספת למציאת תוצאות מובהקות, קל להפוך תוצאה כמעט מובהקת למובהקת. תופעה זו של הסתכלות על תוצאות והתאמת הנתונים לאחר צפייה בתוצאות נקראת p-hacking / fishing - שינוי של ערך Pn לאחר שראינו אותו. המטרה של חוקרים צריכה להיות לתאר את המצב בעולם בצורה הטובה ביותר, מטרה זו מתנגשת עם השאיפה של החוקר לאשש את התיאוריה ולפרסם מאמרים. הרבה ד"רנלים מחפשים לפרסם מאמרים עם תוצאות מובהקות, שמהווה תמריץ לחוקר.

דוגמה - החלטה להוסיף נבדקים לאחר שביצענו בדיקה - חוקרים רבים בודקים לאחר 20 נבדקים ומבינים אם יש אפקט, האפקט לא מובהק אבל בכיוון או האפקט מובהק. אם נמצא שאנחנו בכיוון, במידה ואין אפקט סביר שלא נמצא אותו גם בהוספת נבדקים, אבל אם מוספים נבדקים לאט לאט ובכל פעם בודקים את תוצאות המובהקות הסיכוי הופך גדול מ 5% למציאת אפקט בטעות. שיטה שמביאה לניפוח אלפא. אם נעצור רק כתלות במובהקות התוצאות זה לא ישקף את מה שקורה במציאות וזה פוגע בתוקף המסקנה הסטטיסטית.



חוב החוקרים משערים על קיום קשר ולכן דרגות החופש של החוקר לרוב מובילות לטעות מסוג ראשון - גילוי קשר שלא קיים. לפעמים החוקרים משערים על חוסר קשר ואז דרגות החופש של החוקר יכולות לגרום לטעות מסוג שני - אי גילוי קשר שקיים בעולם.

הפתרון: נבדקים הם משאב יקר, לכן כן סביר לבדוק במהלך הניסוי אבל בצורה מוגבלת ומבוקרת. יש צורך להגדיר מראש מתי עוצרים מראש לבדוק במהלך הניסוי, וכן לחשוף בניסויים שמפורסמים עם שיטת עבודה של בדיקה באמצע. בנוסף יש לדווח איך קבענו את מספר הנבדקים בניסוי, גם לאחר הוספה באמצע.

משבר הרפליקציה:

החל מתחילת 2010 התבהר כי הרבה ממצאים במדעי הפסיכולוגיה אינם ניתנים לשחזור בלתי תלוי (למשל במחקר רפליקציה שיטתית, מתוך 100 ניסויים שנדגמו לשחזור, רק 36 שוחזרו). המשבר מתייחס לתוקף המסקנה הסטטיסטית - לא ניתן לסמוך על התוצאות הסטטיסטיות הללו. כנראה נובע מהטיות בניתוח נתונים, ניפוח אלפא, harking = הצגת מחקרים גישושיים בתור מחקרים אישושיים. המשבר גורם לבעיה מערכתית, חוסר אמונה גם במחקרים שלא ביצעו בעיות מסוג זה.

פתרונות להגדלת תוקף המסקנה הסטטיסטית בתור חוקר:

- צמצום דרגות החופש של החוקר - pre-registration. לדווח מראש על: גודל המדגם, כיצד נחליט על הסרת נבדקים, מה יהיו המבחנים הסטטיסטיים בהם נשתמש... עלינו להתחייב מראש להנחיות המבחן, ולהיצמד אליהן. עם זאת, ניתן לערוך שינויים בתוכנית הניסוי שתכננו, במידה ומצאנו דרכים ליעל או לשפר את הניסוי, אך צריך תמיד לדווח בשקיפות על כל

- שינוי .
- קביעת מקדם טעות מתאים/ תיקון ל P
- רפליקציה

פתרונות מערכתיים לבעיה:

- העלאת המודעות לבעיה (למשל כחלק מקורס לתואר ראשון) תגרום לחוקרים אתיים לפעול לשיפור תוקף המסקנה הסטטיסטית.
- הגברת שקיפות באיסוף הנתונים וניתוחם, כתבי עת מתחילים לחייב על pre-registration.
- Reports Registered לבקש סקירת עמיתים (לבקש ביקורת מחוקרים אחרים) לפני שאנחנו אוספים את הנתונים בניסוי.
- הגברת חשיבותה של רפליקציה, עידוד ביצוע רפליקציה, חיוב סטודנטים לבצע רפליקציה כחלק מהתואר שלהם.
- changing the incentive structure - שינוי מבנה התמריצים, שחוקרים לא יתוגמלו רק על עבודה שמגלה אפקטים חדשים וממצאים חדשים, וכמה שיותר מאמרים, אלא יתוגמלו גם על ביצוע שחזורים, ועל ביצוע מדע מהימן ובר שיחזור.

תוקף פנימי:

תוקף פנימי עוסק בשאלה עד כמה ודאי שבהשוואה בין תנאי הניסוי, המשתנה התלוי נתון אך ורק להשפעת המשתנה הבלתי תלוי. במילים אחרות, האם יש השפעה של משתנים נוספים על המשתנה הבלתי תלוי. מערך ניסויי הוא הכלי היחיד שמאפשר הסקת סיבתיות, מכיוון שבמערכי מחקר אחרים קשה הרבה יותר להגיד שהמשתנה התלוי נתון רק להשפעה של המשתנה הבלתי תלוי. אבל גם במערך ניסויי יכולות להתעורר בעיות בתוקף הפנימי. באופן כללי, כשאנחנו מדברים על תוקף פנימי הכוונה היא לתוקף המסקנה שהמשתנה הבלתי תלוי האופרציונלי הוא היחיד שיכול לגרום בשינויים ברמתו של המשתנה התלוי האופרציונלי, בלי קשר לשאלה כמה המשתנים האלה משקפים את המשתנים התיאורטיים. השאלה עד כמה הם משקפים את המשתנים התיאורטיים קשורה לתוקף מבנה. יש לשים לב שהתוקף הפנימי הוא גם בלתי תלוי בתוצאות המחקר, מחקר בעל תוקף פנימי טוב יהיה כזה בין אם ההשערה שלנו אוששה ובין אם לא.

גורם המאיים על תוקף פנימי:

משתנה מתערב - משתנים לא מתוכננים ולא נשלטים הקיימים בסביבת הניסוי ועשויים להשפיע על תוצאותיו. יש לשלוט בהם כדי למנוע את השפעתם הפוטנציאלית על המשתנה התלוי, בנוסף להשפעתו של המשתנה הבלתי תלוי. דוגמה - אנחנו רוצים לבדוק את השפעת קומוניזם על העוני. אולם, בכל המדינות שבהם נוסה הקומוניזם היה משטר סטאליטרי. לכן, לא נדע אם הקומוניזם או המשטר הסטאליטרי אחראיים למצב הכלכלי.

תוקף מבנה:

תוקף מבנה עוסק בשאלה עד כמה משקפים המשתנים האופרציונליים את המשתנים התיאורטיים. עד כמה מה שנותנים לנבדקים בפועל משקף את מה שרצינו לבדוק מבחינה תיאורטית - עד כמה הניסוי בנוי בצורה שמתאימה למטרות. משתנים תיאורטיים – המשתנים שמעניינים אותנו, שאנחנו רוצים לבחון, שאין לנו בכלל גישה ישירה אליהם. משתנים אופרציונליים – המשתנים שמייצגים בניסוי את המשתנים התיאורטיים. המשתנים שאנחנו מודדים או מתפעלים. השאלה היא האם המשתנה האופרציונלי מתאים למשתנה התיאורטי -

לדוגמא: אם אנחנו מתעניינים בשאלה של מה קורה כשאנחנו מערערים על הדימוי העצמי של הבנאדם? אז אנחנו צריכים למצוא דרך לערער את הדימוי העצמי של הנבדקים שלנו.

- משתנה תיאורטי - איום על דימוי עצמי
 - משתנה אופרציונלי - פידבק שלילי
- צריך להבין שערעור על דימוי עצמי יכול ללבוש הרבה צורות, יתרה מזאת יכול להיות שיש אנשים שפידבק שלילי בכלל לא יערער להם את הדימוי העצמי. כל האפשרויות הללו הם איום על תוקף המבנה. דוגמא נוספת:

- משתנה תיאורטי - אינטליגנציה
- משתנה אופרציונלי – מבחן IQ
- האם IQ מודד את האינטליגנציה?

תוקף מבנה רלוונטי גם למשתנים בלתי תלויים וגם למשתנים תלויים. אחת השאלות העיקריות שאנחנו צריכים לשאול את עצמנו בעריכת ניסוי היא מה הקשר בין המשתנים שבהם אנחנו משתמשים בניסוי לבין המשתנים שאנחנו מתעניינים בהם. לפעמים הדרך הטובה ביותר להקטין את הפער בין משתנה תיאורטי למשתנה אופרציונלי היא לשאול שאלות על המשתנה האופרציונלי עצמו (למשל- אפשר לטעון שאין כזה דבר אינטליגנציה אלא רק הצלחה במבחני IQ וזה מה שמעניין.). בפסיכולוגיה ובתחומים אחרים, לא נצליח להגיע למצב שמשתנה משקף רק דבר אחד.

גורמים המשפיעים על תוקף מבנה:

- מיצוי – האם המשתנה האופרציונלי משקף את כלל ההיבטים של המשתנה התיאורטי. האם תוצאות הניסוי נכונות גם עבור אופרציה אחרת של המשתנה התיאורטי.
 - לדוגמה: מתן פידבק שלילי לא ממצה את כל המשמעות של ערעור דימוי עצמי. יכול להיות שיש היבטים תיאורטיים שונים בערך התיאורטי של ערעור ערך עצמי, ייתכן והמשתנה האופרציונלי משקף רק היבט אחד של המשתנה התיאורטי של ערעור הערך העצמי.
- בלבדיות – יתכן שמהמשתנה האופרציונלי טומן בחובו היבטים שאינם רלוונטיים למשתנה התיאורטי.
 - לדוגמה: פידבק שלילי בניסוי יכול להטיל גם חשד שהנסיינית עובדת עליו. זה לא חלק מערעור על הביטחון העצמי.
 - זה יוצר מצב של סכנת קונפאונד – מצב שבו יש משתנה תיאורטי חלופי שמיוצג באופרציונליזציה, כלומר המשתנה התיאורטי החלופי מוצמד למשתנה התיאורטי שהתכוונו אליו כי שניהם הופעלו על ידי המניפולציה. זה פוגע ביכולת להסיק מסקנות מהניסוי.
- יש מי שמכנה את קונפאונד המאיים על תוקף המבחן – *operational confounding* להבדיל מ- *procedural confounding*: (מפורט שוב בהמשך)
- *operational confounding* – מצב שבו המניפולציה שעשינו מפעילה משתנה תיאורטי אחר מזה שהתכוונו לתפעל במקור, או כאשר המדד התלוי מודד משתנה תיאורטי שונה ממה שהתכוונו למדוד.
- *procedural confounding* – מצב שבו הניסוי מתוכנן באופן לקוי שגורם לכך שיש משתנה מתערב שרמותיו מוצמדות לרמות של המשתנה הבלתי תלוי. זוהי בעיה בתוקף הפנימי של הניסוי. כך, המשתנה אמנם מוגדר טוב אופרציונלית אבל האופן בו מתבצע הניסוי גורם לכך שרמות המשתנה הוצמדו למשתנה מתערב.
- כדאי להבין שההבחנה הזאת תלויה בהגדרה שלנו למשתנה אופרציונלי. הגדרת משתנה אופרציונלי, מוגבלת במה שהחוקר התכוון לתפעל אופרציונלית) למשל לתת פידבק חיובי או שלילי לנבדקים בניסוי) והיא לא כוללת מאפיינים אחרים שהבחינו בין תנאי הניסוי בשל טעות בתכנון או הרצת הניסוי. למשל- דברים כמו פידבק חיובי תמיד יינתן על ידי נסיינית ופידבק שלילי תמיד על ידי נסיין (פידבק שונה ניתן על ידי אנשים שונים = קונפאונד פרוצדורלי). ההבחנה מתבססת על השלב בתכנון הניסוי בו הטעות מתרחשת.
- הגדרת המשתנה התיאורטי – חשוב להגדיר את המשתנה התיאורטי בהירות כדי שנוכל לייצר לו משתנה אופרציונלי בקלות, וגם כדי שחוקרים אחרים יוכלו לייצר משתני אופרציונליים אחרים בניסויים אחרים אם ירצו לבחון את אותו משתנה תיאורטי. כשמדובר במשתנה תיאורטי רחב ואבסטרקטי אז המחקר בוחר חלקים קטנים של המשתנה התיאורטי אשר מרכיבים ביחד את המשתנה השלם. לעיתים לא פשוט להגדיר את משתנה תיאורטי באופן טוב יותר. בהקשר של תוקף מבנה, הגדרה טובה היא הגדרה שקל להחליט אילו משתנים אופרציונליים מתאימים לה.
 - לדוגמה: יכול להיות שיש כמה היבטים לערעור ערך עצמי (פקפוק בחשיבותי, פקפוק במוסריותי, פקפוק ביכולותי) אבל צריך לאחד אותם בהגדרה של המשתנה התיאורטי. ניתן לחשוב על משתנים אופרציונליים שונים המתאימים להיבטים שונים של המשתנה התיאורטי.

תוקף חיצוני:

עוסק בשאלה עד כמה ניתן להכליל ממסקנות המחקר לגבי כלל האוכלוסייה ולגבי מצבים רבים במציאות. אם לא ניתן להכליל מתוצאות הניסוי לגבי המציאות הערך של הניסוי קטן.

לדוגמא: נניח שתפעלנו חרדה באופן נשלט בחדר שקט. זה לא בהכרח יכול להיות מוכלל למצבים רבים בעולם.

קריטריונים שעוזרים בקביעת תוקף חיצוני:

- עד כמה הניסוי דומה לתופעה הנחקרת (אקולוגיות)
- הדמיון בין המדגם לאוכלוסייה
- דגימה - איך אנחנו בוחרים נבדקים למדגם שלנו .
- דגימה מקרית - בד"כ בניסויים נרצה לדגום דגימה מקרית.
- דגימה מייצגת - הרכב המדגם זהה להרכב האוכלוסייה.
- דגימה מקרית היא לא בהכרח דגימה מייצגת. כדי שהיא תהיה מייצגת צריך לדגום מקרית מתוך אוכלוסיות מייצגות.
- קיימים מגוון סוגים שונים של דגימה, נתמקד בדגימה מקרית פשוטה.

השפעת הדגימה על התוקף החיצוני:

איום על יכולת ההכללה של הניסוי נובע מהדרך שבה דגמנו את הנבדקים . לדוגמא: אם לא הצענו לנבדקים שום תמורה להשתתפותם, ייתכן שהנבדקים שלנו שונים מהאדם הממוצע כיוון שהם הסכימו להשתתף בניסוי למרות שהם לא מקבלים על כך כלום. זוהי דוגמה לבעיית דגימה- כשהרכב הנבדקים שבחרנו לניסוי לא נדגם מקרית ולא מייצג את הרכב האוכלוסייה.

יש מקרים שבהם חוקרים בכוונה דוגמים מתוך אוכלוסייה מצומצמת, זה קורה כשאנחנו רוצים להקטין את השונות שנובעת מהבדלים בין אישיים, לפעמים עדיף מדגם הומוגני על מנת לגרום לשאר השונות להיווצר מהמניפולציה או מטעויות מדידה, זה יקל עלינו במציאת אפקטים. חוקרים יעשו את זה בעיקר כשמניחים שאין סיבה להניח שהשונות באוכלוסייה תשפיע על תוצאות הניסוי. לעיתים קרובות פסיכולוגים לא מתעכבים על שאלת הדגימה המייצגת או המקרית, כי הם מניחים שמה שהם חוקרים נכון לכל בני האדם, ומניחים שנבדקים בניסוי ייצגו את האוכלוסייה לגביה רוצים להסיק (למשל סטודנטים לפסיכולוגיה). חשוב לציין שמחקרי המשך ידגמו מאוכלוסיות אחרות. יש לשים לב שקיימים הבדלים תרבותיים שיכולים להשפיע בכל זאת.

תוקף המחקר לא תלוי בתוצאות המחקר - אם נמצא או לא נמצא קשר. מחקר עם תוקף מסקנה סטטיסטית, פנימי, מבנה, ותוקף חיצוני טובים יהיה כזה בין אם ההשערה אוששה או בין אם לא. על כל הדברים האלו צריך לחשוב כשמתכננים ניסוי אבל גם כשקוראים מאמרים.

מערכי מחקר שגויים

מערך מחקר – תוכנית פעולה לבדיקת קשרים בין משתנים. תוכנית פעולה זו מאפשרת בחינת שאלות המחקר.

ניסוי – מחקר שבו החוקר קובע את רמת המשתנה (או המשתנים) הבלתי תלויים של כל נבדק.

רק בניסוי אפשר להסיק שהקשר בין המשתנים הוא סיבתי, כיוון שרק בניסוי יש שליטה על הרמות השונות של המשתנה הבלתי תלוי. מה שגרם להבדלים בין הרמות השונות היא החלטת החוקר, ולכן לא אפשרי שהמשתנה התלוי או משתנה שלישי השפיעו על המשתנה הבלתי תלוי. ההצהרה הזאת נכונה רק אם המשתנה התלוי או משתנה אחר כלשהו לא השפיעו על החלטת החוקר. לכן, הקצאה מקרית חשובה. כמובן ששום דבר לא ודאי, יכול להיות שיש לנו טעות מסוג ראשון (קבענו קשר מובהק למרות שאין באמת) וכאן אנחנו חוזרים לתוקף המסקנה הסטטיסטית.

כדי לבדוק קשר סיבתי נרצה לערוך מערך ניסוי, ונרצה שהמערך יהיה תקף, שיאפשר לנו את המסקנה הסיבית. כעת אנחנו נדבר על מערכי מחקר בהם אין תוקף גבוה כי הם שגויים. היות והמחקרים אינם תקפים הם חשופים לאיומים רבים על התוקף הפנימי.

כדי שתהיה לנו שפה מוסכמת לדבר על מערכי מחקר, נגדיר כמה מושגים וסמלים שיוכלו לייצג את מערך המחקר באופן גרפי:

- X = טיפול ניסויי (מניפולציה) היעדר X זה היעדר מניפולציה
- O = תהליך של תצפית או מדידה. זו הנקודה בזמן בה אנחנו מודדים את המשתנים התלויים שמעניינים אותנו
- כל שורה מייצגת רצף שקרה לאותם נבדקים/ אותה קבוצה
- R = הקצאה מקרית לקבוצה
- כל עמודה מייצגת פעולות שקרו באותו הזמן
- ההצגה משמאל לימין

מערכי מחקר שגויים (להסקה סיבתית) –

1. ניתוח מקרה – **case study** קרתה מניפולציה, ואנחנו מודדים את הקבוצה שלנו לאחר המניפולציה:

X O

דוגמא - מדידת הציונים של סטודנטים לאחר שלמדו אצל מרצה חדש.

הבעיה המרכזית של המערך הזה היא שאין עם מה להשוות את המדידה, אין למשתנה הבלתי תלוי רמות, יש לו רק רמה אחת. יש פה השערה מובלעת שמשווים את המדידה למה שהיה יכול לקרות (מדידה היפותטית). אז מדוגמא כזאת אי אפשר להסיק שום דבר לגבי הסטודנטים או לגבי רמת ההוראה של המרצה, אלא אם כן תהיה לנו יכולת להשוות את הציונים במבחן הזה למשהו הזה. בד"כ לא אוהבים להתעכב על המערך הזה כי הוא כמעט לא באמת קיים, אנחנו תמיד משווים עם משהו. תחום אחד שכן יש בו הרבה שימוש בניתוחי מקרה הוא ניוירו פסיכולוגיה (אנשים עם פגיעות ראש) אנחנו לא יכולים למדוד מה היה קורה אם האדם לא היה נפגע בראשו. בעזרת תיאורי מקרה כאלה אפשר ללמוד בצורה איכותנית על תופעה, ולבנות לה תיאורים טובים, ומכך לנסח תיאוריות, ומהם לגזור פרדיקציות ולהמשיך במעגל בדיקת ההשערות שאנחנו לומדים בקורס.

2. מערך לפני - אחרי עם קבוצה בודדת - **one group pretest-posttest** יש מדידה אחת, מניפולציה, ומדידה שניה לאחר המניפולציה:

O1 X O2

לדוגמא - מדידת חומרת המחלה לפני ואחרי שלוקחים תרופה שלושה ימים. האם אפשר להסיק שהתרופה גרמה לשינוי שמצאתי בין המדידה הראשונה למדידה השנייה? (וגם להפך - אם לא חל שיפור האם אפשר להגיד שהטיפול לא עזר.)? אפשר לחשוב על כמה סיבות מדוע אי אפשר להסיק קשר סיבתי בין הטיפול לבין השינוי או אי השינוי.

איומים על מחקר מסוג זה:

2 הראשונים מתרחשים עקב חלוף הזמן בין המדידות -

1. הסטוריה
2. בשילה
3. השפעת המדידה
4. השפעת המכשור
5. גרסיה לממוצע

נפרט על כל אחד:

1. היסטוריה - משהו **חיצוני** קרה בין המדידות שגרם להבדל בניהן.
2. בשילה - משהו **פנימי** שקרה בין המדידות. הדבר הנוסף שקרה בין שתי המדידות קרה בתוך הנבדקים.

לדוגמא - הגוף למד באופן טבעי לייצר נוגדנים מתאימים וחיסל את המחלה.

שני האיומים האלו הם איומים על התוקף הפנימי – יכול להיות שהמשתנה התלוי (ההבדל בין המדידות) נתון להשפעה ממשנתנים אחרים ולא רק להשפעה של המשתנה הבלתי תלוי (הטיפול).

3. המדידה השפיעה - יכול להיות שהטיפול הראשונה למדידה השניה נובע בכלל מהמדידה הראשונה - יכול להיות שהמדידה הראשונה אימנה אותנו ולכן השתפרנו. כלומר - המדידה עצמה היא משתנה אלטרנטיבי לטיפול, ויכול להיות שזה מה שהשפיע על המשתנה התלוי. איום זה מאיים גם כן על התוקף הפנימי – מאיים על הטענה שהטיפול הוא מה שהשפיע על

מדדים תגובתיים כנגד לא תגובתיים: איום של מדידה מתרחש בעיקר על מדדים תגובתיים.

- מדדים תגובתיים - מדדים שיש סיכוי גבוה שתהליך המדידה שלהם משפיע על הנבדקים ולא רק רושם באופן פסיבי את ההתנהגות, מדדים תגובתיים יכולים להיות הסבר חלופי לתוצאות הניסוי.
- מדדים לא תגובתיים - מדדים שלא משפיעים על המדד. למשל, מדדים בהם הנבדק לא מודע למדידה/התוצאה במדידה ככה שלא יוכל להיות מושפע מהתוצאה.

דוגמאות למשתנים תיאורטיים שניתן למדוד במדדים תגובתיים ולא תגובתיים - בתוכנית הרזיה, עליה על משקל גלוי היא מדד תגובתי, ניתן למדוד גם באופן שסמוי לנבדקים כמדד לא תגובתי. גמילה מאלכוהול, ניתן לשאול באופן מפורש כמדד תגובתי, ניתן לספור בקבוקי אלכוהול מבלי שהנגמל יודע.

4. מכשור - משהו השתנה במכשור שאנחנו עושים בו שימוש כדי למדוד את השינוי, יכול להיות שהשינוי הוא שאחראי על ההבדל בין המדידות ולא הטיפול. זה כולל גם שופטים שונים בין המדידות.
- לדוגמא - למשל מד משקל בתוכנית דיאטה התקלקל והוחלף במד משקל אחר.
- איום זה מאיים גם הוא על התוקף הפנימי כי האיום הזה מציע הסבר אלטרנטיבי למסקנה שהמשתנה הבלתי תלוי השפיע על המשתנה התלוי.

5. רגרסיה לממוצע -

לדוגמא - נרצה למדוד את רמת הדיכאון של תלמידי פסיכולוגיה ולטפל ב 20 התלמידים המדוכאים ביותר. על מנת לעשות את הטיפול האפקטיבי ביותר נבחר את 20 התלמידים עם רמות הדיכאון הגבוהות ביותר ונפעיל עליהם טיפול כלשהו נגד דיכאון. לאחר מכן נמדוד את הדיכאון שלהם שוב. ניתן לנבא שהדיכאון יהיה נמוך יותר בלי קשר לטיפול ואפילו אם לא היה טיפול, וזה נובע מתופעה שנקראת רגרסיה לממוצע. רגרסיה לממוצע נובעת מכך שבכל מדידה יש טעות. כל מדידה/ציון מורכבת מכמה רכיבים:

- ציון אמיתי - הציון המדויק שנבדק במדד, מה שקורה באמת.
- ציון נמדד - הציון שנרשם במדידה של הנבדק.
- טעות מדידה - ההפרש בין הציון האמיתי לציון הנמדד. בכל מדידה יש טעות, במקרה והמדד שלנו טוב, טעות המדידה היא אקראית, לפעמים היא מעל הציון האמיתי, ולפעמים היא מתחת לציון האמיתי. בהנחה של טעויות מקריות בערך 50% מהמקרים יהיו כלפי מעלה ו50% כלפי מטה. הטעות היא מקרית ולכן היא תתקוף את הנבדקים בלי קשר לציון האמיתי שלהם.

מכאן נובע - במדידה אחת, ככל שהציון קיצוני יותר, סביר להניח שהוא כתוצאה מטעות. אם זה ציון מאוד גבוה במדידה יחידה, סביר שיש לנו טעות כלפי מעלה (ולהפך כלפי מטה). במילים אחרות, רוב הסיכויים שהציונים הקיצוניים הם כתוצאה מטעות לכיוון הקיצוני, והציון האמיתי קרוב יותר לממוצע ממה שאנחנו רואים לפנינו. דרך אחרת להבין את זה היא אם יש לי ציון אמיתי 60 ולאדם אחר גם ציון אמיתי 60 זה מאיתנו שהמבחן יעשה לגביו טעות כלפי מטה גבוהה יותר, פשוט יקבל ציון נמוך יותר. מכך נובע שבמדידה נוספת, סביר שהציון יהיה פחות קיצוני. זוהי תופעה סטטיסטית שגורמת לכך שהציון במבחן לא משקף את הציון האמיתי, זה קיים תמיד בפסיכולוגיה, כי אין מבחן מושלם. זה נובע מכך שהמתאם בין שתי מדידות חוזרות של אותו מבחן הוא אינו מושלם.

פתרונות אפשריים לאיום הרגרסיה לממוצע :

1. מדידה נוספת לאחר בחירת הנבדקים: לעשות מדידת baseline ולפי לבחור את הנבדקים. לאחר מכן לקחת מדידה נוספת בתור המדידה הראשונה ולפיה לעשות את הניתוח הסטטיסטי של המחקר. בשל הרגרסיה לממוצע הציונים במדידה הראשונה יהיו קרובים יותר לממוצע מהציונים שבגללם בחרנו את הנבדקים לקבוצה בבייס ליין. בגלל שאנחנו מתעניינים בהבדלים בין מדידה 1 למדידה 2, שיכולים לנבוע מהמניפולציה, אין סכנה שההבדלים בין ממוצע הקבוצה במדידה הראשונה לשנייה יהיו בגלל הרגרסיה לממוצע (כי היא השפיעה כבר על ההבדל בין מדידה 1 לבייס ליין). אבל זה לא הופך את המערך הזה למערך תקין כי הוא עדין סובל מהרבה איומים אחרים שהוצגו מקודם.
2. קבוצת ביקורת - קבוצה שלא עוברת מניפולציה, זה הפתרון לרוב הבעיות שהוצגו עד כה והוא יפורט בהמשך.

גרגסיה לא מסכנת כל קבוצה בעלת ציונים קיצוניים, היא מסכנת כל קבוצה שנבחרה על סמך ציונים קיצוניים כאשר הם מהווים את המדידה הראשונה שאליה משווים בסוף. אם הקבוצה נבחרה על סמך מדדים אחרים ולאחר הבחירה נמצא כי הקבוצה מתאפיינת על ידי ממוצע ציונים קיצוני (נמוך מאוד או גבוה מאוד) יש חשש קטן יותר לתופעה של גרגסיה במדידה החוזרת, כי הגורם לציונים הקיצוניים לא היה מקרי, אלא כנראה נבע מהמדד שעל פיו בחרנו את הקבוצה מלכתחילה. גם כאן הרגסיה לממוצע מאיימת על התוקף הפנימי – כי גרגסיה לממוצע עלולה לגרום לשינוי במשתנה התלוי שלא נובע מהרמות השונות של המשתנה הבלתי תלוי.

3. מערך אחרי ללא הקצאה מקרית – מערך בעל שתי קבוצות. מערך בו לא מקצים נבדקים באופן מקרי. יש לנו טיפול ויש שתי קבוצות – אחת שעברה טיפול והשניה לא עברה טיפול, ולכן יש חוסר רנדומיזציה.

X O1

O2

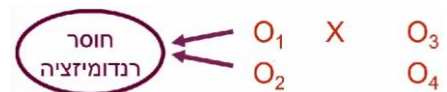
לדוגמה - נרצה לבדוק תרופה נגד דיכאון מול פלסבו. נקצה את 30 הנבדקים שהגיעו ראשונים לקבוצת הטיפול (קיבלו תרופה נגד דיכאון) ו 30 הנבדקים שהגיעו אחרונים הוקצו לקבוצת הביקורת (קבלו פלסבו).

האיום הכי משמעותי למערך כזה נקרא – ברירה: selection המשמעות היא שהסיבה שיש הבדל בין שתי הקבוצות היא לא שאחת מהן עברה טיפול והשנייה לא, אלא שהדרך שבה הקבוצות הללו נבחרו היא שאחראית על ההבדל במשתנה התלוי בין שתי הקבוצות. למשל בדוגמה יכול להיות שנבדקים שהגיעו ראשונים הם פשוט בעלי מוטיבציה גבוהה יותר מהנבדקים שהגיעו מאוחר יותר, ולכן אחרי הטיפול הם יראו רמות דיכאון נמוכות יותר מהקבוצה שקיבלה פלסבו וזה לא קשור לטיפול, אלא כי מלכתחילה הייתה להם יותר מוטיבציה לעשות שינוי.

זה איום על התוקף הפנימי כי הוא מציע שהמשתנה הבלתי תלוי (הטיפול) לא השפיע על המשתנה התלוי, אלא משתנה שלישי (ההקצאה לקבוצות) הוא שהשפיע על המשתנה התלוי.

3 א. מערך לפני אחרי ללא הקצאה מקרית בעל מדידה נוספת – העיקרון הוא שיש קבוצה ואנחנו מודדים אותה לפני המניפולציה ואחרי המניפולציה, ויש קבוצה נוספת שאנחנו מודדים לפני ואחרי ללא העברת מניפולציה. גם כאן יש חוסר רנדומיזציה.

ניתן היה לחשוב שמדידה נוספת לפני במערך ללא הקצאה מקרית תאפשר לנו לדעת מראש אם היה הבדל בין הקבוצות במשתנה התלוי. היתרון של מערך זה הוא שניתן לשלול הבדלים הקיימים בין הקבוצות מראש בהסתכלות על הבייס ליין. (זהו דוגמא למערך מעורב - מערך מחקר בו יש גורם בלתי תלוי תוך נבדקי, וגורם בלתי תלוי בין נבדקי.)



צריך להבין שאפילו אם אנחנו מגלים הבדלים בהתחלה, זה לא ממש תיקן את הניסוי, זה רק מנע מאיתנו להסיק מסקנות פיזיות. שנית, אפילו אם מצאנו שאין הבדל בהתחלה, יכול להיות שההבדלים בין הקבוצות יבואו לידי ביטוי בהמשך (למשל נבדקים בקבוצה אחת מגיבים שונה למדידה מאשר הנבדקים בקבוצה אחרת).

איומים:

- תוקף פנימי - מתקיימים איומים של אינטראקציה בין ברירה לבין בשילה/היסטוריה/גרגסיה/מדידה/מכשור.
- תוקף חיצוני - אינטראקציה בין ברירה לטיפול, יכול להיות שאם היינו בוחרים קבוצה אחרת, הטיפול לא היה משפיע.

לדוגמה - נראה את האיום הנובע מאינטראקציה בין ברירה להיסטוריה. נבחן את השפעת המלגה על סטודנטים. האוניברסיטה מסרבת לתת מלגה באופן אקראי. החוקרים היו צריכים למצוא קריטריון למלגה, והתמקדו בהכנסה של ההורים, כך שסטודנטים שהוריהם הם בעלי הכנסה נמוכה קיבלו מלגה ואלו עם הורים עשירים לא קיבלו מלגה. לקחו קבוצה שהוריהם "עניים" וקבוצה שהוריהם "עשירים" ומדדו את הציונים שלהם בסמסטר א, אחרי שנתנו לעניים מלגה. לאחר מכן מדדו את הציונים בסמסטר ב ומצאו שאכן הקבוצה שקיבלה מלגה, השיפור שלה בציונים היו גבוה יותר מאשר אלו שלא קיבלו מלגה. המסקנה שאפשר היה להסיק היא שמלגה מובילה לציון גבוה יותר.

אי אפשר באמת להסיק שהמסקנה הזאת נכונה: בגלל שמדדנו את השינוי היחסי בציונים, והתחשבנו בציוני הבייס ליין, אנחנו

לא חשופים לאיום של ברירה. עם זאת, ייתכן וקרה משהו בין המדידות שהשפיע על שיפור הציונים של אלו שקיבלו מלגה - יכול להיות שיש עוד תוכניות של האוניברסיטה שעוזרות לתלמידים מרקע סוציו אקונומי נמוך, כמו למשל חונכות של סטודנטים חדשים ע"י סטודנטים לתארים מתקדמים.

כך נוצרת אינטראקציה בין ברירה להיסטוריה, ההקצאה הלא מקרית לקבוצות היא שגרמה לכך שלקבוצות שלנו יהיו היסטוריות שונות, אם היינו מקצים את הנבדקים באופן אקראי לא היינו נתקלים באיום זה. כי החונכות היא משהו שקרה מתחילת הניסוי ועד שמדדנו, והוא עשוי להסביר את הממצא במקום המשתנה הבלתי תלוי.

בנוסף גם ניתן לקרוא לאיום גם איום של אינטראקציה בין ברירה לבין היסטוריה מקומית, כי ההיסטוריה קרתה לקבוצה אחת בלבד.

קונפאונד **confound** – שתי הגדרות שונות:

קונפאונד זה משתנה מתערב, יש שני סוגים שונים של משתנים מתערבים -

- קונפאונד אופרציונלי - כשיש סכנה שמשתנה אופרציונלי מייצג משתנה תיאורטי נוסף (חוסר בלבדיות), שזו סכנה לתוקף המבנה.

- קונפאונד פרודורלי - כשגורם אחר השפיע על המשתנה התלוי. במהלך הפרוצדורה במחקר עצמו משהו אחר התערב. כל מה שדיברנו בהרצאה הזאת הם מקרים שבהם מערך הניסוי פגום והוא זה שגורם לכך שהתוצאות יכולות להיות מוסברות באמצעות משתנה תיאורטי אחר. אלה הם איומים על התוקף הפנימי.

כל האיומים שעסקנו בהם כעת מאיימים על התוקף הפנימי כי הם מציעים משתנה אלטרנטיבי בנוסף למשתנה הבלתי תלוי, בחשש שהמשתנה האלטרנטיבי הוא שגורם לתוצאות שקיבלנו, כיוון שהוא הוצמד למשתנה הבלתי תלוי.

בהינתן הסבר חלופי יכול להיות שהממצא שלנו הוא ארטיפקט (=טעות) – ממצא שנובע מהסבר חלופי שמאיים על התוקף

הפנימי. זה הופך את הממצאים ללא מעניינים כי הם נובעים מסיבות ידועות. כשמבקר של מחקר מציע שהממצא הוא

ארטיפקט הוא בעצם אומר שהניסוי לא כל כך מעניין, כי המשתנה שהשפיע על המשתנה התלוי הוא טריוויאלי. במקרים כאלה

צריך להעדיף את ההסבר החלופי כי הוא פרסימוני יותר - פשוט יותר. במדע אנחנו מעדיפים הסברים פשוטים יותר מאשר

מסובכים בהנחה שהם מסבירים את התופעה.

רגרסיה לממוצע היא ארטיפקט סטטיסטי, בשילה היא ארטיפקט ניסויי - שנקרא גם קונפאונד.

לפעמים לעומת זאת, במקרים בהם יש קונפאונד אופרציונלי (הסברים חלופיים שאומרים שהמשתנה האופרציונלי מייצג

משתנה תיאורטי אחר מזה שהחוקר התכוון לו), בניגוד לקונפאונד פרודורלי, לא אומרים שהמחקר חסר תועלת, פשוט יכול

להיות שמחקר מלמד אותנו משהו שונה ממה שהחוקר חושב, ומוסיף לנו ידע על משתנה תיאורטי אלטרנטיבי. במקרים של

איומים על המשתנה האופרציונלי פחות מקובל להשתמש במושג ארטיפקט, לפעמים זה הסבר חלופי מעניין. זה איום על תוקף

המבנה, כלומר הניסוי פחות טוב כי קשה להסיק עליו מסקנות תיאורטיות.

נהוג להשתמש בשני המובנים של המונח קונפאונד (לפעמים גם ביחד), ההשלכה בפועל היא דומה - ייתכן ולא המשתנה

שרצינו הוא זה שמשפיע על המשתנה התלוי. הדרך להתמודד עם שני סוגי הקונפאונד היא שונה:

בקונפאונד אופרציונלי - נרצה להגדיר משתנה אופרציונלי שלא יוכל להעיד על המשתנה התיאורטי החלופי.

בקונפאונד פרודורלי - נרצה לוודא שאין משתנים שיכולים להשפיע על המדידה שנובעים מהמערך.

מערכי מחקר תקינים

מערכי מחקר שכן מאפשרים לנו להסיק מסקנה סיבתית -

1. מערך לפני אחרי עם קבוצת ביקורת - pretest-posttest control group design

R	O ₁	X	O ₃
R	O ₂		O ₄

מערך שדומה למערך לפני אחרי עם קבוצת ביקורת, רק שיש רנדומיזציה. מציגים בצורה גרפית רק שתי קבוצות (אחת

שמפעילים עליה מניפולציה (X) ואחת שלא) בשביל הפשטות, ניתן לבצע גם ליותר משתי קבוצות, או יותר מטיפול אחד. מפעילים מניפולציה על קבוצה אחת, בהשוואה לקבוצה שלא מפעילים עליה את המניפולציה.

ההקצאה של הנבדקים לקבוצות היא מקרית (R) ההקצאה לא קשורה לשלב שבו אנחנו בוחרים נבדקים באוכלוסיית היעד, זו הדגימה. ההקצאה זו השאלה "איך אנחנו מקצים את הנבדקים שדגמנו לקבוצות השונות בניסוי?" ואנחנו רוצים שזו תהיה הקצאה מקרית, בלי קשר לסדר, שעה ביום וכו'.

דוגמה - אנחנו רוצים לבדוק אם חיבור על אלברט איינשטיין משנה את רמת ה IQ של אנשים. מדדנו IQ של שתי הקבוצות, ביקשנו מקבוצה אחת לכתוב חיבור על אלברט איינשטיין (מניפולציה-X), מדדנו שוב את ה IQ של שתי הקבוצות. ישנם מערכים מורכבים יותר המבוססים על מערך זה, זהו מערך מומלץ כיוון שהוא מפחית הרבה מאוד איומים על התוקף הפנימי של הניסוי.

איומים שלא קיימים על המערך:

- בשילה - לא קיים איום בשילה כי יש לצפות לבשילה זהה בשתי הקבוצות. איום בשילה משמעותו תהליך פנימי שקבוצה אחת עוברת והשניה לא. למשל אם הנבדקים בקבוצה שלנו הם דיירי בית אבות, אז יכול להיות שתהיה הידרדרות של ה IQ בין מדידה אחת לשנייה, אבל אין סיבה להניח שתהיה הידרדרות יותר גדולה בקבוצה אחת מאשר בקבוצה אחרת, לכן לא סביר לייחס הבדלים שנמצא לבשילה.
- היסטוריה - לא קיים איום של היסטוריה, כיוון שההקצאה המקרית דואגת שלא יהיו הבדלים בהיסטוריה של הנבדקים. אבל, ייתכן איום של היסטוריה מקומית (מתרחש כשהקבוצות נפרדות זו מזו), יכול להיות שהיסטוריה של אחת לא תהיה זהה להיסטוריה של הקבוצה האחרת. למשל, אם נמדוד שני בתי אבות, אז במקרה הזה יכול להיות שקרה משהו בבית אבות אחד שלא קרה בשני, למשל אם הייתה שריפה. כדי שמשוהו כזה לא יקרה אנחנו מקפידים שהמדידות יהיו בתנאים וזמנים זהים ככל האפשר. בנוסף, אנחנו מודדים את הנבדקים בקבוצות קטנות ובסדר אקראי. למשל לא נרצה להקצות את כל הבית אבות הראשון לקבוצה הראשונה ואת כל הבית אבות השני לקבוצה השנייה, נרצה לחלק אותם באופן אקראי, ואז נבטיח שאין קשר בין התנאי שאליו משתייך הנבדק (ניסוי מול ביקורת) לבין ההקשר החיצוני- עיתוי או מקום פיזי בו הוא נמצא, ואז אפשר להניח שההיסטוריה של הנבדקים בשתי הקבוצות לא שונה.
- מדידה - לא קיים איום מדידה (איום שבו המדידה היא זו שגרמה להבדלים בין שתי הקבוצות), החלוקה האקראית מפחיתה לאפס את איום המדידה, כך שהמדידה לא יכולה להסביר הבדלים בין הקבוצות. יכול להיות איום של אינטראקציה בין טיפול למדידה, כך שהטיפול הוא זה שמשנה את ההשפעה של המדידה, משהו בטיפול משפיע על איך שנבדקים ממלאים את המדידה העוקבת, אבל זה מקרה נדיר.
- רגרסיה - לא קיים איום רגרסיה, גם אם דגמנו נבדקים קיצוניים, הם הוקצו אקראית לקבוצות והרי הרגרסיה אמורה להתרחש בשתי הקבוצות באותה מידה. למשל גם אם בחרנו ניסוי של נבדקים עם IQ נמוך במיוחד, אפשר לשער שתהיה עלייה בהצלחה במבחן השני בגלל הרגרסיה לממוצע, אבל העלייה הזאת לא תהיה שונה בין הקבוצות, כי חצי מהנבדקים הוקצו לקבוצת המניפולציה וחצי הוקצו לקבוצת הביקורת, והרגרסיה אמורה לקרות בשתי הקבוצות בצורה שווה. אבל, חוקרים עדיין יכולים לטעות בביתוחי המשך ולהסיק מסקנות שגויות בשל התעלמות מתופעת הרגרסיה. למשל להסיק שבתוך קבוצת הניסוי, המניפולציה השפיעה בעיקר על אלו שהיה להם ציון נמוך מההתחלה. איום הרגרסיה קיים רק בביתוחי המשך.
- ברירה - אין איום ברירה כי ההקצאה היא מקרית. אנחנו לא צופים שיהיה הבדל בין שתי הקבוצות בציון ההתחלתי שלהם כי הם הוקצו באופן אקראי. כמובן שזה תלוי בגודל המדגם - ככל שהמדגם גדול יותר, ההקצאה המקרית מונעת הבדלים מקריים בין הקבוצות. במקרה של מדגם קטן נוכל לצמצם את איום הביררה, נוכל להוסיף התאמת (השוואת) נבדקים- כלומר נוכל לוודא שהשונות בין רמות ה IQ בין שתי הקבוצות יהיו קטנים. לדוגמה - נחלק את הנבדקים לזוגות לפי נתוני ה IQ שלהם, $120+119$ ($120+123$ וכו') ואז נוודא שכל אחד מהזוגות מתחלק באופן שווה בין הקבוצות, כלומר אחד מכל זוג ילך לקבוצה אחת והשני לקבוצה השנייה. זו שיטה יפה אבל לא מאוד נפוצה כיוון שדורשת הרבה עבודה, ופעמים רבות לא ברור לפי מה יש לחלק את הזוגות. חשוב לזכור שהדרך הכי טובה להתמודד עם הבדלים אקראיים בין הקבוצות היא הקצאה מקרית ומדגם גדול מספיק.

איומים שכן קיימים על המערך:

- נשירה - נבדקים עלולים לנשור מהטיפול שלנו וזה עלול לגרום לנו לבעיות. קיים איום של נשירה דיפרנציאלית (נפרדת) בין הקבוצות, אך נוכל לבדוק אם אכן זה קרה. למשל, נשירה גדולה יותר בקבוצת הטיפול בגלל הטיפול, וזה עשוי לגרום לבעיות. היתרון הוא שאנחנו יכולים לבדוק אם זה קרה.
- דוגמה- הטיפול (X) הוא דיאטה, רק נבדקים בעלי מוטיבציה ישרדו אותו ויגיעו עד למדידה הסופית. בקבוצת הביקורת לא

תהיה בעיה כזו כי המטופלים לא עברו שום דבר שיכול לאתגר את המוטיבציה שלהם.

יש כמה דרכים להתמודד עם הבעיה-

א. ראשית אפשר למדוד את כל הנבדקים כולל אלו שנשרו מהטיפול, כך נדע שלא ההבדל בין שתי קבוצות הניסוי הוא זה שהוביל לנשירה. כמובן שזה פוגע ביכולת לאמוד את השפעת הטיפול, כי אם נביא למדידה השנייה גם את אלה שנשרו ולא עברו את הטיפול זה פוגע בעוצמה הסטטיסטית. אבל, אם גם אחרי שאנחנו מוסיפים את הנבדקים שנשרו, ההבדל בין שתי הקבוצות נשאר מובהק, אז התוצאה לא חשופה לאיום של נשירה. כלומר הנשירה לא תסביר את תוצאות הניסוי מעבר למניפולציה. יכולה להיות בעיה אפילו יותר גדולה אם לא נצליח למדוד את הציונים של מי שנשר, במקרה הזה המערך הזה יכול לבדוק אם הייתה בעיית נשירה לא זהה בין שתי הקבוצות, אבל יכול להיות שלא נוכל לדעת אם ההבדלים בין הקבוצות נבעו מנשירה או מהטיפול, ולכן זו בעיה בתוקף הפנימי של הניסוי.

ב. דרך נוספת להתמודד עם הבעיה היא להזמין את קבוצת הביקורת לטיפול דומה ככל האפשר ולגרום לאותה סלקציה.

ג. דרך אחרונה להתמודד היא בעזרת אלגוריתמים שאמורים לחזות איך הנתונים החסרים יראו על סמך הנתונים שכבר יש לנו ביד. וכאן נכנסות כל מיני הנחות יסוד של החוקרים, כמו למשל שמי שנשר יראה כמו מי שלא נשר, לכן זו שיטה בעייתית כי קשה להוכיח את ההנחות האלה בצורה מספיק אמינה. עושים את זה רק כשחייבים אבל צריך להכיר במגבלות והחולשות של השיטה הזאת.

- תוקף חיצוני - יכולים להיות מספר איומים על התוקף החיצוני במערך זה:

א. יכול להיות שהתוצאות שאנחנו מקבלים במערך כזה נכונות רק בתנאים הספציפיים בהם נערך הניסוי (הנסיבות), והן לא מייצגות את המציאות. אפשר לחשוב על האיום הזה בטענה שיש אינטראקציה בין תנאי הניסוי לבין תוצאות הניסוי, כלומר להשפעת הטיפול. אם היינו בודקים מערך ניסוי אחר יכול להיות שלא היו מתקבלות אותן תוצאות.

לדוגמא- אולי הייתה לנו נסיינית ג'ינג'ית והתוצאות של הניסוי מתקבלות רק עם נסיינים ג'ינג'ים לצורך העניין.

ב. אינטראקציה בין בשילה לטיפול, כלומר יתכן והטיפול עובד רק כשהנבדקים שלנו משתנים עקב בשילה, זה איום על תוקף חיצוני כי אי אפשר להכליל על נבדקים שלא יחוו את הבשילה.

ג. אינטראקציה בין מדידה לטיפול, כלומר יכול להיות שהניסוי עובד רק כשהנבדקים נמדדים לפני הטיפול, גם זה איום על תוקף חיצוני.

לסיכום, יש מעט מאוד איומים על תוקף פנימי במערך מחקר לפני אחרי עם קבוצת ביקורת. האיומים שכן קיימים על התוקף הפנימי: איומים משניים של היסטוריה מקומית ואינטראקציה של נשירה עם מניפולציה.

לעומת זאת, יש לא מעט איומים על תוקף חיצוני, כי מטבע הדברים יכולת ההכללה של מערך כזה היא מוגבלת בגלל -

- דגימה

- היסטוריה

- מדידה

- עצם ההשתתפות בניסוי מבדילות מהמציאות (למשל אינטראקציה בין מדידה וטיפול)

בד"כ כשמבקרים תוקף חיצוני מתכוונים לכך שלא ברור שהקשר שנמצא בניסוי גם תקף במציאות. פסיכולוגיים בד"כ לא מוטרדים מטענות כאלה, אבל הם מוטרדים מאיומי תוקף חיצוני הפוכים. כלומר מקרים בהם לא הצלחנו למצוא קשר בין המשתנים, אבל אולי המחקר לא מצליח לשקף את המציאות בצורה מספיק טוב ובמציאות יש קשר בין המשתנים.

יש לציין שכל האיומים שהוזכרו בהקשר של תוקף חיצוני, רלוונטיים גם במערכי מחקר השגויים.

2. מערך ארבע הקבוצות של - Solomon

R	O ₁	X	O ₃
R	O ₂		O ₄
R		X	O ₅
R			O ₆

משלב את מערך לפני אחרי עם קבוצת ביקורת (2 השורות הראשונות) יחד עם קבוצת ביקורת וניסוי שלא נבדקו לפני 2 (שורות אחרונות). זה מערך שמיועד להפחית את איומי התוקף החיצוני.

במערך הזה אנחנו מקצים נבדקים באופן מקרי לאחת מתוך ארבעת הקבוצות, הקבוצה השלישית והרביעית מקבילות לקבוצה

הראשונה והשנייה אבל בלי מדידה ראשונה.

כדי למדוד בצורה הטובה ביותר את השפעת הטיפול - משווים את השינוי בין מדידות 1, 3 וגם 2, 4 (שתי קבוצות שאחת עברה טיפול ואחת לא).

כדי לקבל אינדיקציה נוספת ניתן להשוות בין מדידה 5 ל 6 (שתי קבוצות שאחת עברה טיפול ואחת לא) .

במדידות 5 ו- 6 אין לנו חשש שהייתה אינטראקציה בין המניפולציה למדידה הקודמת כי לא הייתה מדידה קודמת.

אם אנחנו רוצים לבדוק אם הייתה אינטראקציה בין המדידה לבין המניפולציה, אנחנו משווים בין מדידה 3 ל 5, ובין 4 ל 6 נובל לראות את ההשפעה של המדידה על התוצאות. החשיבות של מדידה 6 היא שהיא מאפשרת לנו למדוד את ההשפעה של חלוף הזמן, בשילה או היסטוריה על המדד התלוי. אנחנו יכולים להשוות את מדידות 6, 1, 2 כי כולם קבוצות שלא עברו מניפולציה (בשלב זה של הניסוי). זאת כדי לוודא שההקצאה היא באמת מקרית. אם נמצא כי אין הבדל בין 1 ל 2 וכן יש הבדל ביניהם לבין מדידה 6 זה ככל הנראה משהו שקשור לחלוף הזמן - בשילה או היסטוריה.

המערך הזה מאפשר עוד הרבה בדיקות והוא מצוין בלי ספק, אבל הוא מורכב בהרבה יותר ויקר והוא לא פותר את כל איומי התוקף החיצוני.

איומים קיימים על התוקף החיצוני:

- אינטראקציה בין מניפולציה לדגימה: הדגימה היא מאותה אוכלוסייה.

- אינטראקציה בין מניפולציה למצב הניסוי: כל הנבדקים עדיין יודעים שהם בניסוי.

ההמלצה היא להשתמש בו רק כשרוצים לבדוק אם ישנה אינטראקציה של טיפול עם מדידה, ואם יש חשיבות לאפקט בשילה. לפעמים משתמשים בו רק פעם אחת במערך מחקר זאת כדי ללמוד על השפעה המדידה המוקדמת.

3. מערך עם בדיקה אחרי בלבד -

R	X	O ₁
R		O ₂

מערך נפוץ מאוד בפסיכולוגיה חברתית וקוגניטיבית - כמו 2 הקבוצות האחרונות במערך של

סולומון. ההנחה הינה שאין צורך במדידה שלפני כדי להשתכנע שהאפקט של X לא נובע מההבדל הראשוני בין קבוצת הניסוי לקבוצת הביקורת, בגלל הרנדומיזציה. ההקצאה האקראית לקבוצות שאמורה ליצר דמיון בין הקבוצות, אין סיבה חוץ מהמדד התלוי שאמורה ליצר הבדל בין הקבוצות.

אין, baseline אבל עדיין יכולה להיות אינטראקציה בין מניפולציה לבין איומים של חלוף הזמן .

דוגמא -בודקים בקבוצה אחת כתבו בכמה משפטים מה מאפיין את איינשטיין, והנבדקים בקבוצה השנייה נתבקשו לכתוב בכמה משפטים מה מאפיין פרופסורים. לאחר מכן כל הנבדקים ביצעו מבחן טריוויה, הנבדקים שכתבו על איינשטיין היו פחות טובים בטריוויה מהנבדקים שכתבו על פרופסורים. לא נעשתה מדידה מקדימה של יכולת הנבדקים במבחן טריוויה, אבל בכל זאת החוקרים הסיקו שהסיבה להבדלים בין הקבוצות היא המניפולציה הניסויית- החיבור על איינשטיין. יש סיכוי שבאופן מקרי נבדקים חכמים יותר הוקצו לקבוצת איינשטיין מאשר לקבוצת הפרופסורים, אך בהנחה שהנבדקים הוקצו אקראית, הסיכוי ששתי הקבוצות הנבדקות האלו ששייכות לאותה אוכלוסייה יהיו שונות זו מזו מאופן מובהק, מושפע מגודל האלפא שהחלטנו עליו 5% (ברוב המקרים) -ככל שאלפא תהיה יותר קטנה הסיכוי שהקצאה לא מקרית (מקרית אבל לא מקרית בפועל) תשפיע לנו על המסקנות יקטן. לכן חוקרים לא מבצעים בדיקה מוקדמת כיוון שהם מניחים שהסיכוי להבדלים בין הקבוצות קטן מאוד.

יתרונות של המערך: המערך הזה טוב במיוחד במקרים בהם אי אפשר לבצע מדידה לפני, למשל אם היא מסובכת מאוד, או כאשר המדידה לפני עלולה להיות ריאקטיבית, כלומר תשפיע בעצמה על התהליך (חשש לאינטראקציה בין מדידה לטיפול).

למה בכלל יש מערכים עם מדידה לפני המניפולציה, בהנחה שהמערך זה כל כך טוב?

היתרון של מדידה לפני המניפולציה אם היא לא מהווה איום על התוקף, יכולה להקטין את השונות המקרית ולשפר את תוקף המסקנה הסטטיסטית, כי זה Baseline . אפשר לחשוב על מדידות סמויות שנבדקים לא יודעים עליהם.

הסבר למה זה משפר את תוקף המסקנה הסטטיסטית - אם יש לנו מדידת בייס ליון אנחנו יכולים לדעת אם המניפולציה שעשינו עשתה את ההבדל במקרים שהמדד התלוי שלנו "רועש". כלומר, אם נשתמש במשתנה תלוי, שבו יש הרבה הבדלים בין אישיים אז יש הרבה שונות. אפשר לחשוש שהשונות בין הקבוצות תושפע מהשונות הבין אישית והיא תהיה כל כך גדולה, שנצטרך שונות שיטתית ענקית כדי למצוא הבדלים. המדידה אחרי היא אותה מדידה בשתי הקבוצות ובגלל שיש מדידת בייס

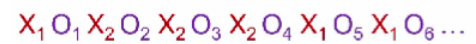
ליין אנחנו יכולים לבדוק מה היו ההבדלים בציונים בין שלב ראשון לשלב השני. מדידת הבייס ליין היא זאת שאפשרה לדעת מה באמת קרה בניסוי, היא מראה שיש הבדל מובהק בין הקבוצות.
 זאת בעיה שלא קשורה לתוקף פנימי או חיצוני, אלא לתוקף המסקנה הסטטיסטית - החשש הוא מטעות מסוג שני: עוצמת מבחן נמוכה מידי כי יש יותר מידי שונות באוכלוסייה.

לסיכום, המערך הנוכחי עמיד בפני איומי תוקף פנימי כמו המערך של לפני אחרי עם קבוצת ביקורת:
 - אין איום של בשילה, היסטוריה, ברירה, נשירה, רגרסיה לממוצע
 - יש איומים משניים של היסטוריה מקומית

האיומים על התוקף החיצוני הם דומים למערך הקודם:
 - הכללה מוגבלת בעקבות הדגימה, ההיסטוריה, או עצם ההשתתפות בניסוי המבדילות מהמציאות.
 איומים על תוקף המסקנה הסטטיסטית - טעות מסוג שני (עוצמת מבחן נמוכה).

עד כה דיברנו על מערכי ניסוי בסיסיים בין נבדקיים - ישנם מערכי ניסוי מורכבים יותר בהם 3 או 4 קבוצות טיפול שונות.

4. מערך תוך נבדקי עם מדידות חוזרות -



כל נבדק מתנסה בכל תנאי הניסוי, לרוב עם מדידות רבות בכל תנאי.
 לדוגמא - במטלת סטרופ הנבדקים מתבקשים לציין את צבע המילה ויש שני תנאים - מילה כתובה מתארת את הצבע וזה תואם, או מילה כתובה מתארת צבע אחר וזה תנאי לא תואם.

כאן אין הקצאה מקרית לתנאי הניסוי, אבל כל הנבדקים עוברים את כל התנאים, וכל נבדק מתנסה בכל תנאי הרבה מאוד פעמים. בד"כ כדי להשוות בין התנאים בניסוי, מחשבים ממוצע של כל המדידות בתנאי אחד, וכל המדידות בתנאי השני, ועורכים את המבחן הסטטיסטי על הממוצעים של התנאים עבור כל נבדק, במבחן T לממוצעים תלויים, או בניתוח שונות לממוצעים תלויים. אפשר לנתח ניתוחים מורכבים יותר, שלוקחים בחשבון את מספר הצעד כמשתנה בלתי תלוי (פחות רלוונטי כרגע). ניתוח שונות תוך נבדקי נקרא לרוב ניתוח שונות של מדידות חוזרות.

אפשר לקחת את המערך הזה למקומות יותר מורכבים:

לדוגמא - ניסוי בו יש שני תנאים ולכל תנאי יש שתי רמות. במטלת סטרופ ניתן לחשב את התנאי בשלב הנוכחי (תואם או לא תואם) וגם את התנאי בשלב הקודם (תואם לא תואם). במקרה כזה כל צעד של הנסיין יחולק לאחד מארבעת האופציות (2 גורמים ולכל גורם 2 רמות, סה"כ ארבע אופציות). נחשב בנפרד לכל תא את הממוצעים, ונבצע ניתוח שונות רב גורמי עם משתנים תוך נבדקים (נלמד ברגרסיה).

אפשר להוסיף עוד גורמים. לדוגמא - הקצאת נבדקים לתנאי מצב רוח חיובי או שלילי, ואז הנבדקים השלימו מטלת סטרופ. נוצר לנו מערך של 2 (מצב רוח: חיובי, שלילי, בין נבדקי) על 2 (תואמות הצעד הנוכחי: תואם, לא תואם, תוך נבדקי) על 2 (תואמות הצעד הקודם: תואם, לא תואם, תוך נבדקי). בתוך כל אחד מהתנאים מחשבים את הממוצעים הרלוונטים.

סדר הצגת התוצאות לא מאוד משנה, אבל בד"כ מה שמקובל זה שברמה הכי גבוה מציגים את המשתנים הבין נבדקים, ומתחתיו את המשתנים התוך נבדקים.

זהו מערך חשוב כיוון שהוא מאוד חזק, אבל בכל זאת יש איומים:

- אפקט סדר - האיום המרכזי במערך זה, מאיים על התוקף הפנימי של הניסוי. איום בו סדר התנאים גרם לאפקט, כלומר הרמות השונות במשתנה הבלתי תלוי הוצגו בסדר מסוים שגרם לתוצאות שאנחנו רואים. זה קיים רק במערך שהוא תוך נבדקי, כי הנבדקים עוברים את כל סוגי המניפולציה. זה סוג של ארטיפקט ניסויי.

דוגמא - אזהרה על עבירה ראשונה, קנס על עבירה שנייה, מאסר על עבירה שלישית. ניתן לחשוב על ממצא תיאורטי -

כמות העבירות יורדת כשהעונש הצפוי גדול יותר. במקרה הזה המשתנה הבלתי תלוי הוא עונש ורמותיו, אנחנו לא יכולים לדעת אם המשתנה התלוי (כמות העבירות) נובע מרמת המשתנה הבלתי תלוי או מהעיתוי של הפעלת המניפולציה. יכול להיות שעונש שני משפיע בצורה יותר טובה פשוט מעונש ראשון. יכול להיות שאם היינו שמים את זה בסדר הפוך אז היינו מוצאים שככל שהעונש קטן יותר יש פחות עבירות.

יש לשים לב כי הסדר יכול לגרום לטעות מסוג ראשון וגם טעות מסוג שני (להרוס אפקט קיים, או ליצור אפקט בגלל הסדר שלא באמת קיים).

לדוגמא - אם עונש המאסר היה על העבירה הראשונה והאזהרה הייתה על העבירה השלישית, היה יכול להיות מצב שאנשים בד"כ לא היו עושים עבירה שלישית ואז גם במקרה כזה עבירות בתנאי אזהרה יהיה קטן. ואז בטעות היינו יכולים להסיק שאין השפעה לגודל העונש, אבל בעצם אפקט הסדר העלים את אפקט העונש כי הוא פעל בכיוון ההפוך.

ישנם שלושה פתרונות לאיום זה :

א. מערך בין נבדקי - כל נבדק מקבל רק רמה אחת של המשתנה הבלתי תלוי. החיסרון בזה הוא שונות טעות גדולה יותר כי יש הבדלים בין אישיים רבים, ואנחנו צריכים מספר גדול יותר של נבדקים ומאבדים משאבים וכן מתקיימים כל האיומים שנגענו בהם .

ב. סדר אקראי, רנדומיזציה - הפתרון הכי מקובל, פשוט להגריל את סדר המניפולציות ולהעביר אותם בסדר אקראי, באופן שונה לנבדקים שונים, כך שהסדר לא יהיה מוצמד באופן שיטתי למשתנה הבלתי תלוי. זה לא פתרון אצל נבדק בודד, אבל מעבר לנבדקים לסדר לא יהיה אפקט שיטתי. חסרונות של פתרון זה:

- חסרון 1: הגדלת השונות המקרית - השונות המקרית גדולה אם לסדר אכן יש השפעה.
- חיסרון 2: צורך משאבים - המערך מורכב ודורש הרבה משאבים.
- חיסרון 3: קושי בזיהוי אינטראקציה עם סדר - במקרה וסדר התנאים הוא באמת חשוב, יש סיכוי גדול שפשוט לא נדע עם זה.

ג. איזון counterbalanced order - שימוש במספר סדרים, כך שכל תנאי יופיע בכל מקום סידורי אותו מספר של פעמים. הפתרון הזה מתייחס למטרה של שימוש בהרבה סדרים, שהיא לוודא שאם יש אפקט של אחד הגורמים, הוא לא יהיה כבול לסדר מסוים, כלומר שנדע שהאפקט קיים מעבר לסדר. יש כמה דרכים להשתמש בכל הסדרים האפשריים .

סדר הצגה				ריבוע לטיני Latin square
1	2	3	4	
A	B	C	D	
C	A	D	B	
B	D	A	C	
D	C	B	A	

לדוגמא "ריבוע לטיני" - בו כל תנאי מוצג פעם אחת בכל מקום סידורי. זה מגביר את האיזון בין התנאים, כך שאם אנחנו מוצאים שלסדר אין אפקט אינטראקציה עם המניפולציה, זה מראה שאין איום של סדר בניסוי.

- השפעה נמשכת - carry over לתנאי נתון יש השפעה על המשתנה התלוי גם בתנאים אחריו, כלומר- יש סכנה שהמניפולציה מהתנאי הקודם תמשיך להשפיע גם על התנאי הבא.

לדוגמא - טיפול תרופתי: האיום הוא שתרופה תמשיך להשפיע גם אחרי שהתפוגגה בגופנו, וכך תשפיע על התנאי הבא. בעיה גדולה במערכים תוך נבדקיים בתרופות.

פתרונות:

א. לעבור למערך בין נבדקי

ב. לתת מרווח זמן גדול יותר בין התנאים

איך נערוך מערך תוך נבדקי בצורה טובה?

- חשוב שסדר התנאים יהיה שונה לכל נבדק, ונימנע כך מאיומי סדר.
- בנוסף חשוב לעשות הרבה מדידות לכל נבדק, זה עוזר לנו לדעת שהמשתנה הבלתי תלוי הוא זה שמשפיע על התנאים.
- במערכים תוך נבדקים אנחנו לא רוצים מדידות תגובתיות, אם מדידה היא תגובתית אנחנו לא יכולים לערוך מחקר תוך נבדקי. לדוגמא -אם המדד הוא שאלון דיכאון, אי אפשר להעביר אותו שוב ושוב אלא אם כן אנחנו עורכים את זה לאורך זמן ממש ארוך. במקרה כזה, יכול להיות שהנבדקים יזכרו את התשובות שלהם מהפעמים הקודמות ויפסיקו למלא את השאלון

ברצינות, וזה מבאס, או שהם יתבאסו שהם מגלים שהם מדוכאים כל שבוע מחדש, ואז הם יתחילו לענות תשובות אחרות שיגרמו לנו לחשוב שהם כבר לא מדוכאים למרות שהמצב הוא שונה, ולכן אנחנו לא רוצים שהמדידה תהיה תגובתית. - לתנאים צריכה להיות השפעה קצרת מועד, אנחנו לא רוצים שתהיה השפעה מתמשכת, כדי שנוכל לשנות את המשתנה הבלתי תלוי שוב ושוב.

לסיכום, למערך הזה יש הרבה יתרונות בכל מה שקשור לתוקף פנימי. אם נמצא קשר בין המשתנה הבלתי תלוי לתלוי, קשה מאוד למצוא משתנה מתערב שבמקרה ישתנה בצמוד למשתנה הבלתי תלוי שלנו ויסביר את המשתנה התלוי) פשוט כי שינינו את המשתנה הבלתי תלוי הרבה מאוד פעמים בתוך אותם נבדקים). כמו כן, באופן ייחודי למערך הזה אין חשש מהיסטוריה מקומית (בערך המערך היחיד בו לא קיים חשש כזה) כי כל המשתנים עוברים את אותם תנאים בדיוק. אין חשש מאינטראקציה של נשירה ומניפולציה כי כולם עוברים את המניפולציה. עם זאת, עדיין יכולים להיות פגמים בתוקף מבנה, ויתכנו לעיתים איומים על תוקף פנימי של - אפקט סדר, מדידה תגובתית ואפקט מתמשך.

איומים על תוקף חיצוני דומים מאוד לאיומים שראינו במערכים תקינים אחרים - הדגימה, ההיסטוריה והאינטראקציה. בנוסף המדידה היא איום חזק יותר כי יש הרבה מדידות חוזרות. איומים שקשורים לחזרה על המניפולציה מן הסתם משפיעים יותר במערך תוך נבדקי כי חוזרים על המניפולציה שוב ושוב, וכאן יש משהו ייחודי למערך תוך נבדקי - אפקט איום על תוקף חיצוני שלפיו הקשר קיים רק כשאנחנו משנים שוב ושוב את המשתנה הבלתי תלוי, על זה נרחיב בהמשך.

איומים נפוצים ופתרונות

איומים על תוקף חיצוני

1. מדידה

דוגמא - נבדוק האם פעילות גופנית משפיעה על חרדה מצבית (=חרדה בזמן נתון, שאינה בהכרח קשורה לחרדה תכונתית קבועה של אדם מסוים). מהלך הניסוי: מדידת חרדה מצבית, מניפולציה של ריצה בפארק, (קבוצת הביקורת, שהוקצתה מקרית, יושבת במעבדה חצי שעה), מדידת חרדה מצבית נוספת. נמצאה רמת חרדה נמוכה יותר לאחר ריצה בפארק.

זו דוגמה לכך שהרבה מהמדידות שנבדקות בניסויים בפסיכולוגיה, לא מתקבלות במציאות. כאן אנחנו מודדים חרדה מצבית בעזרת שאלון, ובמציאות בני אדם לא ממלאים שאלוני חרדה. הבדל מהותי בסיטואציה בניסוי לבין המציאות.

אין כאן איום על תוקף פנימי כמעט בכלל כי שתי הקבוצות עברו את אותן המדידות ולכן אנחנו יכולים להסיק שהריצה היא שעשתה את ההבדל בין שתי הקבוצות. אבל האיום על יכולת ההכללה פה הוא נובע מכך שהריצה אולי עוזרת רק אם מודדים את החרדה לפני שרצים, או שמודדים אותה באופן הספציפי בו מדדנו.

לכן, האיום הוא בעצם אינטראקציה בין מדידה למניפולציה – ייתכן שריצה בפארק עוזרת רק לאחר שחושבים על החרדה, בעקבות שאלון החרדה.

האיום של השפעת המדידה על התוקף החיצוני, נקרא לפעמים "איום סיטואציות המבחן."

פתרונות:

- ניתן לבצע את המדידה בלי ידיעת הנבדקים (מדידה לא תגובתית) - כשאנחנו משתמשים במדידה לא תגובתית, חשוב לוודא שלא מאבדים את היכולת למדוד כראוי את מה שאנחנו מודדים. למשל באמצעות שופטים שישתכלו על סרטונים של הנבדקים לפני ואחרי וינתחו את רמת החרדה שלהם. עם זאת, עשויה להיות בעיה כי מדד החרדה אינו מדויק.
- אפשר למדוד את המשתנה כחלק אינטגרלי מחיי היום יום ולא מהניבוי. למשל אם המדידה היא מבחן בבי"ס או פעולה אחרת שאנשים בכל מקרה עושים כל הזמן. כך אין בעיה להכליל, כי המדידה מתרחשת גם מחוץ לניסוי – צריך לוודא שהמדידה דומה ככל האפשר למה שאנחנו עושים במציאות.

2. היסטוריה

דוגמא - הניסוי התקיים במהלך מבצע צבאי. או שהניסוי התקיים יום לאחר רצח רבין.

אנחנו חושדים שההיסטוריה של כל הנבדקים בניסוי שונה מההיסטוריה של הנבדקים שאנחנו רוצים להסיק באמצעותם על כל האוכלוסייה (נקרא קוהולט אפקט). יכול להיות שהמניפולציה משפיעה על הנבדקים באופן שונה בגלל האינטראקציה שלה עם ההיסטוריה.

כלומר, ייתכן איום של אינטראקציה בין היסטוריה למניפולציה – יש היסטוריה ייחודית משותפת לכל הנבדקים בניסוי, אך לא לשאר הימים בהיסטוריה האנושית. למשל כל הניסויים שהורצו בשנים האחרונות בעולם מושפעים מהקורונה. דוגמא נוספת - ניתן להטיל ספק בניסויים שהתרחשו לפני עשרות שנים, ייתכן והממצאים תלויים בנורמות חברתיות שהשתנו מאז.

3. המצב הניסויי

דוגמא - בחינת עוררות אמפתיה על הנטייה לתרום לצדקה. מעוררים אמפתיה ע"י צפייה ב 10 שעות של סרטונים של כלבלבים חמודים.

הבעיה – נדיר שאמפתיה תתעורר באופן טבעי בצורה זו. זה לא נפוץ באופן טבעי, לעשות את המניפולציה הזו.

זו בעיה מרכזית של ניסויים, הם ניסויים, כנראה שאנשים מתנהגים באופן מאוד שונה בחיים האמיתיים כשהם לא בזמן ניסוי.

כלומר, ייתכן איום של אינטראקציה בין המצב הניסויי למניפולציה – ייתכן והמניפולציה עובדת רק תחת התנאים הספציפיים בהם ביצעו את הניסוי, אבל לא תעבוד במצבים אחרים. זה יכול לקרות כי אנשים מתנהגים בניסוי באופן שונה ממה שהיו מתנהגים במציאות או כי גורמים סביבתיים שונים מגורמים ניסויים וכו'.

פתרונות:

1. לנסות להפוך את הניסוי לדומה ככל האפשר למציאות. למשל בבית ספר להכניס את הניסוי לחלק מסדר היום של התלמידים שלא יבחינו בו.
2. לבצע ניסוי ללא ידיעת הנבדקים (למשל חברות מסחריות שמציעות תפריטים שונים בסניפים שונים ובדקים באיזה סניף המכירות טובות יותר) - יש בזה קושי אתי רב.

4. דגימה

ייתכן איום של אינטראקציה בין דגימה למניפולציה - המניפולציה עובדת, אבל רק עבור האוכלוסייה שממנה דגמנו, ולא ניתן להכליל לאוכלוסיות אחרות.

איומים על תוקף מבנה

1. קונפאונד במניפולציה (אופרציונלי)

הופעל על ידי המניפולציה הניסויית משתנה תיאורטי אלטרנטיבי, שרמותיו מוצמדות לרמות המשתנה הבלתי תלוי. זה משפיע על תוקף המבנה כי אנחנו לא יכולים להסיק שהמשתנה הבלתי תלוי הוא זה שהשפיע על המשתנה התלוי. יש היעדר בלבדיות של המניפולציה. במקרה החמור יותר, המניפולציה הניסויית בכלל לא הפעילה על המשתנה שהתכוונו להפעיל אלא המשתנה התיאורטי האלטרנטיבי הוא זה שהפעיל.

דוגמא - השוואת מהירות הקריאה של מילים גסות מול מילים לא גסות. החוקרים הסיקו כי גסות המילה גורמת לקשיים בזיהויה, מילים גסות גורמות לקריאה איטית. הסבר אלטרנטיבי (האיום) - לא גסות המילה היא שהשפיעה על מהירות הקריאה אלא השכיחות שלה. אי אפשר לדעת מה השפיע על מהירות הקריאה- הנדירות של המילה או הנוחות שלה. המניפולציה של החוקרים היא בחירת מילים גסות לעומת לא גסות, אבל המניפולציה השפיעה גם על גסות וגם על שכיחות ולא ניתן לדעת מה השפיע על התוצאות.

פתרון:

Unconfounding באמצעות התאמה/השוואה - לדאוג שמשנתנה השכיחות ומשתנה הגסות לא יהיו מוצמדים זה לזה. עושים זאת באמצעות התאמה בין השכיחות של כל אחת מהמילים בקבוצת המילים הגסות לבין השכיחות של כל אחת מהמילים בקבוצת המילים הלא גסות. כלומר, נחפש מילים שהשכיחות שלהם בשפה זהות בין שתי הקבוצות: אם בקבוצת הגסות בחרנו במילה "טמבל" נצטרך למצוא מילה מקבילה בקבוצת המילים הלא גסות שהשכיחות שלה בשפה זהה לשכיחות המילה "טמבל". כך נוכל לדעת שאין הבדל במשתנה התיאורטי "שכיחות המילה" בין קבוצת המילים הגסות לקבוצת המילים הלא גסות, ואם נמצא הבדל בין שתי קבוצות המילים האלה נוכל להניח שהשכיחות היא לא הסבר אלטרנטיבי.

2. מערך הסכמה (א)

לבני אדם נטייה לענות בחיוב להיגדים שונים באופן בלתי תלוי בתוכנם.

דוגמא - חוקרים רוצים למדוד פשיזם באמצעות שאלון. בודקים האם הנבדקים מצייתים יותר לנסיין סמכותי לעומת לא סמכותי. נבדקים עם ציון גבוה בשאלון פשיזם לא הראו נטייה לציית יותר לנסיין סמכותי מאשר לנסיין לא סמכותי. לטענת החוקרים: אין עדות כי נטייה לפשיזם מגבירה היכנעות לסמכות.

הסבר אלטרנטיבי (האיום) - הבעיה היא שתמיכה בטענת החוקרים יכולה לנבוע מתמיכה בפשיזם אבל גם מהנטייה הטבעית של אנשים להסכים עם היגדים. אם התופעה הזאת משפיעה על המדד - הנבדק נוטה לומר "כן" על היגדים באופן כללי, המסקנה שלנו על כך שהוא תומך בפשיזם יורדת, כי הרבה מידי נבדקים מביעים תמיכה עם ההיגד רק כי הם מסכימים עם היגדים באופן כללי בלי קשר לתמיכה בפשיזם כן או לא. אז זה מערער את היכולת שלנו למצוא את הקשרים שמעניינים אותנו.

3. מערך הסכמה (ב)

בני אדם שונים זה מזה בנטייתם לענות בחיוב להיגדים שונים באופן בלתי תלוי בתוכנם. לא הנטייה הכללית של בני אדם להגייד כן, אלא השונות הבין אישית בנטייה זו.

דוגמא - נבדקים עם ציון גבוה בשאלון פשיזם נטו לציית יותר לנסיין סמכותי מאשר לנסיין לא סמכותי. לטענת החוקרים: נטייה לפשיזם מגבירה היכנעות לסמכות.

הסבר אלטרנטיבי (האיום): שאלון הפשיזם מדד גם את הנטייה להסכים עם היגדים. לא נדע אם התוצאות נובעות מהסכמה לפשיזם או נטייה להסכים באופן כללי.

פתרון:

צריך להוסיף היגדים לשאלון שהסכמה איתם תבטא את ההפך מהסכמה עם היגדים אחרים. אלה נקראים פריטים בקידוד הפוך - reverse coding, אשר כדי לענות עליהם באופן עקבי עם פריטים אחרים מצריך מענה הפוך. כלומר לפי הדוגמא - צריך להוסיף היגדים אנטי פשיסטיים. נבנה שאלון בו חצי מהשאלות פשיסטיים יסכימו חצי אנטי פשיסטיים יסכימו, נוכל לראות אם מדובר בנטייה הטבעית של אדם להגייד כן, או נטיותיו האידיאולוגיות. אני סכנה לקונפאונד בין הנטייה להסכים לבין הנטייה לפשיזם. כלומר זה איום שהמדד לא מודד את המשתנה התיאורטי שאנחנו רוצים למדוד, אלא הוא מודד גם הסכמה בגלל איך שבנינו את השאלון.

הפתרון מתמודד גם עם בעיה של נבדקים שממלאים את השאלונים באופן רנדומלי.

4. רצייה חברתית

לחץ חברתי, המדד הושפע מרצון הנבדקים להיתפס כחיוביים (בעיני הזולת או בעיני עצמם).

דוגמא - בשאלון נבדקים נשאלו לגבי כמות הפרטנרים המיניים איתם היו, ונבדקים נטו לדווח על מספר גדול יותר של פרטנריות מאשר נבדקות. לטענת החוקרים: לגברים במדגם יש יותר פרטנריות מנשים.

הסבר אלטרנטיבי (האיום) - בגלל נורמות חברתיות גברים רוצים להציג את עצמם כמי שקיימו הרבה, בעוד שנשים רוצות להציג את עצמן באופן הפוך. כלומר ההבדל בין הקבוצות נובע מהנטייה של הנבדקים להציג את עצמם כחיוביים. הרצייה החברתית גורמת לקונפאונד במשתנה התלוי שמאיים על תוקף המבנה, כי אמנם בודקים מספר פרטנרים, אבל המשתנה התיאורטי של 'רצון להיות מקובל בחברה' השפיע על המדד התלוי.

פתרונות:

1. הוספת תנאי אנונימיות - אף אחד לא יקשר בין נתוני הנבדק לבין הזהות שלו. וכך נוכל לקוות שהנבדקים יהיו ישרים. הבעיה היא שלפעמים יש אנשים שעדיין ישקרו כדי להרגיש יותר טוב בעיני עצמם.
2. להשתמש במדד שאינו דיווח עצמי - לשאול את הנבדקים ולבדוק אם שיקרו באמצעות מצלמות נסתרות וכו'...

איומים על התוקף הפנימי

1. רציית הנסיין Demand characteristics

הנבדקים רוצים לרצות את הנסיין, אז הם מתנהגים בהתאם להשערות, כפי שהם מנחשים אותה. זה פוגע בתוקף הפנימי כי זה לא קורה בגלל איך שאנחנו מגדירים את המניפולציה אלא זה בגלל שיש משהו אחר שמוצמד להשערה שלנו.

דוגמא - צוות מטפלים קוגניטיביים-התנהגותיים בדיכאון השוו CBT לעומת טיפול תרופתי. מספר הסימפטומים הדיכאוניים של המטופלים נמדדו לפני ואחרי הטיפול בעזרת שאלון. נמצא ש CBT היה יותר טוב מטיפול תרופתי בהפחתת סימפטומים. מסקנת החוקרים הייתה ש CBT טוב יותר מטיפול תרופתי. אם זו היתה השערת החוקרים, ייתכן והם שידרו את זה לקבוצת הניסוי. נוצר הסבר אלטרנטיבי (האיום) (הנבדקים הבינו שהמטפל מעדיף את טיפול ה CBT והגיבו בהתאם בשאלון. הנסיין יכול גם בצורה לא מודעת) לגרום לנבדקים להתנהג בצורה מסוימת.

ישנו חשש גם לאפקט הפוך- שהנבדקים יעשו "דווקא" ויתמרדו נגד השערת הנסיין. האיום מתרחש כי גורם חיצוני שהוא לא רלוונטי להגדרות של המשתנים מוצמד לרמות המשתנה הבלתי תלוי.

פתרונות:

1. נסיין עיוור להשערה - כך הוא לא יוכל לשדר לנבדקים את הציפיות שלו, והם לא יוכלו לקלוט את הציפיות. זה פתרון חלקי, נסיינים יכולים לנחש את ההשערה בניסוי ולהשפיע גם בלי לדעת על הנבדקים בניסוי.
2. נסיין עיוור גם לתנאי - כלומר, הנסיינים לא ידעו בכלל מה הם התנאים בניסוי. לא תמיד אפשר לעשות את זה, ונסיינים יכולים להמשיך לשער את ההשערה והתנאים ולהשפיע על נבדקים.
3. נבדק עיוור להשערה
4. נבדק גם עיוור לתנאי - לא יודע באיזה קבוצת ניסוי הוא. כמו ניסויים בתרופות עם קבוצת פלסבו.
5. לשקר לנסיינים ו/או לנבדקים - עשוי לפתור את האיום שהנסיינים או הנבדקים ינסו לשער את ההשערה, ככה נדע בדיוק מה ההשערות של הנבדק והנסיין. יש בזה חיסרון כי זה לא אתי, וגם מגביר חשדנות גם כלפי הניסוי הספציפי אבל גם כלפי ניסויים באופן כללי.

blind Double – הנבדק וגם הנסיין עיוורים לפחות לתנאי הניסוי. זה הסטנדרט בניסויי תרופות וחיסונים. יודעים במה עוסק המחקר אבל לא יודעים לאיזה קבוצה הנבדקים מוקצים.

דוגמא נוספת - נבדקים גברים התבקשו לכתוב על משהו עצוב בעברם, וקבוצה שניה התבקשו לכתוב על משהו שמח. לאחר מכן התבקשו להעריך עד כמה מועמדת מתאימה לתפקיד ניהול. נמצא ששמחים מבקרים בפחות חומרה מעצובים. הסבר אלטרנטיבי (האיום) - עצובים הניחו שהנסיין מצפה שהם יהיו ביקורתיים יותר מאשר השמחים וינסו לרצות אותו.

למה זה איום על תוקף פנימי ולא תוקף מבנה - איום על תוקף מבנה קורה כשאנחנו מגדירים את המשתנים התיאורטיים ומהן רמות המשתנים האופרציונליים שמעניינים אותנו. במקרה זה חוקרים הגדירו כתיבה על משהו שמח/עצוב בתור אופרציונליזציה שמח/עצב. אם נחשוב שכתובה על נושא מעוררת את הרגש אז ניתן לחשוב שתוקף המבנה סביר, במקרה כזה ההסבר האלטרנטיבי, שהנבדקים ינסו לחשוב מה הנסיין רוצה, משפיע על התוקף הפנימי ולא על המבנה. יש גבול דק ביניהם.

נחשוב על הפתרון האפשרי לרציית נסיין: נגיד שהנסיין רובוט ויודע לעשות כל מה שנסיין רגיל עושה, נבדקים לא יחשבו שיש לו רצונות. המשתנה האופרציונלי במקרה הזה הוא אותו משתנה, אבל קשה מאוד לטעון שהמשתנה האופרציונלי הושפע על ידי משתנה תיאורטי של רצון הנסיין. סביר להגדיר לכן איומים של רציית נסיין כאיום על תוקף פנימי, ולא על תוקף מבנה. כי אם ננטרל את רציית הנסיין על ידי נסיין רובוט, האיום יתבטל. כלומר- זה לא בעיה בהגדרת המשתנים.

2. הטיית הנסיין

הנסיין רוצה שהשערת המחקר תוכח כנכונה, אז הוא משפיע על תוצאות הנבדקים בהתאם, לעיתים בצורה בלתי מודעת. דוגמה - בוחנים עד כמה נשים יסתכלו בעיניים של נסיין גבר לעומת נסיינית אישה, בעת שיחה. מצאו שנשים יוצרות קשר עין ממושך יותר עם נשים מאשר גברים.

הסבר אלטרנטיבי (האיום) - הנסיינים ידעו מה הייתה ההשערה וחיפשו קשר עין בהתאם להשערה (השפיעו על המדד).

פתרונות:

1. נסיין עיוור להשערה ו/או לתנאי
2. אולי אפילו לשקר לנסיין לגבי ההשערה

3. ארטיפקט סטטיסטי

הקשר בין המשתנים נובע מקשר אלגברי בין המשתנים (לא קשור לתוקף המסקנה הסטטיסטית אלא לתוקף הפנימי כיוון שיש משתנה נוסף - קשר אלגברי, שמשפיע על המשתנה התלוי).

דוגמה - רמת שאיפה (מונח בפסיכולוגיה) - נותנים לאדם תפקיד (למשל לירות למטרה) ולאחר מכן שואלים אותו כמה יקבל בפעם הבאה, רמת השאיפה היא ההפרש בין הציון שלו לשאיפה שלו. אם הציון חיובי מאוד האדם הוא אמביציוזי. ניתן לשאול מה הקשר בין ההישגים ורמת השאיפה של אנשים. רואים כי אנשים בעלי הישגים נמוכים מראים שאיפה גדולה יותר לשיפור בהשוואה לאנשים בעלי הישגים גבוהים.

הישג	שאיפה	שאיפה לשיפור
90	95	+5
80	95	+15
65	85	+20
65	87	+22
78	89	+11
89	96	+7
92	95	+3

הסבר אלטרנטיבי: השאיפה לשיפור מחושבת על סמך רמת ההישגים: כיוון שחלק $b - a$ נמצא במתאם שלילי מושלם עם a ($r_{a,-a} = -1$), מן הסתם שגם $r_{a,b-a}$ יהיה שלילי גם הוא!

הסבר אלטרנטיבי (האיום) - זה ההישגים, b זה השאיפה (כמה היית רוצה לקבל), רמת השאיפה (ההפרש בין מה שהיית רוצה לקבל למה שקיבלת) היא $b-a$. החוקרים הסיקו שככל שההישגים נמוכים יותר רמת השאיפה גבוהה יותר. הבעיה היא שרמת השאיפה מוגדרת ביחס ל a , אז יש משתנה תלוי שהוא $b-a$ ומשתנה בלתי תלוי שהוא a אז בהכרח יהיה מתאם שלילי בין המשתנים האלה בגלל איך שאנחנו מגדירים אותם, ולכן בהכרח נמצא קשר שלילי.

הארטיפקט הסטטיסטי יכול להתרחש כתוצאה מכל קשר אלגברי בין משתנים (ולא רק חיסור כמו בדוגמה).

איזמים על תוקף המסקנה הסטטיסטית

1. דרגות החופש של החוקר

ההשפעה שיכולה להיות על ההטייה של הנסיין בשלבי ניתוח הנתונים, יש המון מקומות בהם הנסיין יכול להשפיע על ניתוח הנתונים. יש המון החלטות לקבל, וכל החלטה כזו מובילה לסדרת החלטות נוספות. למשל:

- מתי להפסיק את איסוף הנתונים?
- האם לנתח את כל המשתנים התלויים? איך לציין אותם?
- בין אילו קבוצות/תנאים יש להשוות? באיזה שיטה/מבחן סטטיסטי להשתמש?
- אילו נבדקים הם חריגים?
- באילו משתני רקע יש "לשלוט" סטטיסטית בניתוח?

כל החלטה קטנה היא עוד הזדמנות למצוא תוצאות מובהקות, ובאופן הזה קל יותר להפוך תוצאה כמעט מובהקת לתוצאה מובהקת.

הבעיה (האיום) - הנסיין עלול לנסות (אפילו באופן לא מודע ולא מכוון) להשפיע על תוצאות הניתוחים הסטטיסטיים כך שיתאימו להשערתו (שיש/אין קשר).

- אם החוקרים משערים שיש קשר בין המשתנים, בעקבות דרגות החופש של החוקר נוצר ניפוח אלפא - אנחנו עושים שימוש בקריטריון שלא מייצג באמת את הסיכוי לטעות מסוג ראשון - משתמשים כל פעם באלפא 0.05 ואנחנו בעצם עושים המון בדיקות עם הקריטריון הזה, ועלולים לגרום לגילוי של קשר כשהוא לא באמת קיים בגלל שעשינו הרבה מאוד ניתוחים.
- אם ההשערה היא שאין קשר, דרגות החופש של החוקר יכולות לגרום לטעות מסוג שני - אי גילוי של קשר שאכן קיים. בכל ניסוי יש סדרה של החלטות כאלה שכל אחת מהן יכולה להשפיע על התוצאה ועל ניפוח האלפא.

דוגמא - הוצאת נבדקים "חריגים". מהו חריג? - פסיכולוגיה 2-3 סטיות תקן מהממוצע. זאת בעיה, כי הוצאת נבדקים באופן הזה מנפחת לנו את הסיכוי לטעות מסוג ראשון – הסיכוי למצוא אפקט למרות שהוא לא קיים בפועל.

פתרונות להוצאת נבדקים חריגים:

- למחוק או לתקן ערכים שגויים (זאת בעיה כי מסירים נתונים)
- במקום סטיות תקן - ממוצע הסטייה של תצפיות מהחציון
- להחליט על הוצאת תצפיות חריגות מראש, ולרשום את הקריטריונים ((pre-registration
- לדווח תוצאות עם ובלי הוצאת התצפיות החריגות
- דיווח שקוף ומלא על הוצאת נבדקים

פתרונות לדרגות החופש של החוקר:

1. נסיין עיוור להשערות המחקר. הבעיה - הנסיין לרוב מבין את המחקר ורוצה לרצות את החוקרת, אז הוא מנתח את הנתונים בהתאם להשערות, כפי שהוא מנחש אותה.
2. רישום מראש של תוכנית ניתוח הנתונים - (pre-registration) זה לא מצמצם את הרצון אפילו התת מודע של הנסיין להשפיע על תוצאות המחקר, אבל זה מצמצם את המקומות בהם אפשר לעשות את זה, כי החלטנו כבר מראש על הרבה מההחלטות שיכולות להתבטא בדרגות חופש ואי אפשר לשנות את הנתונים בהתאם להטיה. *אי אפשר להחליט על הכל מראש, יש הרבה בלתי"מים (נבדק שהגיב בצורה לא הגיונית, מגפה שגורמת לעצירת הניסוי) .
3. הנתונים עיוורים לתנאי הניסוי - כדי לצמצם את השפעת הנסיין על התוצאות בניתוח הנתונים, לשנות את השמות של המשתנים לשמות "בדויים" (ולהחזיר רק לאחר הניתוח הסטטיסטי). למשל ניסוי שכולל גברים לעומת נשים, ניתן לתת לקבוצות שמות a, b ולנבדקים שמות מומצאים. כך נוכל להיות בטוחים שההחלטות של הנסיין בעת ניתוח הנתונים לא מושפעות מההשערות. זה מאפשר איזון בין הנוקשות של תכנון מראש לבין הטיית הנסיין. פותח פתח לטעויות שיוך ויש להיות זהירים.
4. הצהרה על טוהר המידות - המסר של ההצהרה מבהיר שהיינו ערים לאופציה של דרגות החופש של החוקר. גם ראוי לאפשר לקוראי המאמר גישה לחומרים של הניסוי שבהם עשינו שימוש (שאלונים, נתונים וכו'). זה חשוב גם בשביל שיוכלו לעשות רפליקציה לניסוי ולבדוק השערות נוספות .

איומים נוספים:

צריך לעשות הבחנה בין מחקר אישוש למחקר גישוש -

Confirmatory Research - מחקר אישוש

- יוצא מתוך תיאוריה (או תיאוריות)
- נעשה בד"כ כדי לבחון תיאוריה/יות
- יש השערות ספציפיות

Exploratory Research - מחקר גישוש

- בודק "מה יקרה אם"
- אין תיאוריות רלוונטיות דיין
- אין השערות ספציפיות

מחקר אישוש כיום יותר שכיחים ממחקר גישוש. באיומים על תוקף המסקנה הסטטיסטית יש השלכות סטטיסטיות על ההבחנה בין מחקר אישוש למחקר גישוש.

איומים במחקר גישושי

1. במחקר גישוש צריך יותר לחשוך בתוצאות כי יש סכנה של ניפוח אלפא (כתוצאה מהשוואות מרובות).

דוגמא- מערך 2 (מין הדמות) על 2 (סוג התמונה - פנים/גוף). תפעלו את מין הדמות וסוג התמונה - הציגו לנבדקים פנים של גבר/אישה או גוף של גבר/אישה וטקסט קצר עליהם. מדדו 6מדדים שונים - עד כמה לדעתך הדמות מסוגלת לתכנן, לשלוט בעצמה, להתנהג מוסרית, לחוש הנראה, לחוש רב, לחוש תשוקה .

במערך כזה יש חשיבות מאוד גדולה להאם תכננו מערך אישוש או גישוש בנייתוח התוצאות. אם אנחנו רוצים לבדוק את ששת המדדים התלויים אנחנו יכולים לעשות 6 ניתוחי שונות דו גורמים -אחד לכל אחד מהמדדים. במקרה כזה גם אם שני המשתנים הבלתי תלויים לא משפיעים על אף אחד מהמשתנים התלויים, יש סיכוי לא רע שנמצא אפקט מובהק כלשהו (עיקרי או אינטראקציה). (במחקר גישוש כל תוצאה מובהקת יכולה להיות מעניינת, אבל תוקף המסקנה הסטטיסטית של תוצאה מובהקת במבחן אחד יהיה נמוך יחסית, כי היה כאן ניפוח אלפא. יש 18 מבחנים סטטיסטיים לכל אחד מהמדדים, אז הסיכוי שלפחות מבחן אחד יצא מובהק הוא סיכוי של 60% .

לעומת זאת - אם החוקרים תכננו מראש את התיאוריה להתמקדות בגוף, זה מגדיל את הנטייה לייחס תבניות פחותה לאדם וחייתיות מוגברת. אז יש השערה כיוונית ברורה -בתנאי הפנים הדמות תדורג כבעלת יכולת גבוה יותר לתכנן, לשלוט בעצמו ולהתנהג באופן מוסרי מאשר בתנאי הגוף. במחקר אישוש הסיכוי למצוא בדיוק את מה ששיערונו קטן יותר מהסיכוי למצוא אפקט כלשהו, בהשוואה למחקר גישוש, אז יש איום קטן יותר על התוקף הסטטיסטי במחקר אישוש מאשר במחקר גישוש.

פתרונות:

1. שימוש באלפא קטנה יותר
2. שחזור

- HARKing - hypothesizing after results are known. 2. בעולם יש מי שטוען שגם אם הממצאים שהשגנו הם מפתיעים, אז בסדר חשוב לחשוב בדיעבד על השערות שתומכות בתוצאות, ולדווח על התוצאות כאילו זה מה שניבאנו.

אבל אחרים טוענים (וניב מסכים איתם) שזה נוהל פסול שנקרא HARKing וזה פוגע בתוקף המסקנה הסטטיסטית -כי אנחנו מציגים מחקר גישוש בדיעבד כמחקר אישוש כי אנחנו מוצאים הסבר בדיעבד לתוצאות שמוצג כניבוי שלנו, זה סוג של זיוף. יש מי שטוענים שאפשר לעשות את זה רק במידה ושחזרנו את התוצאות אחרי הניסוי הראשון. אבל באופן כללי עדיף להיות ישרים ולספר שלא ניבאנו את המסקנות מראש ולהציג את השתלשלות האירועים בסדר בו היא קרתה, מהסיבה שהתיאוריה יותר משכנעת אם היא אכן הביאה לניבוי מראש מאשר אם היא מסבירה בדיעבד תוצאות שלא שיערונו. האיכות של תיאוריה נבחנת בעיקר ביכולת שלה לנבא מראש קשרים חדשים במציאות, ולא ביכולת שלה להסביר בדיעבד קשרים שנמצאו (פוסט הוק).

HARKing פוגע במסקנה הסטטיסטית כי אנחנו מאמינים יותר לתוצאות שנגזרו מתיאוריות, וזה גורם לנו להאמין בממצא כזה יותר מאשר אם הוא היה מוצג כממצא שנמצא במחקר למרות שלא שוער עליו. האמון שלנו במסקנה הסטטיסטית במקרה כזה גבוה יותר ממה שהיה צריך להיות.

פתרון: רישום מראש של ההשערות כדי להימנע מה"פיתוי". - pre-registration.

סיכום איומים על תקפים ופתרונות:

סוג התוקף	איום	פתרון
תוקף חיצוני	מדידה	1. למדוד ללא ידיעת הנבדקים (מדידת לא תגובתית) 2. למדוד באופן אינטגרלי בחיי הנבדקים
	הסטוריה	
	המצב הניסויי	1. לדמות את הניסוי למציאות כמה שניתן 2. ניסוי ללא ידיעת הנבדקים (לא אתו)
	דגימה	

Unconfounding באמצעות התאמה/השוואה	קונפאונד במניפולציה	תוקף מבנה
reverse coding - הוספת פריטים בקידוד הפוך	מערך הסכמה א+ב	
1. הוספת תנאי אנונימיות 2. להשתמש במדד שאינו דיווח עצמי	רצייה חברתית	
1. נסיין עיוור להשערה ו/או לתנאי 2. נבדק עיוור להשערה ו/או לתנאי 3. לשקר לנבדקים (לא אתי)	רציית הנסיין	תוקף פנימי
1. נסיין עיוור להשערה ו/או לתנאי 2. לשקר לנסיין לגבי ההשערה	הטיית הנסיין	
	ארטיפקט סטטיסטי	
1. נסיין עיוור להשערה ו/או לתנאי 2. pre-registration 3. נתונים עיוורים לתנאי 4. הצהרת טוהר מידות	דרגות החופש של החוקר	תוקף המסקנה הסטטיסטית
1. שימוש באלפא קטנה יותר 2. שחזור	ניפוח אלפא במחקר גישושי	
pre-registration	במחקר גישושי HARKing	

מדידה

מהימנות המדידה

בכל סוג של מחקר כמותי יש לפחות משתנה נמדד אחד, גם אם את המשתנה הבלתי תלוי אנחנו מבצעים בעזרת מניפולציה, את המשתנה התלוי אנחנו נמדוד, ולכן שאלת המדידה היא רלוונטית לכל מחקר.

מדידה = ייחוס שיטתי של מספרים למאורעות או אובייקטים.

שני קריטריונים להערכת כלי מדידה:

1. **מהימנות reliability** – עד כמה המדידה נקייה מטעויות מקריות. מהימנות היא תכונה של משתנה אופרציונלי, לא של משתנה תיאורטי. מהימנות של משתנה תיאורטי יכולה להיות שונה במדידות שונות (לדוגמה - בדיקת קורונה בעזרת מטוש בעלת מהימנות X ובדיקה של קורונה באמצעות בדיקת דם בעלת מהימנות Y).
2. **תקפות validity** – עד כמה המדידה מודדת את מה היא "אמורה" למדוד. התקפות היא למעשה תוקף מבנה של המשתנה הנמדד - עד כמה המדד בו אנחנו משתמשים (המשתנה האופרציונלי) מתאים למשתנה התיאורטי שאנחנו רוצים למדוד.

בלי מהימנות ובלי תקפות לא נדע שאנחנו מודדים את מה שאנחנו רוצים למדוד.

יש קשר מתמטי בין מהימנות לתוקף – מהימנות מגבילה את חוזק התוקף. שני הדברים הללו תלויים- התקפות תלויה

מקורות השונות במדידה

- שונות מקרית – ההבדלים בין התצפיות שהם מקריים. זו שונות שאנחנו רוצים לצמצם.
- שונות שיטתית – ההבדלים בין התצפיות הנובעים מההבדלים בתכונה/ סט תכונות כלשהן.

מהימנות מגבילה את תוקף המדידה, אם המדידה לא מהימנה אז בהגדרה יש בה הרבה טעויות שנובעות משונות מקרית, והן לא יכולות לייצג שום משתנה. ברגע שהכל מושפע משונות מקרית המדד לא מודד שום דבר. מהימנות היא בעצם מדד של עד כמה המדידה מגיעה מטעויות מקריות. מדידה עם מהימנות אפסית תמדוד רק רעש, בלי שום שונות שיטתית. מכאן, היא לא תקפה (היא לא מודדת את מה שהיא אמורה למדוד).

קל לזכור:

- מהימנות - האם יש גורם שיטתי כלשהו שמשפיע על המדידה?
- תוקף - האם הגורם השיטתי הזה הוא הגורם שאנחנו רוצים שהמדידה תמדוד?

טעויות מדידה

טעות מדידה זו סטייה של הערך הנמדד מהערך האמיתי שלו. חשוב להבין שלא כל טעות מדידה נובעת משונות מקרית, כדי להבין למה צריך להפריד בין שני סוגים של טעויות מדידה –

1. טעות מקרית – סטיות בלתי צפויות לכל הכיוונים האפשריים מהערך האמיתי. הטעויות המקריות לא ניתנות לניבוי והן לא נגרמות ע"י גורם קבוע אחד. הן יכולות להיגרם ע"י שינויים זמניים במצב האובייקט הנמדד, הפרעות סביבתיות שונות וכו'. יש סדרה של דברים שיכולים להשפיע באופן מקרי על המדידה, ומה שמשותף ביניהם הוא שהם מקריים.

אם גורם מסוים משפיע על כל הנבדקים באופן דומה ולא מקרי בכל מדידה ומדידה (למשל עייפות גורמת לכולם לטעות באופן דומה) אז זו לא טעות מקרית אלא טעות שיטתית. במקרה של טעויות שיטתיות- אין פגיעה במהימנות, יש פגיעה רק בתוקף.

2. טעות שיטתית – ישנה טעות בציון של הנבדק, אבל הטעות תמיד זהה או תמיד באותו כיוון.

לדוגמא - אם נרצה למדוד גובה של ילדים, אנחנו יכולים לבקש מהם להימדד במד גובה אבל טעות שיטתית היא כאשר שבחנו לבקש מהם להוריד את הנעליים ואז הטעות תמיד תהיה לכיוון למעלה (הגובה יהיה יותר גבוה) אבל תשתנה בהתאם לגובה הסוליה של אותה הנעל. אם לכל הילדים יש את אותו גובה של סוליה, אז הטעות תהיה טעות קבועה – מספר קבוע שנוסף או יורד מהמדד.

טעות שיטתית היא לא שונות שיטתית, טעות שיטתית היא קבועה, אין בה שונות.

טעויות מקריות וטעויות שיטתיות הן בלתי תלויות זו בזו, הן יכולות להתרחש ביחד או לחוד.

תיאוריית המדידה ששולטת היום נקראת "תורת המבחנים הקלאסית", מניחה שאם יש לנו טעויות גם אם הן שיטתיות אבל הן מגיעות מגורמים לא ידועים, אז כל הגורמים הלא ידועים הללו גורמים ביחד בממוצע לטעות אקראית, כלומר לכך שתוחלת הטעות מעבר לכלל הבדיקות תהיה 0 ואז אפשר לחשוב עליהן כטעות מקרית- שהתוחלת שלה היא אפס אבל השונות שלה גדולה מ-0 ומשתנה בהתאם למדידות שונות.

הציון האמיתי של אדם במדד כלשהו הוא הציון שלו ללא טעויות מדידה. במדידה בודדת מוזדדים ציון אמיתי + טעות.

ממוצע הציונים על פני מדידות רבות יהיה קרוב לציון האמיתי, כי הטעות המקרית תסכם ל 0 (כלומר, בתוחלת הציון הנמדד יהיה קרוב מאוד אם לא בדיוק הציון האמיתי).

מהימנות מוגדרת כתשובה לשאלה "עד כמה המדידה מושפעת מטעויות מקריות?", כלומר מה שמעניין אותנו בהקשר של מהימנות הן הטעויות המקריות.

מקרה לדוגמא - נמדוד את היכולת המילולית של אדם באמצעות ציון במבחן פסיכומטרי מילולי. נניח שלכל אדם יש ציון אמיתי

שאותו היינו מסוגלים למדוד אם לא היו טעויות מדידה.

אם מהימנות המדד היא מושלמת, אם לאדם א' יש ציון נמוך יותר מאדם ב', מדידה אמיתית תראה רק את הקו האנכי בגרף, בכל מדידה נקבל את הציון האמיתי ונדע להפריד בין הציונים.

אבל אם מהימנות המדד אינה מושלמת, כשנמדוד את הציון של האדם נקבל ציון שונה בקצת מהציון האמיתי שלו. אם נערוך את המדידה שוב ושוב עבור אותה תכונה (למשל כושר קריאה) נקבל לפעמים סטיות חיוביות (ימינה) או סטיות שליליות (שמאלה) ולכן לכל אדם תתקבל התפלגות של ציונים סביב הציון האמיתי.

סטיות מקריות מתקזזות כך שממוצע המדידות של כל אדם יהיה הציון האמיתי שלו, כי הטעויות הן מקריות, ממוצע הטעות המקרית הוא 0.

התפלגות הציונים מסביב לציון האמיתי יכולה להיות צרה או רחבה - פיזור גדול או קטן. פיזור רחב של ציונים, או שונות גדולה, שקולים כנגד מידה רבה של טעות. פיזור קטן, התפלגות צרה זו מדידה נמוכה של טעות.

השונות של הציונים היא מדד למידת הטעות המיוצגת במדידה שלפנינו וזו נקראת "טעות התקן" של המדידה. גודל טעות התקן הוא מדד ישיר של הטעות המעורבת בשימוש במכשיר המדידה המסוים, ככל שהמדידה מהימנה יותר ההתפלגות צרה יותר. במדידה מושלמת אין בכלל טעות ולכן ההתפלגות הכי צרה שיכולה להיות ואין שונות.

$$x = x_T + e$$

ציון נצפה ← x ← x_T ציון אמיתי ← e טעות מקרית

ציון של אדם מורכב מציון אמיתי ומגורם טעות .

תכונות הטעות המקרית לפי תורת הטעות הקלאסית:

- תוחלת הטעות (e) שווה 0. כי הטעויות הן מקריות ומעבר להרבה מדידות הן יתקזזו.
- שונות הטעויות (S) (בריבוע) שונה מ0. שונות הטעות המקרית תהיה 0 רק כשאין טעות מקרית בכלל, וברגע שיש טעות מקרית יש שונות בטעות המקרית.
- המתאם בין הטעויות לבין משתנה אחר (r) שווה 0. זה מה שאומר שהטעויות הן מקריות ולא ניתנות לניבוי.

$$s_x^2 = s_T^2 + s_e^2$$

שונות הציונים הנצפים ← s_x^2 ← s_T^2 שונות הציונים האמיתיים ← s_e^2 שונות הטעויות

מכאן נובע שהשונות של התפלגות הציונים הנצפים, היא הסכום של שני רכיבים: שונות הציון האמיתי ושונות הטעות.

זה קורה כי שונות של סכום ציונים (משתנים) שווה לסכום השונות של כל אחד מהם פחות השונות המשותפת שלהם (זה נובע מהתכונה השלישית מתורת המבחנים הקלאסית). זה תלוי בהנחה שטעויות המדידה לא מנבאות שום דבר.

יש לתיאורית המבחנים הקלאסית אלטרנטיביות מודרניות, אבל יותר קל להבין את הבסיס של טעויות למידה באמצעות התיאוריה הזאת.

חישוב מהימנות

*לא צריך לזכור את החישובים, המטרה היא לתת אינטואיציה איך לגשת למדידת מהימנות.

מהימנות כמתאם בין שתי מדידות

דרך מקובלת לבדוק מהימנות של מדד היא לחשב את המתאם בין מדידות חוזרות. למשל- למדוד את כל המדגם פעמיים (עם מרווח זמן מסוים ביניהם) ולבדוק עד כמה שתי המדידות דומות זו לזו. ככל שהמתאם בין שתי המדידות יהיה גבוה יותר כך המדידה תהיה מהימנה יותר, כלומר שיעור הטעויות המקריות יהיה קטן יותר.

מה יקרה כשהמדידה מושלמת? – מדידה מושלמת, מהימנה לגמרי ללא טעויות מקריות תיתן לנו את אותו הציון לכל נבדק בכל המדידות. כלומר, המתאם בין שתי המדידות החוזרות יהיה בדיוק 1.

מדידה גרועה באופן מוחלט - היא אקראית לחלוטין, בעלת המהימנות הכי גרועה. המתאם בין שתי המדידות במצב זה יהיה 0 (בתוחלת של כלל המתאמים).

פעמים רבות לא ניתן לחזור על אותה מדידה בדיוק (למשל במבחנים מסוימים, הנבדקים כבר יכירו את השאלות), במקרה כזה

צריך לייצר מבחנים מקבילים - שתי מדידות או יותר לאותם נבדקים, שהן שונות אופרציונליות אך זהות תיאורטית (למשל מבחן מועד א' ומועד ב.')

למה המתאם בין שתי המדידות נותן מדד למהימנות המדידה?

תזכורת - מתאם:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n \cdot s_x \cdot s_y} = \frac{cov(x, y)}{s_x \cdot s_y}$$

מתאם הוא השונות המשותפת חלקי מכפלת סטיות התקן של שני המדדים, כשהשונות המשותפת היא ממוצע מכפלת ההפרשים של כל ציון מהממוצע של המדגם שלו.

מתאם הוא הקשר בין שני משתנים, אם יש קשר חיובי חזק בין שני המשתנים זה אומר ששני המשתנים עולים/יורדים יחד - X עולה/יורד אז גם Y עולה/יורד. בקשר שלילי בין שני משתנים אחד מהמשתנה עולה והשני יורד - X עולה אז Y יורד ולהפך - X יורד אז Y עולה. בחישוב המתאם מחלקים את השונות המשותפת במכפלת סטיות התקן כי אנחנו צריכים להיות מסוגלים להיות בלתי תלויים ביחידות שבהן אנחנו עושים שימוש במדד, אנחנו לא רוצים שמדידה אחת תשפיע על חישוב המתאם ולכן אנחנו "מתקנים" את הציונים.

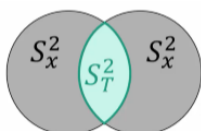
אנחנו רוצים לדבר על ציון באותו מבחן שמודדים אותו בשתי פעמים נפרדות:

	שונויות במבחנים	ציונים במבחנים
שני הציונים בנויים מהציון האמיתי ועוד ציון הטעות.	$s_{x_1}^2 = s_T^2 + s_{e_1}^2$	$X_1 = X_T + e_1$
השונות של שני הציונים מורכבת משונות הציון האמיתי ועוד שונות הטעות.	$s_{x_2}^2 = s_T^2 + s_{e_2}^2$	$X_2 = X_T + e_2$

יש לשים לב כי הטעויות שונות אולם הציונים האמיתיים זהים. המבחן הוא אותו מבחן ולכן אנחנו מניחים שהציונים זהים בשניהם, ומה ששונה הוא הטעויות.

המתאם בין שתי המדידות -

$$r_{xx} = \frac{\sum(X_{1i} - \bar{X}_1) \cdot (X_{2i} - \bar{X}_2)}{n \cdot s_1 \cdot s_2} = \frac{cov(X_1, X_2)}{s_1 \cdot s_2}$$



אפשר לפתח את הנוסחה הסתמכות על העובדה שהשונות של כל מדידה שווה לשונות הציונים האמיתיים ושונות הטעות, שהיא שונה ממדידה למדידה. מהדבר הזה נובע שהשונות המשותפת בין שתי המדידות היא שונות הציונים האמיתיים בלבד.

מכאן - מהימנות (מתאם בין שתי מדידות) היא שונות הציונים האמיתיים חלקים מכפלת סטיות התקן של שתי המדידות -

$$r_{xx} = \frac{s_T^2}{s_1 \cdot s_2}$$

אפשר להניח אם מדובר במדידות חוזרות אבל גם אם מדובר במבחנים מקבילים, שהשונות של הציונים האמיתיים שלהם זהות אבל גם שהשונות של הטעות שלהם זהה.

בגלל שיש לנו שני מדדים לאותה תכונה, אנחנו יכולים להניח שהשונויות שלהן שוות, ואז אפשר להגיד שהמתאם בין שני המבחנים שווה לשונות הציונים האמיתיים חלקי שונות המדד כולו -

$$r_{xx} = \frac{s_T^2}{s_X^2}$$

מקודם הראנו ששונות הציונים הנמדדת שווה לסכום של השונות האמיתית של הציונים ועוד שונות טעויות המדידה, אז במילים אחרות המתאם בין שתי מדידות מבטא את היחס בין השונות של הציונים

$$r_{xx} = \frac{s_T^2}{s_T^2 + s_e^2}$$

האמיתיים לבין כלל השונות של הציונים הנמדדים (כולל את השונות של הציונים האמיתיים ואת שונות טעויות המדידה). כלומר, מתוך כלל השונות במדד יש שונות שנגרמת מגורם כלשהו (הציון האמיתי) ויש שונות שנגרמת כתוצאה מטעות. המדד מהימן יותר ככל ששונות הטעות קטנה יותר והשונות המוסברת על ידי המשתנה שרצינו למדוד גדולה יותר. כלומר, ככל שחלקה של השונות המוסברת מכלל הטעות גדול יותר כך המהימנות גדולה יותר.

לדוגמא - אם מצאנו מהימנות של 0.8 זה אומר ש%80 מהשונות במדד הזה מוסברת על ידי ציונים אמיתיים.

חשוב מאוד להבין שראינו כאן פיתוח תיאורטי כשאנחנו מחשבים את המתאם בין שתי מדידות במדגם אנחנו רק אומדים את המהימנות, כל הפיתוחים שנראו כאן נכונים רק מעבר לאינסוף מתאמים בין שתי מדידות שונות אם מצב כזה היה אפשרי. אבל בתוחלת הפיתוח הזה בהחלט נכון - המתאם בין שתי מדידות הוא היחס בין שונות הציונים האמיתיים לשונות הכללית שמורכבת משונות הציונים האמיתיים ושונות הטעות. הפיתוח הזה הראה לנו איך מתאם בין שתי מדידות בעצם משקף לנו ציון מהימנות, עד כמה השונות השיטתית מסבירה את השונות שאנחנו רואים במדד?

מהימנות כחסם לתוקף עליון

$$r_{xx} = \frac{S_T^2}{S_X^2}$$

עוצמת הקשר עם מדדים אחרים מוגבלת על ידי מהימנות.

תזכורת: מקדם המהימנות הוא החלק של שונות הציון האמיתי מכלל שונות המדידה -

האם מתאם עם מדד אחר יכול להיות גבוה יותר ממקדם המהימנות? או במילים אחרות, האם ייתכן שמתאם בין X ל Y יהיה נמוך יותר ממתאם בין X ל Y? התשובה היא כן. זה יכול לקרות כתלות בקשר בין הציונים האמיתיים של שני המדדים וגם במהימנות של כל אחד משני המדדים. אם למדד השני יש מהימנות טובה יותר זה יכול לקרות.

הטעות המקרית פוגעת ביכולת של המדידה להיות במתאם עם מדידה אחרת, כי לטעות מקרית אין מתאם עם שום דבר. אז במתאם בין מדידות חוזרות, אותה שונות של טעות מקרית פוגעת במתאם פעמיים - פעם במדידה הראשונה ופעם במדידה השנייה. אם נחליף את המדידה השנייה במדידה מושלמת שאין בה שום שונות של טעות, אז השונות המשותפת של שתי המדידות תגדל, כלומר המתאם יגדל גם הוא.

* זה כמובן לא אומר שהמתאם של המדידה עם כל מדידה של משתנה אחר עם מהימנות מושלמת, יהיה גבוה מהמהימנות של המדידה המקורית. יכול להיות שאין באמת קשר בין שני המשתנים.

כמה גבוהה יכולה להיות הקורלציה של מדידה עם מדידה אחרת (**Rxy**) בהינתן מהימנות מסויימת - **(Rxx)** ברור לנו שמהימנות המדידה צריכה להגביל את המתאם בין המדידה לבין מדידות אחרות, עכשיו ננסה לחשב את החסם העליון שהיא מציבה על מתאם שכזה, כלומר ננסה לחשב את ה Rxy המקסימלי, כאשר Y הוא מדידה היפותטית שהמתאם שלה עם המדידה שלנו - X הוא המתאם הגבוה ביותר ש X היה יכול לקיים עם מדידה אחרת.

איזה Y יאפשר Rxy מקסימלי? -

1. צריך שהמתאם האמיתי בין X ל Y ללא טעויות מדידה יהיה מקסימלי (1) - הקשר בין הציונים האמיתיים של X ו- Y יהיה מושלם. זה קורה בד"כ אם $X = Y$ - הם אותו משתנה.
2. למדידה של Y צריכה להיות מהימנות מושלמת (1).

כלומר, המועמד הטוב ביותר ל- Y הוא הציונים האמיתיים של X. כלומר, המתאם המקסימלי שיכול להיות למדידה מסוימת של X עם איזשהו משתנה אחר הוא מתאם בין המדידה של X לציון האמיתי של X.

$$r_{tx} = \frac{cov(X_T, X)}{S_T \cdot S_X}$$

כדי לחשב את המתאם בין הציונים האמיתיים לציונים הנצפים במדידה נשתמש בנוסחה:



$$cov(X_T, X) = s_T^2$$

שונות הציון הנמדד תמיד מורכבת משונות הציון האמיתי ועוד שונות הטעות המקרית. שונות הטעות המקרית בטוח לא קשורה לשונות הציון האמיתי, אז השונות המשותפת בין הציונים

האמיתיים לנמדדים היא שונות הציון האמיתי.

מזה אפשר להגיע לנוסחה -

$$r_{tx} = \frac{S_T^2}{S_T \cdot S_X} = \frac{S_T}{S_X}$$

המתאם בין הציונים אמיתיים
הציונים חלקי מכפלת סטיות התקן
יכולים לצמצם את סטיית התקן
שהמתאם בין מדידה לציון האמיתי הוא היחס בין סטיית התקן של הציון האמיתי לסטיית התקן של הציון הנמדד וזה המתאם המקסימלי בין ציון מדידה לכל מדידה אחרת.

עכשיו אנחנו רוצים להשוות את המדד ההיפותטי למהימנות המדד שמעניין אותנו עם המתאם המקסימלי של המדד שמעניין אותנו. עוצמת הקשר המקסימלי בין מדד אחד למדד אחר, היא בעצם שורש המהימנות של אותו מדד.

$$r_{tx} = \sqrt{r_{xx}}$$

המתאם בין המדד לבין הציון האמיתי שלו:
הראנו קודם שהמתאם בין שתי מדידות של אותו משתנה:
שורש של מספר שקטן מ-1 (למשל אם מקדם המהימנות המקסימלי בין המדידה לכל שורש של 0.64).

מוביל למספר גדול יותר
הוא 0.64 אז המתאם
מדידה אחרת יהיה 0.8

זה מראה שמדידה יכולה להיות במתאם גבוה יותר עם מדידה אחרת מאשר עם עצמה, במידה והמהימנות של המדידה האחרת טובה מספיק. אם המהימנות של מדד מסוים נמוכה, יהיה קשר למצוא קשר למדדים אחרים, אז המתאם המקסימלי בין המדד שלנו לאחר יהיה נמוך.

זהו גם התוקף המקסימלי - נרחיב על זה בהמשך.

עוצמת הקשר עם מדד אחר מוגבלת ע"י המהימנות. כדי שמדד יהיה מהימן הוא חייב להיות בעל שונות, אם אין שונות במדד אז אי אפשר לייחס הבדלים במדידה להבדלים בתכונה הנמדדת. אז בעצם, מהימנות מוגבלת ע"י אפקט רצפה/תקרה - אפקט המתקבל כאשר המדידות נוטות להיות קרובות לחסם התחתון/העליון של הציונים האפשריים. מהימנות מוגבלת על ידי חסמים כאלה.

דוגמא - בבגרות במתמטיקה, הוחלפו בטעות המבחנים במבחני חשבון לתלמידי כיתה ג' (המבחנים היו ממש קלים) סביר שאם תהיה שונות בציונים היא תנבע כולה מטעויות מדידה מקריות. זה דוגמא לאפקט תקרה, כל הציונים יהיו גבוהים מאוד, ואם תהיה שונות זה יהיה מקרי לחלוטין. לכן, יהיה לנו קשה להעריך את המהימנות, כי אין לנו דרך להעריך את שונות המדידה, ואז המהימנות תהיה נמוכה ולכן גם התוקף יהיה נמוך. פתרונות לכך היא להשתמש במדד רגיש יותר שיפיק יותר שונות, או להשתמש באוכלוסייה אחרת (למשל האוכלוסייה לה נועד המבחן במקור - תלמידי כיתה ג')

הדוגמא הפוכה לכך היא תלמידי כיתה ג' שקיבלנו בטעות בגרות למתמטיקה. זו דוגמא לאפקט רצפה - כל הציונים יהיו נמוכים מאוד כי התלמידים לא יודעים איך לפתור את המבחן.

כשאין שונות במדד לא נוכל לחשב מהימנות וזה גם יגביל את התוקף.

לקח נוסף מהדוגמאות הללו הוא שצריך להתאים את המבחן לאוכלוסיית הנמדדים.

סוגי מהימנות

יש כמה דרכים לבדוק מהימנות של מדד, על 2 הראשונות כבר דיברנו:

1. מבחן חוזר **test-retest** - מדידה חוזרת של אותו מדד בדיוק. ככל שהמתאם בין המדידות גבוה יותר כך גם המהימנות גבוה יותר.

בעיות במדידה חוזרת:

- השפעות של המדידה על איכות המדד carryover - למשל אם מדובר במבחן יכולת שבו מודדים כמה זמן לוקח

לפתור שאלות אז בפעם השניה הנבדקים יענו על המבחן יותר מהר, במקרה זה השונות השיטתית מצטמצמת בגלל שהתשובות מבוססות על זיכרון טוב של השאלות ולא על יכולת. זה קורה במדדים תגובתיים. במקרה כזה המדידה השניה תהיה כושלת, ולכן תעריך הערכת יתר את המהימנות ויכול להיות שלא יהיה מתאם משמעותי בין שתי המדידות ולא נוכל ללמוד מכך על מהימנות המדד.

- לפעמים אנחנו לא רוצים לבדוק את איכות המדד אלא את איכות שיטת המדידה, אנחנו רוצים לדעת אם הדרך בה חיברנו את המבחן היא דרך נכונה. ואז אנחנו צריכים להרכיב יותר ממבחן אחד ולמדוד את הקשר בין שני המבחנים האלה (זה מוביל אותנו לדרך נוספת למדוד מהימנות)

2. טפסים/ מבחנים מקבילים - שני מדדים או יותר שאמורים למדוד את המשתנה התיאורטי באופן דומה. נהוג להגיד ששני המבחנים הם מקבילים אבל לא זהים.

בשני המקרים האלו כשמדובר במדידה חוזרת אחת בלבד, חישוב המהימנות נעשה כמו שלמדנו עד עכשיו.

3. מבחן חצוי - split half method - שיטה שדומה למבחנים מקבילים, אבל במקום לבצע שתי מדידות, חוצים את המבחן ל2 ובודקים מתאם בין שני חלקי המבחן. המקרה הזה מצריך הנחה מוקדמת ששני חלקי המבחן מצליחים למדוד את כל המשתנה התיאורטי.

לדוגמה - מודדים את המשתנה התיאורטי "דיכאון" שנמדד באמצעות שאלות. אין שאלה אחת שעוזרת לנו לדעת כמה האדם דיכאוני. נוכל להשתמש בשיטת המבחן החצוי רק אם אנחנו מאמינים שאפשר לחצות את המבחן לשניים בצורה כזאת שבכל חצי יהיו מספיק פריטים כדי למדוד דיכאון.

- אורך מגדיל מהימנות - ככל שיש יותר פרטים שמרכיבים את הציון, הסיכוי שהטעות המקרית תשפיע על ציון המבחן קטנה.

לדוגמה - האם נעדיף מבחן שיש בו הרבה שאלות או מבחן בעל שאלה אחת בלבד? אם יש רק שאלה אחת הסיכוי שנקבל ציון שלא משקף את היכולות האמיתית שלנו גדול ולכן עדיף הרבה שאלות. זה אותו היגיון שעומד מאחורי העובדה שמדגם גדול יותר משקף יותר טוב על האוכלוסייה.

במקרה של מבחן חצוי, המתאם בין שני חלקי המבחן הוא לא ההערכה המדויקת של מדד המהימנות, בגלל שבמבחן ארוך יותר יש סיכוי טוב יותר להיות מהימן ואז אם נבחן את המתאם בין שני החצאים נקבל הערכת חסר של המהימנות של המבחן כולו. אז צריך ליצור תיקון לבעיה הזאת, זה קורה באמצעות נוסחת ספירמן-בראון.

הרעיון הכללי הוא להשתמש במתאם בין שני החצאים כדי לנבא את המתאם שהיה למדידה כולה עם מדידה היפותטית דומה באורך זהה. משתמשים בנוסחה הזאת כשרוצים לחשב את המהימנות של המבחן כולו באמצעות שני חצאים. יש לשים לב שהנוסחה מניחה זהות בין שני חצאי המבחן.

$$r_{xx} = \frac{2 \times r_{12}}{1 + r_{12}}$$

מהימנות המבחן כולו

מתאם בין שני החלקים

4. התאמה בין פריטי המבחן - הגישה הקלאסית - לקחת בחשבון את כל פריטי המבחן ולחשב את ההתאמה ביניהם, לבצע שכלול של המתאם בין כל הפריטים לבין עצמם. אפשר לחשוב על כך כאילו היינו לוקחים את כל החלוקות האפשריות של הפריטים לשני חצאים, ועושים ממוצע של מקדמי ספירמן-בראון לכל חלוקה אפשרית.

הנוסחה לביצוע התאמה בין פריטי המבחן: זאת שיטה שהומצאה על ידי קרונבך והיא נקראת גם האלפא של קרונבך.

$$\alpha_c = \frac{k}{k-1} \times \left(1 - \frac{\sum s_i^2}{s_x^2} \right)$$

Cronbach's α

מספר הפריטים במבחן

שונות פריט i

שונות כולל (המבחן) (סכום הפריטים)

(מקובל לטעון שאלפא קרונבך שווה לממוצע מקדמי ספירמן-בראון של כל החצאים האפשריים של המבחן, אבל זה נכון רק בהנחה שסטיית התקן של כל הפריטים זהה).

הבעיה באלפא קרונבך היא שבפועל היא לא מעריכה בצורה נכונה לגמרי את המהימנות. יש אלמנטים בנוסחה שגורמים להנחת חסר של המהימנות, ויש אלמנטים שגורמים להערכת יתר של השונות השיטתית של המבחן. כלומר, עם אלפא

קרונבאך לא ניתן דעת מה המהימנות האמיתית של המבחן הנמדד. הנוסחה מניחה שהציון האמיתי של כל אחד מהפריטים זהה, אין רגישות לתרומות שונות של פריטים שונים. בפועל, זה לא המצב והנוסחה נותנת לטעות מקרית "להיכנס בדלת האחורית." הבעיה הזאת ועוד בעיות גרמו לכך שסטטיסטיקאים ממליצים לא להשתמש בנוסחה הזאת, ובמקום זה מעדיפים להשתמש באומגה של מקדונלד - ω (ניתן לחשב בקלות ב jasp וכו'). לא נכנס לחישוב הזה.

5. מהימנות בין שופטים - כאשר המדידה היא הערכה של שני שופטים (או יותר). מספר דרכים לחשב:
- מתאם פשוט בין ציוני השופטים (מדד כמותי) - ככל שהמתאם גדול יותר המהימנות של המדידה גבוה יותר. כיוון שזה מעלה את הסיכוי שהשונות נובעת מהנבדקים ולא מהשופטים עולה.
 - מקדם Kappa (מדד איכותי) - מדד שבודק למשל את המתאם בין ציונים שהם לא מספריים, לא צריך לדעת איך לחשב אותו, פשוט חשוב לדעת שאפשר לחשב מתאם גם למדדים איכותניים (כמו סגנון התקשורת) ולא מספריים.

דרכים למדידת מהימנות - סיכום

- חשוב להבין, לא כל הדרכים למדידת מהימנות מוזדדות את אותו סוג של מהימנות, אפשר לדבר על שני סוגי מהימנות – מדידות חוזרות ועקיבות פנימית. יכולים להיות מבחנים בהן יש מהימנות נמוכה במדידות חוזרות אבל עקיבות פנימית גבוה או להפך. ישנם מקרים שבהם יהיה לנו חשוב יותר שהמדידות החוזרות יהיו מהימנות ופחות אכפת לנו מהעקיבות הפנימית ולפעמים להיפך. מצד שני, יש גם דמיון גדול בין שני סוגי המהימנות כי בשני המקרים מקדם מהימנות גבוה מרמז לנו שהמדד מושפע באופן גבוה מתכונה/סט תכונות כלשהו, כלומר שהוא לא בנוי מטעויות מקריות בלבד.
- שתי השיטות הראשונות (מבחן חוזר ומבחנים מקבילים) שייכות לקבוצה שמודדת מהימנות של מדידות חוזרות – עד כמה כשניתן לאותו אדם לעשות את אותה מדידה הוא יקבל ציון דומה.
 - השיטה השלישית והרביעית (מבחן חצוי והתאמה בין פריטי המבחן) מודדות עקיבות פנימית- עד כמה חלקים שונים של מדד מסוים מוזדדים בדיוק את אותו הדבר.
 - השיטה האחרונה- מהימנות בין שופטים, בהקשר הזה היא מקרה מיוחד שרלוונטי רק כששופטים הם מקור השיפוט שלנו, ואפשר לחשוב על מצבים בהם אפשר למדוד מדידות חוזרות של שופטים, וגם מצבים בהם אפשר למדוד עקיבות פנימית של שופטים.

המשך מדידה

תוקף המדידה

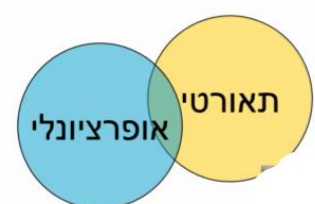
תוקף מבנה = מידת ההלימה בין ההגדרה האופרציונלית למבנה התיאורטי.

תוקף המדידה היא המידה בה המדידה מודדת את מה שאנחנו חושבים שהיא מודדת, או עד כמה ההגדרות האופרציונליות של המשתנה מייצגות בצורה נאמנה את המושגים התיאורטיים שהתכוונו לחקור. דרך נוספת לחשוב על זה היא לחשוב על חפיפה בין המשתנים האופרציונליים לבין המשתנים התיאורטיים.

במצב של תוקף טוב:

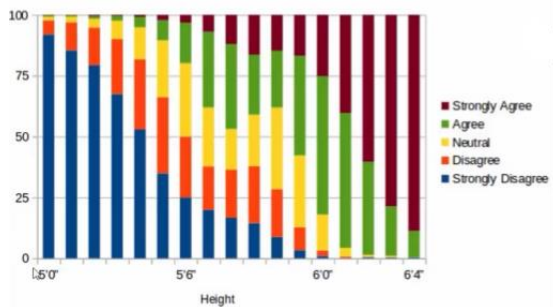


במצב של תוקף לא טוב:



נתמקד בתוקף מבנה של משתנה נמדד (ולא מתופעל).

תוקף היא תכונה של משתנה אופרציונלי (ולא תיאורטי). אופרציונליזציות שונות של אותו משתנה תיאורטי יכולות להיות בעלות תוקף שונה.



לדוגמא -משתנה תיאורטי גובה. מדד 1- מרחק בס"מ בין העקב לקודקוד הראש. מדד 2- ההסכמה עם ההיגד "אני גבוה." על ציר ה X רואים את הגובה האמיתי, על ציר ה Y אנחנו רואים את אחוז העונים לכל אחת מקטגוריות ההסכמה (מאוד לא מסכים, מאוד כן מסכים וכל מה שביניהם). אפשר לראות מהגרף שהמדד השני סובל מתוקף נמוך יותר, כי אנשים עם מענה דומה להיגד (ציונים דומים) אינם בהכרח בעלי אותו גובה.

תשובה אחת לא אומרת לנו מה הוא הגובה. בנוסף, באותו גובה יכולים להיות אנשים עם ציונים שונים. למשל ב- 180 ס"מ, אנשים שונים מסכימים בצורה שונה עם ההיגד לגבי ספציפית הגובה הזה. כנראה שההיגד הזה מודד משהו אחר שהוא לא הגובה ולכן יש לו בעיה בתוקף.

תיקוף – איך אנחנו בודקים תוקף?

למשתנים התיאורטיים במדעי החברה אין בד"כ הגדרה אופרציונלית חד משמעית, ולכן יש קושי בתיקוף המשתנים האלו. למשל - ההבדל בין גובה לגובה נתפס, אפשר לחשוב על כמה דרכים שונות למדוד גובה נתפס, וקל לשער שתהיה שונות גדולה יותר בין אנשים שיציגו מדד לגובה נתפס לבין המדד של הגובה שלהם. גובה הוא בעל יותר תוקף, וזה מדגים את האתגר של פסיכולוגיים ומדענים חברתיים בתיקוף מדידות. ישנן כמה שיטות לתיקוף לפי כמה גישות שונות.

הגישה הפונקציונלית

מקור הגישה בביהביוריזם הרדיקלי. צריך להסביר כל ממצא של ניסוי באמצעות מאפיינים של הסביבה שנצפתה על ידי החוקרים, צריך להסביר כל ממצא בניסוי באמצעות מה שאפשר למדוד.

לדוגמא - אנשים מדווחים שהם מכבדים יותר אדם כאשר הוא מצולם מלמטה. מסקנה לפי הגישה המקובלת - הגדלת גובה הנתפס של אדם תגדיל את הכבוד כלפיו. מסקנה פונקציונלית תהיה עדינה יותר - צילום אדם מלמטה יגדיל את הכבוד שאנשים מדווחים שיש להם כלפיו. הגישה הפונקציונלית שמה דגש על הגדרות באמצעות המדידה.

לכאורה, נראה שאין דרך להתקדם כשמשתמשים בגישה הזאת, כי קשה להכליל בניסויים בלי להשתמש במושגים תיאורטיים. אבל בפועל, מי שמתמש בשיטה הזאת כן מסוגל לייצר משתנים כללים יותר מהמשתנים האופרציונליים.

לדוגמא -

- ניסוי:1 אנשים שנחשפו למילים הקשורות לזקנה הולכים לאט יותר לאחר החשיפה.
- ניסוי:2 אנשים שנחשפו למילים הקשורות בפרופסורה מצליחים יותר בבחינת ידע.
- ניסוי:3 אנשים שנחשפו למילים הקשורות בחוצפה מפריעים יותר.

מסקנה מכלילה לפי הגישה המקובלת (האינטואיטיבית) - הגברת הנגישות של קבוצה חברתית בזיכרון גורמת להתנהגות הסטריאוטיפית של הקבוצה. לעומת זאת, מסקנה מכלילה לפי הגישה הפונקציונלית - חשיפה למילים הקשורות לקבוצה חברתית גורמת להתנהגות הסטריאוטיפית של הקבוצה, אין פה את החלק של המשתנה התיאורטי של נגישות בזיכרון.

הגישה הפונקציונלית מסיקה שיש קבוצת מניפולציות (חשיפה למילים שקשורות לקבוצה חברתית כלשהי) שגורמת לקבוצה של התנהגויות (שקשורות לאותה קבוצה חברתית).

הגישה הפונקציונלית לא טוענת שאי אפשר להשתמש במשתנים תיאורטיים, אלא שהם צריכים להיות תכונות משותפות/תכונות של גירויים ותגובות. למשל משתנה תיאורטי לפי הגישה הפונקציונלית יכול להיות גירוי שמופיע בתדירות גבוהה, אפשר למדוד אותו.

ניתן לטעון שההכללות האלה, "קבוצה של", הן משתנה תיאורטי, ולבדוק מה ההבדל בין הגישה הפונקציונלית לשאר העולם. לטענת החוקרים שתומכים בגישה, הגישה מבוססת על יותר ממצאים ומניחה פחות הנחות ולכן היא טובה יותר. כלומר- במקום להניח משתנים תיאורטיים ולחפש להם אופרציות כמו שעושים בגישה המקובלת, הגישה הפונקציונלית טוענת שצריך לבחון אופרציות (סדרה של אופרציות) ומהן להסיק משתנים מקבצים, שהפ למעשה משתנים תיאורטיים. בסופו של דבר

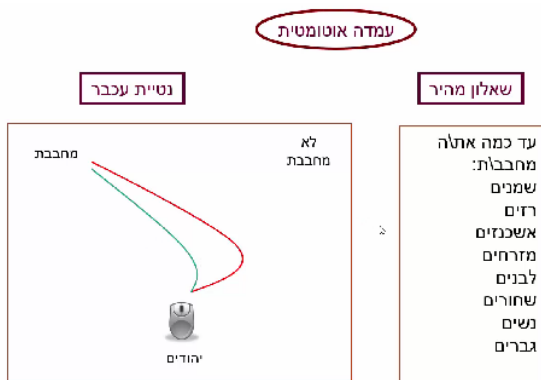
ההבדל בין הגישה הפונקציונלית לשאר העולם הוא שהיא מנסה לתת משקל לממצאים ולזכור כל הזמן שהמשתנים התיאורטיים הם השערות, והעובדות נוגעות למשתנים האופרציונליים. הגישה הזאת מהווה מיעוט בקרב קהילת החוקרים בפסיכולוגיה.

גישת הפשר החלקי

באופן מעשי, מי ששואף להיות פונקציונליסט מתקשה לעשות זאת כי עולם המדע מדבר במשתנים תיאורטיים. לפי הגישה שהציגו קרונברך ומיל ב-1955, שהיא הגישה המקובלת על רוב החוקרים היום, החוקרים מניחים משתנה תיאורטי ומתקפים אותו באמצעות אופרציות. הכוונה במתקפים היא שהם מוכיחים שהמבנה התיאורטי באמת קיים וניתן להראות השפעה שלו על משתנים אחרים. לכן, לא ניתן להפריד בין תיקוף האופרציה לתיקוף המבנה התיאורטי. תיקוף האופרציה אומר שצריך להראות שהיא מבטאת את המשתנה התיאורטי, ואי אפשר לעשות את זה בנפרד זה מזה.

בד"כ אחרי שאנחנו מייצרים משתנים אופרציונליים למשתנים תיאורטיים, אנחנו לומדים דברים חדשים על המשתנה התיאורטי עצמו ומעדכנים את ההבנה לגביו ולכן הגישה נקראת גישת הפשר החלקי.

לדוגמה - נניח שחוקרים משערים שקיים משתנה תיאורטי שנקרא "עמדה אוטומטית" (=הערכה לא נשלטת אל אובייקט כלשהו, כלומר הנטייה הלא נשלטת לשייך לאובייקט רגש חיובי/שלילי). המשתנה התיאורטי מוגדר כהערכה כלפי אובייקט אחר שמופעת ללא שליטת אדם. ניתן לפתח 2 מדידות - שני משתנים אופרציונליים.



מדד 1- שאלון מהיר- יש שניה אחת להשיב עד כמה מחבבים קבוצות חברתיות שונות (שמנים, רזים, אשכנזים מזרחים וכו. ('ההנחה היא שתגובה מהירה מושפעת יותר מעמדה אוטומטית מאשר מעמדה נשלטת.

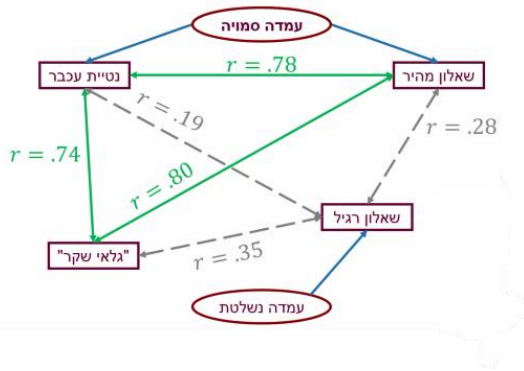
מדד 2 – נטיית עכבר- במדד הזה כשנראה שם של קבוצה חברתית נצטרך לסמן עם העכבר "מחבב/לא מחבב." ההנחה היא שהעמדה הנשלטת תשפיע על מה נבחר אבל העמדה האוטומטית תשפיע על המסלול שנעשה עם העכבר.



ההשערה שקיימת עמדה אוטומטית מביאה להשערה שיהיה מתאם חיובי בין שאלון מהיר לבין נטיית עכבר. אם באמת נמצא מתאם בין שני המדדים (למשל מתאם גבוה של 0.78) אז הממצא הזה יחזק את האמונה שעמדה אוטומטית קיימת וגם את האמונה שלנו ששני המדדים האלה באמת מודדים עמדות אוטומטיות. כלומר, התיקוף הזה משכנע אותנו שהמשתנה התיאורטי קיים וגם שהמשתנים האופרציונליים שהגינו אכן מודדים אותו.

נניח ונוסיף מדידת עמדות בעזרת שאלון רגיל בלי לחץ זמן, תוספת כזאת שוב תחזק את ההנחה שלנו שקיים משתנה תיאורטי של עמדה אוטומטית. אנחנו רואים שהמדדים המשוערים קשורים זה לזה, אבל אם הם לא יהיו קשורים לשאלון עמדות גלוי בצורה חזקה, זה יחזק לנו את האמונה שעמדה אוטומטית קיימת בלי קשר לעמדה נשלטת, וכך נחזק את האמונה שלנו שקיים

כזה דבר "עמדה אוטומטית". הוספת השאלון הרגיל פה זה תיקוף של משתנה תיאורטי עם תוקף מבחין.



אם נוסיף עוד מדידה עם גלאי שקר - בדיוק כמו השאלון הרגיל רק שמחברים לנבדקים מכשור פיזיולוגי ואומרים להם שזה גלאי שקר ועל כן הם צריכים להיות כנים ככל האפשר. זאת מניפולציה שנמצאה כמפחיתה שקרנות. כשנחשב מתאם עם המדדים נגלה שעמדה שאנחנו מודדים עם גלאי השקר קשורה יותר לעמדה האוטומטית מאשר מתאמי העמדה הנשלטת. יכול להיות שבגיע למסקנה שהמדדים האלה לא בדיוק מודדים עמדה שפועלת בלי שליטה, אלא מה שגלאי השקר יגלה הוא שהמדדים האלה מודדים עמדה שאנשים מחביאים- עמדה סמויה. כי מדדנו באמצעות אותו כלי כמו שאלון רגיל רק שהוספנו גלאי שקר, כך שזה לא בהכרח עמדה אוטומטית אלא עמדה סמויה.

זאת דוגמה למקרה שבו התיקוף של משתנה תיאורטי מלמד אותנו דברים חדשים על המשתנה התיאורטי וגורם לנו לעדכן אותו ואת ההבנה שלנו של כל מה שלמדנו עד עכשיו באמצעות המדדים. כך אנחנו מגיעים למסקנה שתיקוף של משתנה תיאורטי הוא בלתי נגמר - כל פעם שנוסיף אופרציה ונמצא עוד ממצאים, יכול להיות שנלמד דבר חדש על המשתנה התיאורטי ונעדכן אותו.

שיטות לתיקוף - תוקף המשתנה התלוי

באופן כללי ניתן וכדאי להשתמש בכמה שיטות שונות או בשילובים שונים לתיקוף מדד. להלן שיטות לתיקוף שעליהן נדבר:

1. תוקף תוכן ופנים
2. תיקוף באמצעות קריטריונים:
 - תוקף ניבוי
 - תוקף בו זמני
 - הבדלים בין קבוצות
 - הבדלים בין מאורעות
3. תיקוף באמצעות מדדים אחרים:
 - תוקף מתכנס
 - תוקף מבחין
 - ניתוח גורמים
4. חקר התהליך המשפיע על המדידה

תוקף פנים / Face validity

על פניו: האם המשתנה האופרציונלי נראה מתאים למשתנה התיאורטי? האם נראה לנו הגיוני שהמשתנה האופרציונלי מתאים למשתנה התיאורטי? אין שום כלי אנליטי, רק הגיון.

תוקף תוכן / Content validity

לפי מומחה, האם התוכן של המשתנה האופרציונלי מתאים למשתנה התיאורטי? מתאים= האם התוכן ממצה? בלבדי?

לדוגמה - אם מישוהו מחליט שהתשובה לשאלה תה/קפה היא מדד לדיכאון קליני, אז נראה שהשאלה הזאת סובלת מתוקף תוכן ותוקף פנים נמוך, כי לא הגיוני שהתשובה קשורה לדיכאון קליני וגם התוכן לא נראה מתאים.

ההבדל בין תוקף פנים לתוקף תוכן - תוקף פנים קצת יותר אינטואיטיבי וכללי מתוקף תוכן. יש מי שמשווה תוקף פנים לשאלה האם האדם הפשוט יסכים שהמשתנה האופרציונלי קשור על פניו למשתנה התיאורטי. לעומת זאת, תוקף התוכן יהיה קשור לשאלה האם מומחים בנושא יסכימו שהמשתנה האופרציונלי נראה קשור למשתנה התיאורטי. בנוסף, בתוקף תוכן הרבה פעמים מתכוונים לבחינה מפורטת יותר של הפריטים בשאלון על מנת לבדוק האם נראה שכל המשתנים התיאורטיים נמדדו בשאלון.

תיקוף עם קריטריון

בחינת קשר עם משתנים אחרים שאמורים להיות קשורים למשתנה התיאורטי. אין הרבה תועלת במשתנה תיאורטי שמנותק ממשתנים אחרים, זאת שיטה חשובה לתיקוף המדד והיא צריכה להיות מבוססת על תיאוריה לגבי אותו המשתנה התיאורטי, במיוחד אם כבר הודגם בעבר שהקשר קיים עבור משתנים אופרציונליים אחרים של אותו משתנה תיאורטי. ישנם כמה סוגים של תיקוף עם קריטריון-

1. תוקף ניבוי - **(predictive)** האם המדידה מנבאת משתנה אחר? היכולת של המדידה לנבא משתנים אחרים שימדדו בעתיד. נרצה לבדוק האם מדידה אחת בזמן א' מנבאת מדידה אחרת בזמן ב'.
 2. תוקף בו זמני - **(concurrent)** מדידת הקריטריון מתרחשת באותו זמן של מדידת המשתנה האופרציונלי שמעניין אותנו. דומה לתוקף ניבוי רק שהמדידה של הקריטריון מתרחשת באותו זמן בדיוק כמו מדידת המשתנה האופרציונלי. למעשה הרבה משתמשים במושג תוקף ניבוי גם במקרה של תוקף בו זמני ולכן ההבחנה ביניהם היא לא מאוד חשובה אבל כדאי להכיר אותה.
 3. הבדלים בין קבוצות - לפעמים תיאוריה משערת שיש הבדל בין אנשים מסויימים בערך המשתנה התיאורטי, תיקוף עם הבדלים בין קבוצות הוא בחינת קשר בין שיוך קבוצתי לבין המשתנה האופרציונלי.
 4. לדוגמה - האם אנשים עם ADHD מראים ציונים נמוכים יותר בשאלון של תפקודים ניהוליים.
 4. הבדלים בין מאורעות - בחינת הקשר בין מאורעות לבין המשתנה האופרציונלי.
 4. לדוגמה - האם ציונים בשאלון חרדה מצבית גבוהים יותר במצבים מלחיצים יותר.
- תוקף ניבוי נמדד בעזרת מתאם בעוד תוקף של הבדלים בין קבוצות/מאורעות נמדד בעזרת מבחן לבדיקת הבדלים.

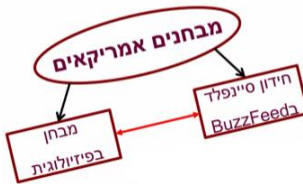
תיקוף באמצעות מדדים אחרים

אפשר לתקף מדידה גם באמצעות השוואתה עם מדידות אחרות של אותו משתנה תיאורטי, ישנם כמה דרכים שאפשר לעשות באמצעות תיקוף עם מדדים אחרים:



1. תוקף מתכנס - השוואה בין שני משתנים אופרציונליים שמוזדים את אותו המשתנה התיאורטי. נצפה למתאם גבוה בין שני המשתנים. נרצה מתאם גבוה ככל האפשר כדי להוכיח שהן מודדות אותו דבר, ולתקף את שתיהן. אז מתאם גבוה בין ציוני הנבדקים באופרציות השונות יעיד על תוקף מתכנס גבוה.

2. תוקף מבחין - כדי להראות שאנחנו מוזדים את המשתנה התיאורטי שרצינו ולא משתנים תיאורטיים אחרים (בלבדיות) חשוב שיהיה מתאם נמוך יחסית בין המדידה שלנו למדידה בשיטות שונות של משתנה אחר, וזה יהיה תוקף מבחין שבודק באיזו מידה האופרציה בודקת רק את מרכיבי תכונה X ולא את תכונה Y.



בדוגמה - נבדוק את המתאם בין מבחן בפיזיולוגיה שבודק ידע בתחום, לבין חידון על סיינפלד. שניהם מבחנים אמריקאים. מתאם נמוך בין ציוני אופרציות מדידה דומות של משתנים תיאורטיים שונים יעיד על תוקף מבחין גבוה. הוא יעיד שהמבחן בפיזיולוגיה לא בוחן יכולת להיבחן במבחנים אמריקאים, כי גם חידון סיינפלד הוא מבחן אמריקאי.

המתאם המבחין לבדו לא מספיק כדי לחזק את תוקף המבנה, זה רק מתאם נמוך בין שני מדדים. מתאם נמוך יכול להיות גם תוצר של מהימנות נמוכה באחד או שניים מהמדדים.

הגיויני להשוות לפעמים את התוקף המתכנס עם התוקף המבחין, אנחנו רוצים מתאם גבוה יחסית בין שתי שיטות מדידה לאותה תכונה, ומתאם נמוך בהרבה בין שתי שיטות מדידה דומות לשתי תכונות שונות.

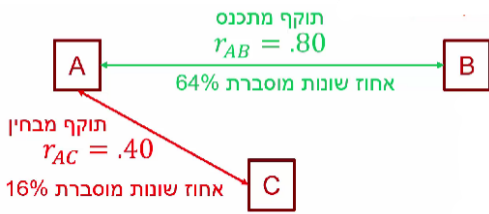
לדוגמה - נרצה לבחון יצירתיות, נבחן

- A - מבחן אמריקאי ליצירתיות
- B - מבחן השלכתי ליצירתיות
- C - מבחן אמריקאי לזיכרון

מתכנס, ונבחן עם מבחן אמריקאי על מדד אחר בשביל התוקף המבחין.

ריבוע כל אחד מהמתאמים זה אחוז השונות המוסברת -

בתוקף המתכנס בין מדידה B ו A רואים כי אחוז השונות המוסברת משמעותה הגורמים המשותפים שהשיעור על 2 המדדים A B מסבירים 64% מהשונות ביניהם. בתוקף המבחין בין C ו A רואים כי אחוז השונות המוסברת משמעותה שהגורמים המשותפים שהשיעור על 2 המדדים מסבירים 16% מהשונות ביניהם (העובדה שמדובר במבחן אמריקאי היא בעיקר הגורם המשותף).



זה יוביל אותנו לתהות אם בהינתן מקדם מהימנות מסוים, מהו התוקף המתכנס המקסימלי?

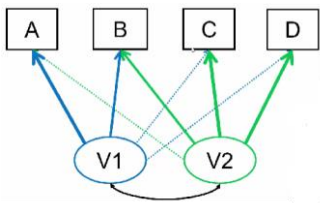
אנחנו צריכים לדעת מה מהימנות המדידה כי מקדם המהימנות שווה לחלק של שונות הציון האמיתי מתוך כלל שונות המדידה. המתאם המקסימלי של המדידה עם איזושהי מדידה אחרת הוא מתאם בין הציון האמיתי לציון הנצפה. לכן התוקף המתכנס המקסימלי הוא שורש המהימנות. כך אנחנו מזכירים שוב כי שורש המהימנות הוא חסם עליון לתוקף.



שורש המהימנות הוא חסם עליון לתוקף $r_{tx} = \sqrt{r_{xx}}$ לכן, $|r_{AB}| \leq \sqrt{.81} = .90$

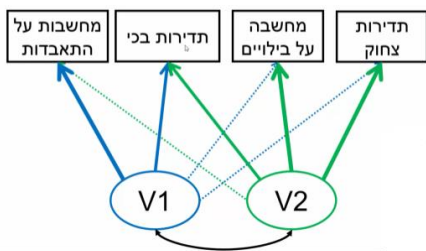
3.

ניתוח גורמים - **factor analysis** בניתוח גורמים אנחנו מודדים את הנבדקים בכמה מדדים (ריבועים) ובעזרת שיטת ניתוח סטטיסטית מחלצים מהשונויות המשותפות בין המדדים כמה גורמים (עיגולים) משותפים. כלומר מסתכלים על כל השונות שנוצרה מארבעת המדדים האלה ומחפשים דפוסים של שונויות משותפות, מחפשים חלקים מהשונות במדדים השונים שהולכים ביחד. הקווים בין הריבועים לעיגולים מחולקים לקווים חזקים-שמשקפים קשר חזק, וקווים חלשים- שמשקפים קשר חלש.



הרעיון הוא שאפשר להסיק מהמתאמים בין המשתנים הנמדדים שיש גורמים משותפים ביניהם. הגורמים המשותפים הם למעשה משתנה תיאורטי כי הם לא נמדדים באופן ישיר, אבל הם הסיבה לשונות המשותפת בין המדדים. זה קצת דומה לתוקף מתכנס - המתאם בין שני מדדים שונים לאותה תכונה נגרם מכך שיש מקור משותף לשונות שהוא המשתנה התיאורטי.

לדוגמא - נמדוד 4 גורמים שונים, מוצאים כי יש שונות משותפת בין תדירות צחוק, מחשבה על בילויים ותדירות בכי. אבל הם לא חולקים אותה עם מחשבות על התאבדות. ניתן לפרש את 2V1 1V2 - נראה כי 1V קשור לדיכאונות. 2V קצת יותר מורכב, לא לגמרי ברור מהו הגורם המקשר ביניהם.



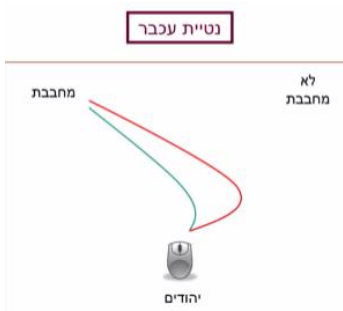
שיטת תיקוף אחרונה - חקר המדידה כתופעה פסיכולוגית:

חקר התהליך המשפיע על המדידה

תיקוף באמצעות חקר התהליכים הפסיכולוגיים שמשפיעים על המדידה.

ניתן לבצע ניסויים כדי להבין מה בדיוק משפיע על התנהגות האדם בעת המדידה. כך אנחנו יכולים לבודד גורמים שונים שמשפיעים על המדידה, זה עוזר לנו להבין מהו בעצם המשתנה התיאורטי שאנחנו מודדים עם השיטה. זה גם עוזר לנו למצוא גורמים המשפיעים על המדידה שלא משקפים את המשתנה התיאורטי.

לדוגמא - מדידה עמדה אוטומטית באמצעות נטיית עכבר. ההנחה היא שהעמדה הנשלטת תשפיע על מה תבחר אבל העמדה האוטומטית תשפיע על מסלול העכבר. אנחנו יכולים לבצע מחקר שינסה להבין מה הם התהליכים המשפיעים על תנועת העכבר בניסוי הזה. למשל- יכול להיות שיתגלה שאנשים שלא משתמשים בעכבר לעיתים תכופות יראו מסלול פחות ישיר אל הבחירה שלהם. זאת תהיה שונות שלא נובעת מטעות מקרית, מאוד יעזור לנו לגלות אותה כי אולי נצליח לשלוט בה באמצעות מבחן נוסף שיבדוק את יציבות השליטה בעכבר של כל נבדק בלי קשר לעמדה האוטומטית שלו. מבחן כזה יעזור לנו לבודד את השונות שנובעת מהניסיון עם העכבר ולנקות אותה



מציון העמדה שניתן לכל נבדק. אז בעצם השונות שאנחנו מודדים (של X) אנחנו יכולים לחשוב עליה כמורכבת מהשונות של העמדה + שונות של משתנה השליטה בעכבר + שונות של טעות. אם ננכה את שונות השליטה בעכבר יש לנו מדד יותר טוב לשונות העמדה.

זה תיקוף כי זה עוזר לנו להבין מה המדידה מודדת ולשפר את מדידת המשתנה התיאורטי שמעניין אותנו, במקרה הזה- העמדה האוטומטית.

שיטות לתיקוף – תוקף המשתנה הבלתי תלוי

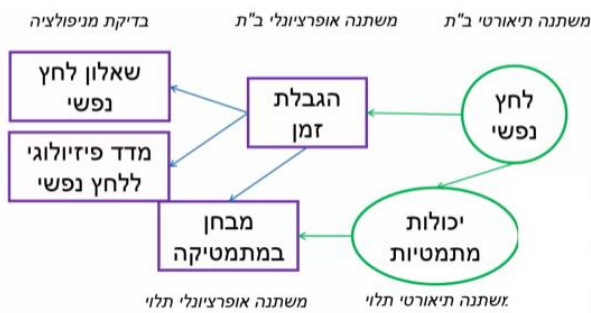
עד כה ענינו על השאלה איך אפשר לדעת שהמשתנה הנמדד מייצג את המשתנה התיאורטי. ננסה לענות על השאלה איך ניתן לדעת שהמשתנה המתופעל מייצג את המשתנה התיאורטי.

$$s_x^2 = s_c^2 + s_m^2 + s_e^2$$

↖ ↖ ↖
עמדה שליטה בעכבר טעות

תוקף המניפולציה

אנחנו רוצים לשאול האם המניפולציה שהפעלנו אכן הפעילה את המשתנה שאותו התכוונו להפעיל.



לדוגמא - אנחנו רוצים לבדוק את ההשפעה של לחץ נפשי (תיאורטי בלתי תלוי) על יכולות מתמטיות (תיאורטי תלוי).

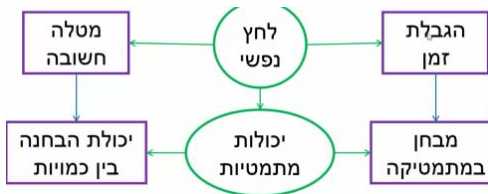
דרך אחת שאנחנו יכולים לעשות היא באמצעות בדיקת מניפולציה - אנחנו מוסיפים מדד שמשמש כמשתנה אופרציונלי בלתי תלוי נמדד - אנחנו בודקים איך המשתנה האופרציונלי הבלתי תלוי נמדד.

נגיד שנתפעל לחץ באמצעות הגבלת זמן (משתנה אופרציונלי בלתי תלוי) ויכולות מתמטיות באמצעות פתרון מבחן במתמטיקה (משתנה אופרציונלי תלוי). איך נדע שהמניפולציה של הגבלת הזמן באמת תפעלה את הלחץ הנפשי? או בניסוח אחר- אם לא נמצא השפעה איך אפשר לדעת שזה קרה בגלל שלא הייתה השפעה או בגלל שהמניפולציה שלנו לא עבדה?

הפתרון הוא להוסיף מדד שבדוק אם אכן המשתנה שתפעלנו תופעל. למשל ע"י הוספה של שאלון לחץ, או מדידת מדד פיזיולוגי ללחץ נפשי וכו'. זוהי בעצם בדיקת מניפולציה- הוספנו מדד שבדוק אם הרמות של המשתנה הבלתי תלוי שונות בין התנאים השונים. אפשר לבצע גם בדיקת מניפולציה במחקר נפרד במקרה שהמדידה למשל עשויה לפגוע בניסוי במהלך הניסוי, ואז מחקר נפרד יכול להוכיח את תקפות המניפולציה, מחקר כזה נקרא מחקר תיקוף.

שחזור קונספטואלי

בודק האם הממצאים משתחזרים גם עם משתנים אופרציונליים חלופיים. אם נמצא תוצאות דומות אז גדל הסיכוי שאכן הקשר שמצאנו בשני הניסויים מייצג קשר בין המשתנים התיאורטיים שרצינו לבחון. אם לא נמצא תוצאות דומות, אז יכול להיות שנוכל להבין טוב יותר את המשתנים האופרציונליים, לעדכן את ההגדרות של המשתנים התיאורטיים וכתוצאה מכך לעדכן את התיאוריה.



בדוגמא- השפעה של לחץ על יכולות מתמטיות- הניסוי הראשון בדק זאת באמצעות הגבלת זמן על מבחן במתמטיקה, בניסוי של שחזור קונספטואלי נוכל להגיד לנבדקים שלמבחן שהם הולכים לבצע יש השלכות חשובות (לתפעל לחץ נפשי) ולבדוק את יכולת ההבחנה שלהם בין כמויות (דרך אחרת לבדוק את המשתנה התיאורטי של יכולות מתמטיות).

שחזור קונספטואלי באופן הזה יחזק את תוקף המבנה, אבל עדיין לא יהפוך אותו לבלתי שביר - יכול להיות שיהיה אפשר להסביר את התוצאות של שני המחקרים האלה באמצעות הסבר חלופי של משתנה תיאורטי נוסף שמשפיע. לדוגמא - הגבלת זמן ומטלה חשובה דורשים מאמץ. יכולות מתמטיות דורשים כוחות שאזלו במאמץ. אם מאמץ הוא שהוביל להפחתה בביצועי הנבדקים, ייתכן והמאמץ הוא מה שתופעל במניפולציה, וזו פגיעה בבלבדיות. כלומר, שחזור קונספטואלי לא מגן באופן מלא בפני איומים על התוקף מבנה של המשתנה הבלתי תלוי.

מערכים דמויי ניסוי

כשלא ניתן להשתמש במערך ניסויי לצורך ביצוע המחקר, נשתמש במערכים דמויי ניסוי.

במערך ניסויי יש שליטה מלאה על המשתנים, הסביבה הניסויית ותנאי העברה, יש לו הרבה יתרונות במובן של הסקה סיבתית אבל בד"כ יש לו תוקף חיצוני מאוד חלש וזה חיסרון מאוד גדול - אי אפשר להכליל בוודאות ממחקרי מעבדה על השטח.

היתרון של מערך מחקר דמויי ניסויי - התוקף החיצוני שלו גבוה, כשנרצה להסיק על תנאים מציאותיים. עם זאת, מערכים דמויי ניסוי הרבה פעמים מגיעים עם בעיות תוקף מסוימות שמקשות על הסקה סיבתית, אך הם מאפשרים הסקה טובה יותר ממערכים ניסויים. החיסרון העיקרי של מערכים דמויי ניסוי הוא שהם לא שוללים הרבה ארטיפקטים.

דוגמה נפוצה (אבל לא היחידה) למערכים דמויי ניסוי היא מחקרי שטח - ממש עושים בפועל, בשטח את מה שאנחנו רוצים לבדוק. אפשר להגיד שלחוקרים במערכי דמויי ניסוי יש יכולת מוגבלת לשליטה בתנאי הניסוי, מה שמקטין את היכולת להסיק מסקנות על סיבתיות. יכולות להיות כל מיני בעיות, הנפוצות הן- בעיות בהקצאה אקראית לקבוצות, או שליטה בתפעול המשתנה הבלתי תלוי.

נקודה שחשוב לזכור - לא כל מחקרי השטח הם דמויי ניסוי. אפשר לעשות מחקרי שטח ניסויים, אם שולטים על ההקצאה המקרית מספיק טוב.

דוגמה - נבחנת שיטת לימוד חדשה בתנאי מעבדה מבוקרים, והסתבר שיש לה יתרון על שיטת הלימוד השכיחה היום. האם השיטה הזאת תהיה יעילה גם במערכת החינוך? לא בהכרח, תנאי המעבדה מלאכותיים ולכן כדאי לבחון את יעילות השיטה גם מחוץ למעבדה.

סיבות עיקריות לביצוע מחקרי שדה (דמויי ניסוי -)

1. כדי לבחון את התוקף החיצוני של תוצאות אשר הושגו במחקר מעבדתי: לשפר הכללה מעבר למצבים/נבדקים/זמן (כמו בדוגמא)
2. כאשר לא ניתן "להכניס את הניסוי למעבדה" או לתפעל את המשתנים. (למשל, השפעה של רפורמת ענישה וקנסות על מספר התאונות הקשות שיתרחשו)
3. על מנת להעריך/לקבוע השפעה/ תוצאה של אירועים המתרחשים בשטח, שעליהם אין לנו שליטה (למשל, לבדוק השפעה של שינוי בתנאי העסקה של עובדים על היעילות שלהם).

מוטיבציות עיקריות לביצוע מחקרי שדה

- מערכים דמויי ניסוי נועדו לאפשר לחוקר להסיק על סיבתיות למרות שאינם מתקיימים במעבדה
 - מנסים להתמודד עם העדרם של הקצאה מקרית וקבוצת ביקורת שקיימים במערך ניסויי
- למערכים דמויי ניסוי יש מבנה בסיסי כמו מערך ניסויי - השערה, מניפולציה בין שני תנאים או יותר. היתרון של מערכים דמויי ניסוי הוא שאנחנו מצליחים להתמודד עם חלק מהאיומים (confound), אולם כמובן שאינם מאפשרים להסיק על סיבתיות כמו מערך ניסויי תקין.

דוגמאות למערכים דמויי ניסוי

תזכורת למושגים וסמלים לייצוג המחקר בצורה גרפית:

- X = טיפול ניסויי (מניפולציה) \boxtimes היעדר X היעדר מניפולציה
- O = תהליך של תצפית או מדידה. זו הנקודה בזמן בה אנחנו מודדים את המשתנים התלויים שמעניינים אותנו
- כל שורה מייצגת רצף שקרה לאותם נבדקים/ אותה קבוצה
- R = הקצאה מקרית לקבוצה
- כל עמודה מייצגת פעולות שקרו באותו הזמן (ההצגה היא משמאל לימין)

יש רצף של מדידות, כל הנתונים נמדדים כמה פעמים לפני אירוע כלשהו ואחרי אותו אירוע, או מניפולציה. זה דומה למערך של לפני אחרי ללא קבוצת ביקורת, אבל טוב יותר כי יש יותר מדידות לפני ויותר מדידות אחרי. מספר המדידות יכול להשתנות, בכל מקרה צריכות להיות לפחות 2 מדידות לפני ולפחות 1 אחרי.

O_1 O_2 O_3 X O_4 O_5

היתרון במדידה חוזרת הוא שזה מגביר את הסיכוי ששינוי בעקבות המניפולציה אכן קרה בגלל המניפולציה.

לדוגמה - נבדוק האם חוק חדש שחייב הורים להושיב ילדים מתחת לגיל 4 בכיסאות בטיחות הצליח להביא להפחתה של מספר אירועי הפציעה של ילדים מתחת לגיל 4 בתאונות דרכים. דוגמה אפשרית לתוצאות המחקר - החלת הרפורמה הובילה לירידה משמעותית במספר הפגיעות של ילדים מתחת לגיל 4 בתאונות דרכים.

על פניו, נראה כי הרפורמה היא שגרמה לירידה במספר הנפגעים. אך ייתכן הסבר אחר לתוצאות - כלומר איום על התוקף הפנימי של המחקר (אין הקצאה מקרית ואין קבוצת ביקורת).

האיומים שמערך מחקר זה מתגבר עליהם:

- מדידה ורגרסיה לממוצע - יש מספר מדידות עד לקבלת דפוס יציב, ולכן לא סביר כי איומים אלו יוכלו להסביר את השינוי בין המדידות לפני ואחרי.
- בשילה - אם היא קיימת ניתן לעיתים קרובות לזהות אותו במדידות המרוכות, מאחר ונראה שינוי קודם למניפולציה - לאורך זמן (למשל מהדוגמה, במחקר על תאונות הדרכים היה אפשר לראות ירידה בכמות תאונות הדרכים בגלל שמשהו קרה באוכלוסייה ולא בגלל המניפולציה. במערך כזה נדע אם היה איום בשילה).

האיומים שמערך זה לא מתגבר עליהם:

- היסטוריה ומכשור - אם יהיו הבדלים במכשור או שיתרחש משהו בין המדידות זה עשוי להסביר את ההבדל בין המדידה לפני ואחרי. (למשל, שינויים במחירי הדלק שחלו במקביל להחלת הרפורמה וגרמו לאנשים לנסוע לאט יותר ולכן יש פחות נפגעים בכבישים).
- ברירה - אינה רלוונטית משום שלא מדובר בשתי קבוצות.

איך נתגבר על האיומים?

אפשר להוסיף קבוצת ביקורת אחת או יותר.

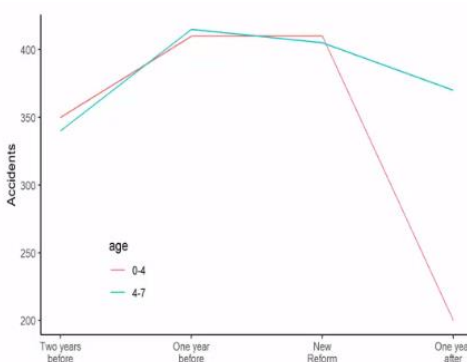
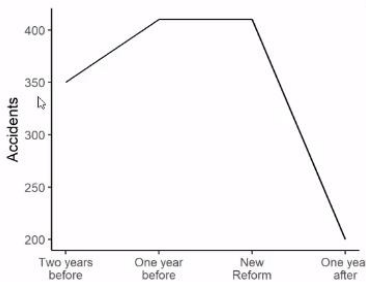
לדוגמה - נבדקו שתי קבוצות ביקורת :

1. נבדקו רישומים ביחד לפגיעות של ילדים בין הגילאים 4-7 אשר לא חוקק חוק עבורם. נמצא כי בעוד שעבור ילדים צעירים מגיל 4 חלה ירידה במספר הפגיעות, לא היה שינוי עבור קבוצת הגיל הבוגרת יותר.
 2. נבדקו רישומים של מדינה שכנה אשר בה לא הוחל החוק, לא נצפתה במדינת הביקורת ירידה בפגיעות ילדים מתחת לגיל 4.
- הוספת 2 קבוצות הביקורת מחזקת את הביטחון בהשפעה של החוק החדש, ושלא הייתה השפעה של מכשור או היסטוריה.

קבוצות הביקורת עדיין לא הפכו את המחקר לניסוי - התשובה היא לא, אין הקצאה מקרית.

איום אחד שעדיין יכול לקרות בגלל שזה לא מחקר ניסויי הוא איום של אינטראקציה בין ברירה למניפולציה. כלומר, משהו ספציפי בקבוצת הגיל הזאת היה שונה ביחס לקבוצות גיל אחרות וזה גרם לאפקט להיות שונה בקבוצת הגיל הזאת מאשר בקבוצות אחרות. במחקר ניסויי החוקרים שולטים בערך המשתנה הבלתי תלוי של כל נבדק, ולכן מערך מחקר דמויי ניסויי עדיין לא שולט על כל הבעיות אבל הוא מצמצם חלק גדול מהבעיות שדיברנו עליהם במערכי מחקר לא טובים.

מערך עם קבוצת ביקורת לא שווה



מעריך שבו יש מניפולציה ושתי קבוצות, אבל אין הקצאה מקרית לקבוצות השונות.

X O₁
O₂

ההבדל הבולט בין מעריך זה למעריך של סדרת זמן מתמשכת הוא שבמעריך זה אין מדידות לפני האירוע/מניפולציה ובסדרת זמן מתמשכת יש.

לדוגמא -מחקר שמבקש לבחון את ההשפעה של עישון על ריכוז. לא ניתן להקצות נבדקים באופן מקרי לקבוצת מעשנים ולא מעשנים בשל מגבלות אתיות. לכן, ההקצאה של הנבדקים לקבוצות איננה מקרית.

שתי בעיות מרכזיות במעריך זה:

1. ייתכן שהקבוצות שונות במשתנה התלוי עוד לפני ביצוע המחקר. זה מתקיים כשאין לנו מדידה לפני הפתרון - לעשות מדידה מקדימה, למדוד את המשתנה התלוי לפני ואחרי המניפולציה (ואז זה בעצם מעריך לפני-אחרי ללא הקצאה מקרית, ויש לו את אותם האיומים כמו של מעריך זה).

O₁ X O₃
O₂ O₄

2. ייתכנו משתנים נוספים שהקבוצות שונות בהם, שיהוו הסבר אלטרנטיבי לתוצאות (איומי ברירה, נשירה אינטראקציה של ברירה עם הדברים האלה). מספר פתרונות-

- להגדיר מראש אילו משתנים יכולים להיות מעורבים ולשלט עליהם עד כמה שניתן, לחשוב מראש בצורה שיטתית איך ניתן לגרום לקבוצות להיות דומות כמה שניתן. ניתן לעשות זאת על ידי התאמה = הפעולה בה דואגים להתאמה של המדידות/נבדקים בין הקבוצות, כך שיהיה ייצוג שווה לרמה מסוימת בכל הקבוצות.
- הגדרת משתנה תלוי נוסף שלא אמור להשתנות מהמניפולציה, **nonequivalent outcome**. המשתנה הזה אמור להישאר דומה בין שתי הקבוצות אחרי המניפולציה ובכך נדגים שהאפקט שלנו הוא יחסית ספציפי למניפולציה. כתוצאה מכך מתחזק התוקף המבחני.

רגרסיה חסרת המשכיות

קבוצה אחת עוברת מניפולציה והשנייה לא עוברת אותה. הקבוצות לא מוקצות אקראית, אלא לפי קריטריון (יש איזושהי שליטה על ההקצאה) - מי שעומד בקריטריון(+C) מוקצה לקבוצת הניסוי, ומי שלא עומד בקריטריון (-C) מוקצה לקבוצת הביקורת.

[C+] O₁ X O₃
[C-] O₂ O₄

נשתמש במעריך זה כאשר אין לנו הרבה משאבים ואנחנו לא יכולים לספק את המשאבים לכל הנבדקים, אבל אנחנו עדיין רוצים להעריך את האפקט ואת הסיבתיות.

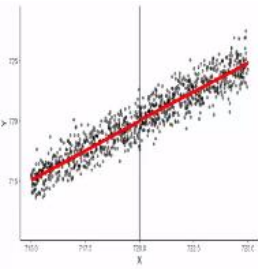
דוגמא -מחקר הבוחן את השפעה מלגה על הצטיינות בלימודים. המלגות ניתנות לתלמידים שקיבלו ציון גבוה במבחן כניסה. כעבור שנה נמצא כי אכן התלמידים שיפרו את ציוניהם.

עדיין אי אפשר לטעון כי בהכרח למלגות ישנה השפעה על הציונים - יתכן שהתלמידים האלו היו מוצלחים יותר גם ללא קבלת המלגה, כי הם מוכשרים. אם היו יורדים הציונים, עדיין אי אפשר לטעון כי המלגה מחבלת בהצלחה - יכול להיות שמשנתה אחר גרם לירידה בציונים.

הדרך הטובה ביותר לבדיקת ההשפעה של המלגה - הקצאה מקרית של תלמידים לקבוצות של מקבלי מלגה וללא מלגה. הבעיה היא שאי אפשר לחלק מלגת הצטיינות באופן מקרי מאחר והיא מיועדת כגמול להצטיינות. הפתרון (החלקי) הוא מעריך רגרסיה חסרת המשכיות.

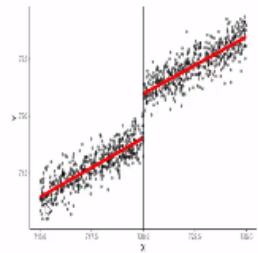
סדר הפעולות במעריך זה:

- נקבע קריטריון מסוים (ציון) שממנו ואילך קיימת זכאות לקבלת מלגה
- ניקח קבוצה מצומצמת מעל ומתחת לקריטריון ומתוכה דוגמים אקראית מספר שווה של סטודנטים מעל הקריטריון (יקיבלו מלגה) ומתחת לקריטריון (לא יקיבלו מלגה). הקבוצה מצומצמת מבחינת טווח הציונים, כדי שיהיה דמיון ככל האפשר בין מי שמקבל את ה"טיפול" לבין מי שלא.



תוצאות אפשריות:

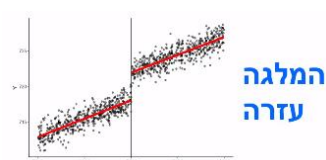
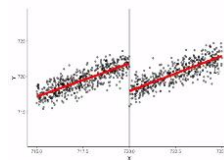
- אם המדידה לאחר קבלת המלגה לא תראה הבדל ניכר בין שתי הקבוצות, נוכל להסיק שהמלגה לא משפיעה על הציונים. אין הבדל בין הקווים לפני ואחרי הקריטריון/חתך (הקו הוא רציף, אם היתה השפעה למלגה היא היתה גורמת להסטה של הגרף).



- אם המדידה לאחר קבלת המלגה תראה יתרון למקבלי המלגה נוכל לחשוף שיש אפקט - שיש השפעה למלגה. קטיעה ברורה ברצף הרגרסיה, שלא ניתן להסביר על ידי הרגרסיה הנורמלית של מדידה שאחרי על זו שלפני.

כיוון ההבדל חשוב:

- המלגה עזרה - הסיטה לכולם את הציונים כלפי מעלה.
- המלגה הזיקה - שתי הקבוצות נראות אותו דבר. הקו היה אמור להיות רציף, בעצם המלגה הסיטה לכולם את הציונים כלפי מטה.



נבחן האם גם בלי מלגה (בהנחה שפסיכומטרי מנבא הצלחה), יכול להיות שקבוצת הסטודנטים שקיבלה את המלגה היו אמורים להיות מוצלחים יותר מלכתחילה ולכן אין השפעה בכלל למלגה.

יש שתי דרכים לבדוק האם קיים הבדל מובהק:

1. לבדוק האם הקשר שנמצא בקבוצת הביקורת בין הפסיכומטרי לציונים, מנבא היטב גם את הציונים של קבוצת הניסוי. אנחנו רוצים לבדוק האם חישוב הציון של קבוצת הניסוי לפי הקשר הסטטיסטי שנמצא אצל קבוצת הביקורת יביא להערכת חסר של ציוני קבוצת הניסוי. נבנה משוואת רגרסיה לכל קבוצה בנפרד ולהשוות ביניהם, זה נקרא ניתוח מיתון.
2. דרך נוספת היא לבדוק האם מלגה מנבאת ציונים מעבר לפסיכומטרי. נעשה זאת באמצעות בניית משוואת רגרסיה עם שני מנבאים: ציון הפסיכומטרי וקבלת מלגה. נרצה לבדוק האם מעבר לציון הפסיכומטרי, כאשר הוא קיים במודל, אפשר לנבא ציונים בעזרת מלגה, או האם מקדם המנבא של מלגה הוא מובהק. אם מקדם הרגרסיה של מלגה מובהק אז אפשר להסיק שיש הבדל בין הקבוצות שלא מוסבר על ידי ציון הפסיכומטרי אבל כן מוסבר על ידי המלגה.

איומים עיקריים הנובעים מחוסר ההקצאה האקראית:

1. משתנה אחר השפיע על קבוצת הניסוי. לדוגמה - אולי סטודנטים עם פסיכומטרי מעל 720 מוחרמים על ידי שאר הסטודנטים (הדוגמה קצת מגוחכת, זה סביר יותר בקריטריון פחות שרירותי מ-720 פסיכומטרי).
2. אולי הפונקציה המקשרת בין הקריטריון לציון אינה מתאימה לרגרסיה לינארית, אולי הניבוי הוא לא לינארי. לדוגמה - אולי מעל הקריטריון אפילו בלי המלגה הגרף יראה כמו רגרסיה קטועה. כלומר, הערך הטבעי באוכלוסייה גם יציג קטיעה. בעיקרון הקצאה מקרית הייתה פותרת את הבעיה, אבל מכיוון שאנחנו לא יכולים לעשות את זה אפשר להשתמש בקבוצת ביקורת כדי להפחית את האיום הזה. למשל, לתת את המלגה לסירוגין שנה כן שנה לא ולראות אם בשנים שבהם לא נתנו מלגה אנחנו מוצאים את אותו קו רגרסיה קטוע, אם נמצא שכן גם בקבוצת הביקורת, נוכל להניח שהמניפולציה של מתן

מלגה לא השפיעה.

לסיכום - מה יקבע מהו מערך דמוי ניסוי טוב?

- תשומת לב גדולה מאוד למשתנים מתערבים אפשריים.
 - תשומת לב למאפייני הדגימה, צריכים להיות דומים ככל האפשר בין הקבוצות.
- אם נשמור על העקרונות הללו בצורה שיטתית וננסה לנטרל איומים עד כמה שאפשר, אפשר בזהירות להסיק סיבתיות גם במערכים דמויי ניסוי, ככל שיש לנו כמה מערכים דמויי ניסוי ששולטים על בעיות שונות- זה יותר טוב.

מערכים מתאמיים

מערכים מתאמיים ללא יכולת הסקה סיבתית - לפעמים אנחנו רוצים לחקור תופעות שמתרחשות בעולם באופן טבעי ואין לנו אפשרות לדמות מצב של מניפולציה, במקרים כאלה נפנה למערכים מתאמיים. כאמור, קשה להסיק סיבתיות ולשלול הסברים אלטרנטיביים במערכים מתאמיים. ובכל זאת יש מצבים בהם לא תהיה לנו ברירה אלא להשתמש במערכים מתאמיים. נרצה לנטרל את ה confounds כמה שאפשר.

סוגים של מערכים מתאמיים-

- מחקר אורך (longitudinal) (בו בודקים את הנבדקים בכמה נקודות לאורך הזמן. דומה לסדרת זמן מתמשכת עם היתרונות והחסרונות שלה.

O_1 O_2 O_3 O_4 O_5

- מחקר חתך (cross sectional) (בו נדגום את כל הנבדקים באותה נקודת זמן ובכך ננטרל את איום ההיסטוריה והבשילה. האיום המרכזי על מערך כזה הוא ברירה - יש כמה קבוצות ואין הקצאה מקרית.

O_1
 O_2
 O_3
 O_4
 O_5

מערך ניסויי לנבדק בודד

Single-Subject experimental designs ניסויי בו נבדק בודד עובר סדרת מדידות חוזרות ומניפולציות שונות.

כמו מערך ניסויי - נערכת מניפולציה במשתנה הבלתי תלוי ונמדדת ההשפעה שלו על המשתנה התלוי .
במובנים מסוימים, זו הרחבה של מערכים תוך נבדקיים עם מדידות חוזרות כי הנבדק מופיע בכל אחד מהתנאים בניסוי.
במובנים אחרים, זו הרחבה של מערך מסוג סדרת זמן מתמשכת - הנבדק נמדד כמה פעמים לאורך פרק זמן.
מערכים מהסוג הזה התפתחו מאוד בעשרים שנה האחרונות והם לעיתים משמשים כאלטרנטיבות למערכי מחקר קבוצתיים שכיחים עליהם דיברנו בעבר.

רוב הניסויים הם השוואות על קבוצות. מתי נעדיף מערך ניסויי בודד על פני מערך של השוואה בין קבוצות:

1. כאשר נרצה להעריך שינוי אצל נבדק בודד - למשל אם נרצה לדעת מה עוזר לנו ספציפית להירגע אז ננסה על עצמנו כל מיני שיטות. במקרה הזה, ההכללה פחות מעניינת אותנו.
2. כאשר הפרוצדורה (מניפולציה ו/או מדידה) מורכבת יחסית וקשה להעביר אותה על הרבה נבדקים. למשל, אם קל לי יותר לערוך ניסוי על הכלב שלי מאשר על כלבי השכונה.

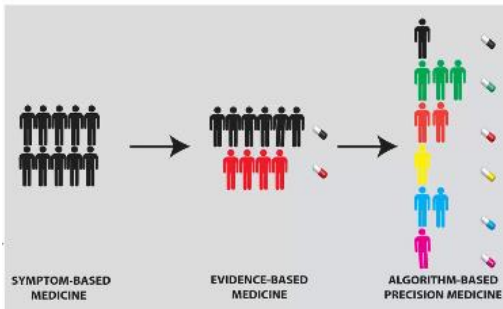
Single Subject: הבעיה העיקרית במערך ניסוי לנבדק בודד - עשוי להיות בעלי תוקף חיצוני חלש, אנחנו לא יודעים אם זה יכול להיות מוכלל לעוד אנשים. אבל אם המחקר נערך טוב, ייתכן שיהיה לו תוקף פנימי גבוה ונוכל להסיק סיבתיות היטב לגבי הנבדק הספציפי.

Group Comparison: מחקר המשווה בין קבוצות - מספק מידע על ביצוע קבוצתי ולכן עשוי לאבד אינפורמציה חשובה הנוגעת

לביצוע של אדם ספציפי. כלומר - אנשים מגיבים באופן שונה לאותו טיפול, גם אם הטיפול נראה אפקטיבי בממוצע, הוא אינו בהכרח אפקטיבי עבור כל המשתתפים.

יש פה אלמנט הסתברותי וגם אלמנט של התאמה אישית – בני אדם הם מערכות ביולוגיות מורכבות.

הדגש על הנבדק הבודד בא לידי ביטוי גם ב - **precision medicine** התגובה של אדם לתרופה כלשהי תלויה בשילוב של הרבה



גורמים ספציפיים למערכת הביולוגית שלו, בגלל שהמערכת הביולוגית שתומכת בגוף האדם היא מורכבת, יכול להיות מספר ענק של שילובים של הרבה תגובות ייחודיות מעבר לנבדקים. בעצם, בעולם כזה אנחנו בעולם מערכי המחקר לנבדק בודד. בסכימה אפשר לראות איך רפואה מבוססת סימפטומים/ראיות עושה השוואות ברמת קבוצה. היום יש מעבר הולך וגובר להתאמה אישית של טיפול לנבדקים אינדיבידואליים, כך שלכל אחד יכולה להיות התרופה המתאימה לו. המעבר הזה הוא מסובך מאוד ברמה חישובית, אין לנו הרבה פעמים את כל המידע, ולכן מתחילים להיות כלים אלגוריתמים שמאפשרים בהקשר הרפואי להתחיל לחזות תגובות אינדיבידואליות לתרופות גם בלי לערוך מערך מחקר נפרד לכל נבדק.

בפסיכולוגיה זה לא קיים וכנראה לא יהיה בקרוב, כי אין מודלים אלגוריתמיים טובים לפסיכולוגיה האנושית.

לא מדובר במערכי **Single Case Studies** – בהם אין ניסיון לבקר את המשתנה הבלתי-תלוי. החוקר צופה ומסיק, אך אינו שולט במשתנה הבלתי תלוי ולכן לא ניתן לשלול הסברים אלטרנטיביים לתוצאות. לעומת זאת, במערכי נבדק בודד - החוקר שולט ומתפעל את המשתנה הבלתי תלוי, אשר השפעתו על המשתנה התלוי נמדדת.

מערכי case studies משמשים היום בעיקר למחקרי עומק קליניים שמתארים מקרים יחידים של אינדיבידואליים, מהם יוצרים תיאוריות חדשות אבל לא בוחנים אותן. שני המקרים מכילים נבדק אחד בלבד אבל ההבדל הקריטי הוא השליטה במשתנה הבלתי תלוי.

הכוח של מערכי נבדק בודד -

- שליטה במשתנים הבלתי תלויים
- הפחתת הפוטנציאל להימצאותו של גורם מסביר נוסף

כך אנחנו מגדילים את התוקף הפנימי ואת היכולת להסיק על סיבתיות.

הבעיה בסיבתיות במערכים האלו -

- אין קבוצת ביקורת, משווים בין התנהגותו הראשונית של הנבדק לפני מתן הטיפול לבין התנהגותו לאחר מתן הטיפול (כמו מערך לפני אחרי נבדק בודד).

לכן, נסיק סיבתיות תחת שני תנאים:

1. התנהגות הנבדק משתנה באופן קבוע בכיוון המצופה כאשר הטיפול נוכח בהשוואה למצב בו אין טיפול
2. ניתן לפסול קיומם של רוב המשתנים המתערבים

שני התנאים האלו יאפשרו לנו להסיק סיבתיות ממערכי מחקר של נבדק יחיד .

שלבים בביצוע מערך ניסוי לנבדק בודד - יעזור לנו להסיק על סיבתיות.

1. הגדרה אופרציונלית של התנהגות היעד – אנחנו רוצים להגדיר מהו המשתנה התלוי. אנחנו צריכים להגדיר מראש את המשתנה התלוי במחקר בצורה אופרציונלית. דוגמא - אם את רוצה להפסיק לעשן, אנחנו רוצים מדדים אופרציונליים שיאפשרו לנו למדוד את זה - מספר סיגריות שהנבדק מעשנת כל 24 שעות, הערכת מידת ההשתוקקות לסיגריה וכו'. נניח שמטרת המחקר היא לבדוק את השפעת מדבקת ניקוטין על מספר הסיגריות המעושנות ועל העוצמה בה הנבדק משתוקק לסיגריה, אנחנו ניתן מדבקה ונבדק את ההשפעה שלה על המשתנים התלויים. התנהגות היעד היא המשתנים התלויים במערך .
2. מדידת תקופת ה – **Baseline** אנחנו רוצים לבדוק את המשתנים התלויים בזמן הבייס לין =משך הזמן שבו אנחנו בודקים את ההתנהגות הראשונית, כל פרק הזמן מתחילת הניטור עד תחיל ביצוע המניפולציה. משך הזמן של הבייס לין משתנה

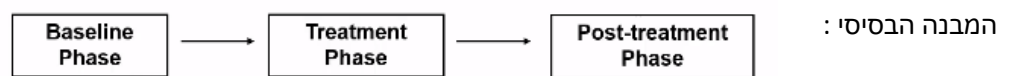
בהתאם להתנהגות הספציפית שאנחנו בודקים, בד"כ מדובר במספר ימים - שבוע. אנחנו רוצים לקחת בייס ליין מייצג ולהימנע מכזה שלא מייצג בו ההתנהגות משתנה בצורה קיצונית וחד פעמית .

בדוגמא - הבייס ליין יכול להיות תקופה של שבוע שבה הנבדק מציין כמה סיגריות הוא מעשן כל יממה, וכמה הוא משתוקק לעשן סיגריות בכמה נקודות זמן במשך היום. הבייס ליין משמש לאותה מטרה כמו קבוצת ביקורת במערך בין נבדקי.

3. יישום המניפולציה הטיפולית – ניישם את המניפולציה הטיפולית - נבצע את הניסוי. הטיפול הוא המשתנה הבלתי תלוי במערך והוא חייב להיות מוגדר מראש בצורה אופרציונלית.

בדוגמא - זה יהיה מדבקת ניקוטין שנועדה להפסקת עישון, אז היישום הוא מתן מדבקת ניקוטין ותחילת השימוש של הנבדק. אנחנו מוודים גם את ההתנהגות הזאת לאורך תקופה.

4. ניטור שינויים בהתנהגות – נמדוד את השינויים בהתנהגות במהלך הטיפול. נשווה את המידע שנאסף בתקופה שלפני הטיפול לזה שנאסף אחרי הטיפול. החוקר ירצה לבחון את הפרש במשתנה התלוי (התנהגות היעד) לבין שני ערכי המשתנה הבלתי תלוי (לפני ואחרי הטיפול). יש לנו 2 מדידות ממושכות בזמן - כמה פעמים בייס ליין וכמה פעמים לאחר המניפולציה.



סוגי מערכי ניסוי לנבדק בודד:

ABA reversal designs

הרעיון העיקרי הוא שאנחנו רוצים לבדוק את האפקט של משתנה בלתי תלוי על המשתנה התלוי בצורה מעגלית: אנחנו מוודים את המשתנה התלוי עם טיפול (מניפולציה) ובלי טיפול, וחוזר חלילה, בנקודות זמן שונות.

המערך המינימלי מסוג זה הוא מערך בו ישנו:

1. פרק זמן בו אין כל טיפול (בייס ליין) ובמהלכו התנהגות היעד מופיעה - נגדיר את הבייסליין בתור פרק זמן שקודם למניפולציה ובו אין שום מניפולציה, אבל יכולות להופיע התנהגויות שונות.

2. פרק זמן בו מופיע הטיפול (מניפולציה)

3. חזרה (Reversal) לפרק הזמן בו אין טיפול

לרוב אנחנו משתמשים במערך מהסוג הזה רק כשאפשר לצפות שהמשתנה התלוי יכול לחזור למצבו הקודם בכל פעם שמשנים את רמת המשתנה הבלתי תלוי לבייס ליין. הרעיון הוא שיש מדידה בכל אחד מהשלבים שצוינו למעלה. לרוב, הסדרה הזו תסתיים עם הופעתו של תנאי B נוסף (ABAB)

נמדוד את האפקטים של המשתנה הבלתי תלוי על התלוי - כלומר, ההשפעה של הטיפול על ההתנהגות נמדדת על סמך השינויים שחלים בהתנהגות במעבר בין השלבים השונים. אם כל פעם שנפעיל את B יהיה שינוי בהתנהגות ואז כשמחזירים ל A השינוי יעלם זה ביטוי ל reversal ABA.

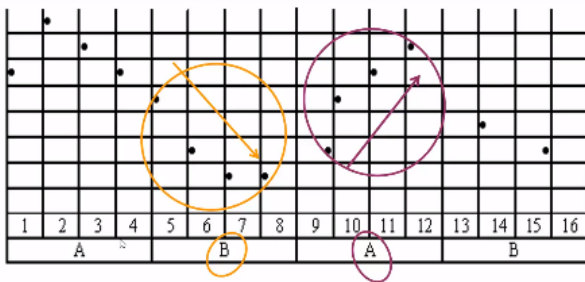
דוגמא- דני בן ה 3 נוהג להתעורר במהלך הלילה בצרחות עזות, ומסכים להירגע רק כאשר הוריו מעבירים אותו למיטתם. כאשר דני נרגע ומוחזר למיטתו חוזרת ההתנהגו (מתעורר שוב בצרחות).

השערת הניסוי היא שההתנהגות של הוריו של דני מחזקת ומשמרת את התנהגותו. הוריו של דני החליטו לערוך ניסוי לבחון את ההשערה. במקרה הניסוי הזה:

- A תנאי baseline - ההורים לוקחים את דני למיטתם.

- B תנאי הטיפול- ההורים ניגשים לדני רק כאשר הוא נרגע - הם נותנים לו חיזוק אשר מספק יחס ותמיכה בכל פעם שהוא חוזר בו מההתנהגות שהציג .

- ההתנהגות הנמדדת - הצרחות של דני .



כל כמה לילות, מחליפים את התנאי A או B, כלומר - כל כמה לילות תגובת ההורים משתנה.

בטבלה זו תצוגת התוצאות: ציר X זה מספר הלילות (16 סהכ), כל 4 לילות השתנה התנאי, ציר ה Y הוא מדד הצרחות בכל לילה.

- בשלב A ההורים ניגשים לדני בכל פעם שהוא צורח והוא מתעורר בצרחות פעמים רבות במשך הלילה.
- בשלב B ההורים ניגשים לדני רק כשהוא נרגע, ואכן רואים ירידה במספר הפעמים שהוא מתעורר בצרחות.

נתונים מהסוג הזה מציעים קשר סיבתי, בין היחס של ההורים (המשתנה הבלתי תלוי) להתנהגות של דני (המשנה התלוי). על פניו, נראה שיש עדות לקשר סיבתי בעיקר בגלל החזרתיות.

איומים על המערך:

1. כמו מערך תוך נבדקי עם בדיקות חוזרות. תוקף חיצוני נמוך (מאוד) בגלל שיש נבדק אחד, יש סיכוי ש"נפלנו" על נבדק שלא מייצג את כלל האוכלוסייה.
2. המבחנים הסטטיסטיים חזקים פחות בד"כ, אם בכלל יש צורך לקיימם. אי אפשר לתת אומדנים לממוצעים באוכלוסייה והערכות סטטיסטיות אחרות שאנחנו רגילים אליהם, כלומר - המבחנים הסטטיסטיים חלשים יותר, וזה נובע מכך שגם יש פחות מדידות מאשר בניסוי רב נבדקים.

לסיכום, המערך הזה הוא לא רע כמחקר חלוץ, בעיקר במקרים בהם עלות הטיפול גדולה.

Randomized time series

משתמשים במערך זה כאשר אי אפשר להפעיל ולהפסיק את המניפולציה, אפשר להפעיל את המניפולציה רק פעם אחת ואי אפשר להפסיק אותה (כלומר אי אפשר לחזור לבייס ליין). לדוגמה - אפשר לחתוך את הקורפוס קלוסום כדי להוריד תדירות התקפים אפילפטיים רק פעם אחת, ואי אפשר להחזיר את המצב לקדמותו.

במערך זה עורכים מדידות רבות של המשתנה התלוי לפני ואחרי המניפולציה. כך ניתן להעריך את השינוי בממד זה לאורך זמן, ולראות אם האפקט של זמן על הממד השתנה בעקבות המניפולציה.

חסרונות -

- כמו מערך לפני אחרי עם קבוצה אחת (איומים של היסטוריה, בשילה, מדידה, רגרסיה לממוצע - תלוי איך נבחר הנבדק, מכשור).
- מערך זה התוקף החיצוני נמוך יותר.

יתרונות-

- המדידות הרבות מקטינות מעט את האיומים הללו (בהשוואה למדידה אחת).
- התמקדות בנבדק אחד, אם הדבר מתאים לצרכים של החוקר.

Multiple baseline designs

המערך מתאים למקרים בהם ה Reversal - של מערך ABA אינן ישימות או אינן אתיות.

למשל, כאשר התנהגות הבייס ליין של ילד היא פציעה עצמית, או שהחוקר הצליח לשפר הישגי תלמיד. במקרים אלו, לא הגיוני "להפוך" את התנאים על מנת לבחון את הסיבות לשיפור.

זוהי משפחה של מערכים שונים, שרק בחלק מהם יש נבדק אחד בלבד. במקרים רבים מהם מדובר למעשה במספר נבדקים שעל כל אחד מהם עושים ניסוי של single subject.

הרעיון הכללי – אין לנו דרך לתפעל את המשנה הבלתי תלוי מספר פעמים לנבדק אחד, אז אנחנו רוצים למצוא דרכים להגביר

את הביטחון שלנו בכך שההבדלים בין המדידות לפני ואחרי המניפולציה נגרמו בשל המניפולציה.

סוגים של מערך מחקרי זה:

1. **Across behaviors** - התנהגויות שונות של אותו נבדק משוות.

נבדק אחד בלבד, בייס ליין מרובה - יותר ממשתנה תלוי אחד (מגוון מדדים תלויים).

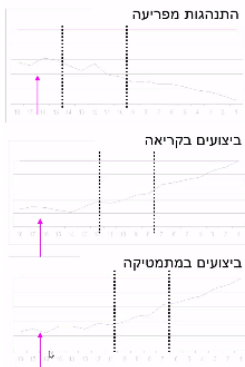
דוגמא- הנבדק הוא יוני, נער בעל ביצועים נמוכים במתמטיקה וקריאה, אשר נוטה להפריע בשיעור. הפסיכולוגית משערת שהמורה מתייחסת אל התלמיד רק כשהוא מפריע ולא מחזקת אותו כשהוא מצליח. נרצה לבחון את השפעת היחס של המורה על התנהגות התלמיד.

המשתנה הבלתי תלוי (יחס המורה) מוצג בשלוש רמות :

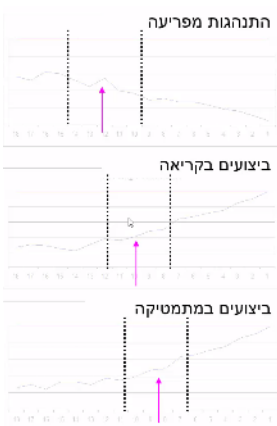
1. התייחסות להפרעות, התעלמות מהישגים
2. התעלמות מהפרעות, התייחסות להישגים רק במתמטיקה
3. התעלמות מהפרעות, התייחסות לכל ההישגים

כדי לעשות את זה נצטרך לכלול שלושה משתנים תלויים: כלומר ניצור הגדרות אופרציונליות לשלושה משתנים שונים. בהקשר הזה, ריבוי בייס ליינים בא לידי ביטוי במדידה של כמה משתנים תלויים.

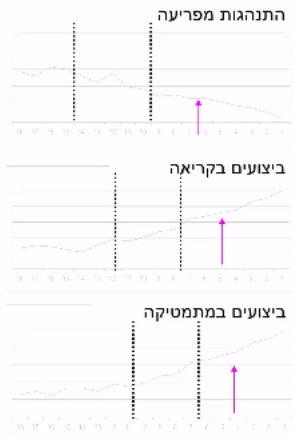
1. ההתנהגות השלילית של יוני
2. ביצועיו במתמטיקה
3. ביצועיו בקריאה



בשלב הראשון - הבייס ליין: המצב הנוכחי הוא שהמורה מענישה את יוני על התנהגות שלי ומתעלמת מהתנהגות חיובית, נרצה למדוד את כל המשתנים התלויים כאשר המורה ממשיכה בהתנהגות שלה (משתנה בלתי תלוי). ניתן לראות כי ההתנהגות השלילית גבוהה והביצועים במתמטיקה וקריאה נמוכים (לפי המדידה).



בשלב השני - נשנה את המשתנה הבלתי תלוי (התנהגות המורה). המורה כעת תתעלם מההפרעות ותייחס רק להישגים במתמטיקה. ניתן לראות לפי המדידה כי ההתנהגות המפריעה פוחתת בעוד שהביצועים בקריאה ומתמטיקה משתפרים.



בשלב השלישי - נשנה שוב את המשתנה הבלתי תלוי (התנהגות המורה). (הפעם, היא תתעלם מההפרעות ותתייחס להישגים. במדידה רואים כי ההפרעות פוחות וההישגים גוברים.

ניתן לראות שהתוצאות מתאימות בדיוק להשערה - יחס המורה משמעותי להתנהגות הילד. אמנם לא היו מניפולציות חוזרות אבל ההתאמה המושלמת של כל המשתנים מחזקת את אמונתנו בבכונות ההשערה של הפסיכולוגית.

- היתרון המרכזי - ריבוי התנהגויות מקטין חשש למקריות, כל התנהגות מוסיפה מידע עד כמה סביר שהשינויים שראינו בהתנהגויות האחרות מקריים.
- החיסרון המשמעותי - יתכן שרק קשר המדדים למניפולציה מקרי, ייתכן שהמדדים קשורים זה לזה, והקשר שלהם למניפולציה מקרי לניסוי.
- החיסרון הזה יכול למשל להיות ברור בדוגמה שנתנו - לדוגמה היה שיפור גם בקריאה וגם במתמטיקה למרות שבשלב השני היו חיזוקים רק להישגים במתמטיקה ולא בקריאה, ולכן יכול להיות שהמדדים קשורים זה לזה באיזשהו אופן שלא תומך בהשערת המחקר.

נקודות חשובות:

- להפעיל מניפולציות (ולהפסיק לבסוף את המדידה) רק אחרי שנראה שהמדדים הגיעו ליציבות.
- לבחור נקודת זמן מקרית לתחילת הטיפול, שכן היא מקטינה איומים של בשילה או היסטוריה.

2. Across settings and time - על אדם בודד, כאשר הטיפול ניתן במקומות/זמנים שונים.

דוגמה - נבדק הסובל מפיגור שכלי, עבר טיפול להפחתת ריוח. הטיפול בוצע בכיתה, אך אז נמצא שינוי רק בכיתה. אז הרחיבו את הטיפול גם למצבים שבהם הייתה לו אינטראקציה חברתית בקהילה, ולאחר מכן גם לשיעור בישול בו השתתף. כלומר, השתמשו באותו מערך מעבר לנסיבות ומעבר לזמן.

התוצאות הראו שמייד לאחר הטיפול בכל קונטקסט בנפרד, נצפתה התחלה של ירידה בריוח. תוצאות אלו מחזקות את תוקף המסקנה שהטיפול עזר - כי הפעלנו מניפולציות בזמנים משתנים במקומות שונים ובכל התנאים האלו הייתה ירידה בריוח. במערך זה הזמנים השונים מקטינים את איום ההיסטוריה.

- היתרון - ריבוי ההקשרים מקטין חשש מקריות או משתנים מתערבים.
- החיסרון - בעיקר בתוקף החיצוני והעוצמה הסטטיסטית- אין כמעט משניהם, לרוב, אין ממש הרבה הקשרים, אז הפרדת המקריות עדיין דורשת אפקט חזק ומשכנע במיוחד.

3. Across individuals - על מספר נבדקים, שעוברים את הטיפול בזמנים שונים.

דוגמה - ארבעה נבדקים עברו טיפול נגד עישון באינטרנט, לאחר הטיפול הייתה עוד תקופה שבה קיבלו חיזוקים חיוביים על ההתמדה. לאט לאט הם הפסיקו לקבל חיזוקים, ולבסוף החיזוק נעלם לחלוטין (נקרא (shaping). ארבעת הנבדקים עברו את הטיפול בזמנים שונים, על מנת להקטין את איום ההיסטוריה והבשילה.

זוהי תבנית ABA ולא בדיוק AB, אבל פעמים רבות זה המצב כי אי אפשר להמשיך בטיפול בלי סוף (למשל במקרים של טיפול פסיכולוגי).

התוצאות - אחד הנבדקים הראה ירידה בעישון עוד לפני הטיפול בשלב מדידת הבייס ליין, אם היינו מסתכלים רק עליו היינו מתקשים לדעת מה השפעת הטיפול, כיוון שהוא התחיל לדת כבר לפני הטיפול. אבל, כשמסתכלים על כל התוצאות של כל הנבדקים נראה שבאופן כללי, התחילה ירידה עם זמן הטיפול.

אפשר להתייחס למערך זה באופן עקרוני כאל מערך לפני-אחרי עם קבוצת ביקורת עם הקצאה אקראית. אבל יש נקודות זמן בהן חלק מהנבדקים עברו טיפול וחלק לאף וביצוע הטיפול הוגרל אקראית.

- במערך זה יתרונות רבים של המערך התקין לפני - אחרי עם הקצאה מקרית (בשילה, היסטוריה, ברירה, נשירה ורגרסיה). מעשית, יש מעט נבדקים ובוחרים נקודות זמן שונות רבות (ויש לא מעט רעש במדידה של האפקט בשל כך).
- חסרונות נוספים הם כמות הנבדקים הקטנה שמקטינה את העוצמה הסטטיסטית, ומאפשרת איזומי היסטוריה ובשילה שהשפיעו על חלק מהנבדקים וכן הטו את התוצאות.

סיכום מערכי ניסוי לנבדק בודד:

יתרונות עיקריים

- בזכות מדידות חוזרות רבות, היכולת להסיק סיבתיות חזקה למדי
- נשתמש כאשר מעניין אותנו אדם אחד בלבד

חסרונות עיקריים

- קושי להכליל לכלל האוכלוסייה
- עוצמה סטטיסטית נמוכה לעיתים קרובות

חשוב להגיד שלפעמים העובדה שאנחנו מודדים הרבה פעמים בתוך כל נבדק, זה נותן לנו יתרון בעוצמה הסטטיסטית, לעומת מחקר תקין עם מדגם קטן ומדידות מועטות.

איך נתמודד עם החסרונות ?

שחזורים, יש לציין שזה לא פתרון מושלם ולעיתים הוא יקר. אבל שחזור הם אחד העקרונות העיקריים במדע, תמיד נרצה שהמחקרים שלנו יוכלו להשתחזר, ובמקרה הזה נרצה להסיק יותר מאשר על מקרה בודד בפעם אחת. אם ניסויים לא אפשריים לשחזור, אנחנו נטיל בהם ספק, לא נוכל להסיק או להכליל ללא שחזור (גם אם הטיפול עזר במקרה בודד).
כשנעשה שחזורים במערך ניסוי לנבדק בודד נוכל לשאול כמה סוגי שאלות :

1. אם נחזור על הטיפול עם אותו אדם, האם נוכל לשחזר את התוצאות? האם במשך הזמן נוכל להכליל על אדם זה?
נעשה Direct replication
2. האם הטיפול יהיה אפקטיבי עבור אנשים שונים בעלי אותה הבעיה? האם נוכל להכליל את הטיפול מעבר לאדם הספציפי?
נעשה Systematic Replication
3. האם הטיפול יהיה אפקטיבי בתנאים שונים? האם יפעל אצל מטפלים אחרים או בעיות פסיכולוגיות שונות מהבעיה הספציפית עבורה נבחן?
נעשה Clinical replication

הקהילה המדעית בפסיכולוגיה

האנשים

מי מעורבים בתהליך ייצור הידע:

- נבדקים - הנבדקים הם אבן הבניין החשובה ביצירת ידע מדעי הפסיכולוגיה. מאיפה מגיעים הנבדקים:

- האוכלוסייה הנפוצה הם סטודנטים לפסיכולוגיה כחלק מדרישות קורסים. win-win.
- סטודנטים נוספים מחוגים אחרים עם תמורה מסוימת.
- אנשים נוספים, על ידי גיוס בפאנלים באינטרנט וכו'. גיוס באינטרנט הולך ותופס תאוצה - נרשמים לאתרים, מבצעים מטלות תמורת תשלום.

מחקר שבו משתתפים נבדקים שגויסו באינטרנט חשוף לאיזומים: איום על תוקף המסקנה הסטטיסטית (חשיפה לטעות מסוג שני) כי הנבדקים מבצעים את הניסויים בסביבות שונות, מחשבים שונים עם מסכים שונים שגורם

- להבדלים בין הנבדקים. איום על תוקף חיצוני כי הנבדקים מגויסים באינטרנט מגיעים מאוכלוסייה עם נגישות לאינטרנט ולאתרים הללו ולא משקפת את כלל המין האנושי.
- ניסויים בהם החוקרים עצמם הם הנבדקים - מחקרים על כאב, מצבים בהם זה לא אתי לבקש מנבדקים לחוות את המניפולציה.
- אוכלוסיות מיוחדות (ילדים, אוכלוסיות קליניות וכו'..)
- בעלי חיים - סוגיה מורכבת.

רוב הניסויים שנמצא, בעיקר מחקרי עבר, התבצעו על נבדקים מאוכלוסייה זו. כיום יש הבנה שהרקע התרבותי משפיע על התוצאות. גם על ממצאים קוגניטיביים קלאסיים כמו תפיסה (מולר ליר).
 • WEIRD – Western, Educated, Industrialized, Rich & Democratic
 ייתכן שהרבה מאוד מהידע הפסיכולוגי שלנו מוגבל לאוכלוסייה בלבדית אחת, ייתכן והמסקנות לא תקפות לכל המין האנושי.

● **נסיינים** - מתחלק לחוקרים, סורקים ועורכים. מיהם הנסיינים:

- סטודנטים במהלך או לאחר תואר ראשון בפסיכולוגיה.
- נסיינים בתשלום.
- החוקרים עצמם.

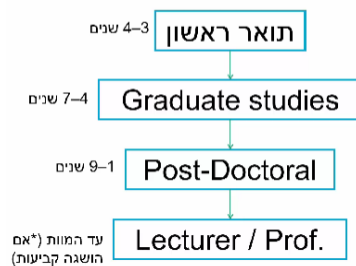
חוקרים:

- ראשי מעבדה (מרצים), הרבה פעמים אין הכשרה בהוראה. המטרה היא להכשיר את הדור הבא.
- חוקרים בתשלום במעבדה (פוסט דוקטורנטים)
- תלמידי מחקר graduate students - דוקטורנטים, מאסטרנטים
- לפעמים גם תלמידי תואר ראשון undergraduates

סורקים: reviewers חוקרים בתחום, בהתנדבות. מאמר צריך לעבור סקירת עמיתים לפני הפרסום.
עורכים: editors תלוי כתב עת, לעיתים בהתנדבות (מהתחום) ולעיתים בתשלום (עורכות מקצועיות).

תהליך ההכשרה

מסלול התפתחות מקובל, יש המון יוצאי דופן:



תואר ראשון:

- ידע בסיסי בפסיכולוגיה (מבואות)
- ידע מתקדם (קורסי בחירה, סמינר, סדנאות)
- כלים לעבודת מחקר (סטטיסטיקה, ניסויית, מחקר מודרך, סדנאות)
- תואר שני ודוקטורט:
- ידע מתקדם בפסיכולוגיה (קורסים, קריאה)
- כלים מקדמים לעבודת מחקר (סטטיסטיקה, מחקר מודרך, סדנאות)
- ניסיון מעשי בעבודת מחקר
- תרומה משמעותית לקהילה המדעית (תזה) פחות מקובל לעשות דוקטורט כדי לקבל את התואר וזהו, נהוג להמשיך לחקור. פוסט דוקטורט: בהתאם לאיכות החוקר (לפי מספר ואיכות הפרסומים), התאמה, משרות פתוחות.
- עיסוק במחקר, לצבירת עוד ידע וניסיון
- היכרות עם מעבדות אחרות
- עצמאות מחקרית

חוקר במעבדה - מרצה: משרה באוניברסיטה (לרוב) בה החוקר מקבל מעבדה, אחריות להנחיית תלמידים ובדכ התחייבות לעסוק בהוראה. לרוב הקריטריונים הם מספר הפרסומים המדעיים, התאמה למחלקה, job talk. בהתחלה כמרצה, לאחר מכן כמרצה בכיר. הקביעות מאפשרת חופש מדעי-אקדמי.

מי משלם על המחקר - תקציבים מהמדינה, מלגות מגופים פרטיים...

תהליך ייצור הידע

סוגי פרסומים שונים מתאימים למחקרים שונים - ספרים, מאמרי סקירה, מאמר דיווח על מחקר, מאמרי תיאוריה...

הפרסום המדעי (מאמר) - הדרך לשתף את הממצאים והתיאוריות עם שאר הקהילה המדעית.

איך נולד פרסום מדעי:

- מרעיון למאמר - רעיון למחקר, תכנון הניסויים, ביצוע ניסויים וניתוח התוצאות, הסקה מהתוצאות ובתיבת המאמר.

- מבנה המאמר המדווח על מחקר:
- כותרת
 - תקציר
 - מבוא
 - רקע (תיאוריה + ממצאים קודמים)
 - תיאור המחקר הנוכחי
 - ההגיון מאחוריו (כיצד יתרום?)
 - השערות אפשריות
 - השיטה (מה עשיתי ואיך)
 - תוצאות המחקר (מה מצאתי)
 - מסקנות המחקר (מה למדת/ מה לא למדתי)
 - ביבליוגרפיה



- ממאמר לפרסום - נהוג לפרסם בירחונים מדעיים ולא בבלוגים. הרבה פעמים מפרסמים pre-print לפני ההגשה הרשמית, כדי לקבל חווד ראשוני לפני ההגשה.

- הגשת המאמר - בחירת כתב העת (journal) - עד כמה הדורנל נחשב נקבע לפי ה impact factor שלו - ציון

שנקבע לפי מספר המאמרים שמצוטטים מתוך הדורנל, מראה עד כמה הדורנל משפיע בתחומו.

- ביקורת עמיתים - המאמר מגיע לאחר ההגשה לעורך ראשי, מחליט אם שווה להתעמק בו או לא (desk rejection) -> לאחר מכן מגיע לעורך בפועל, אם הוא חושב שהמאמר שווה -> הוא שולח לסוקרים. הן משתמשות בידע שלהן אם המושגים במאמר תקפים מספיק, אם המסקנות מחדשות מספיק בהתאם לסטנדרטים של הדורנל. -> הביקורת מגיעות לעורך בפועל שמעביר את התשובה לכותבי המאמר -> אם יש הרבה מידי תיקונים ואי אפשר לפרסם או מהם התיקונים שיש לבצע כדי לתקן.

תהליך הביקורת תומך בחיזוק הממצאים והמסקנות. זהו תהליך ארוך.

- אחרי הפרסום - הביקורת נמשכת (חיובית ושלילית). בעיקרים מאמרים אחרים שמתייחסים למאמר שלנו, אבל גם post publication peer review. ככל שהמאמר מצוטט יותר הוא נחשב חשוב יותר ותורם יותר למדע. אבל גם טעות חשובה תהיה מצוטטת. לא משתמשים הרבה בהבחנה בין ציטוט שמפריך לבין ציטוט שמחזק.

איך משתפים ידע מדעי:

- פרסום מאמר מדעי
- שיתוף מידע בכנס מדעי. יש גם כנסים וירטואלים, קצת פחות אפקטיבי אבל עם הרבה פוטנציאל.
- pre-print
- רשתות חברתיות (יש גם ייעודיות למחקר בתחומים שונים).

התפתחויות אחרונות

יש הרבה ביקורת על תהליך הפרסום:

1. איטיות
2. מהימנות נמוכה בין שופטים (reviewers)

פתרונות:

- **post publication peer review** - מנסה להתגבר על שתי הביקורות הראשונות.

3. העדפה לתוצאות מובהקות על פני תוצאות לא מובהקות גורמת להטיה בפרסום (בעיית המגירה) - מחקרים רבים שבהם לא נמצאו אפקטים מובהקים, או שהמדגם היה קטן מידי כדי לקבל השערת אפס (בעיה מסוג שני). ייתכן יש הרבה חוקרים שנמצאים באותו המצב לגבי אותם הנושאים ואם החוקרים היו משתפים אחד את השני גם את התוצאות בעלות ההסקה הסטטיסטית הנמוכה היה לקחת בחשבון את כל המחקרים ואולי ניתן היה לשרר את היכולת להעריך את הקשר בין המשתנים למציאות.
פתרונות לבעיית המגירה:

- מטה אנליזה - מחברים מספר גדול של ניסויים שבדקו קשר בין משתנים ומנסים להסיק מהם את גודל האפקט האמיתי, ואיזה משתנים ממתנים את הקשר. מחפשים את כל המאמרים הרלוונטיים וגם מפיצים הודעות לחוקרים כדי למצוא מחקרי מגירה. הבקשות לא תמיד נענות, לפעמים חוקרים לא טורחים לשלוח וכו'...

אפילו שלא נלקחים בחשבון כל מחקרי המגירה זה כן עוזר להתגבר על הבעיה כי מחשבים גם את חוזק העדויות שיש עד כה, מקבלים ממצא שאומר כמה מחקרי מגירה בהם לא נמצא אפקט, נדרשים על מנת להפריך את האפקט שנמצא.

- **registered reports** -יפורט בהמשך

4. מעבר לכך - יש העדפה למסרים חד משמעיים ברורים, הפרסום נעשה סלקטיבי ומוטה.

5. הפרסום נעשה היעד ולא ההתקדמות בתחום המדעי - **incentive structure** - קריטריונים מרכזיים לקבלת משרה, קביעות וקידום הם מספר המאמרים, מספר המענקים ומספר הציטוטים. קבלת מענק מחקר או פרסום מאמר נקבע פעמים רבות לפי כמה המחקר חדשני.

זה נותן תמריץ לפרסם כמה שיותר מאמרים, כמה שיותר מחדשים. מעודדים מדע לא טוב, דרגות החופש של החוקר, הטיות...

פתרונות:

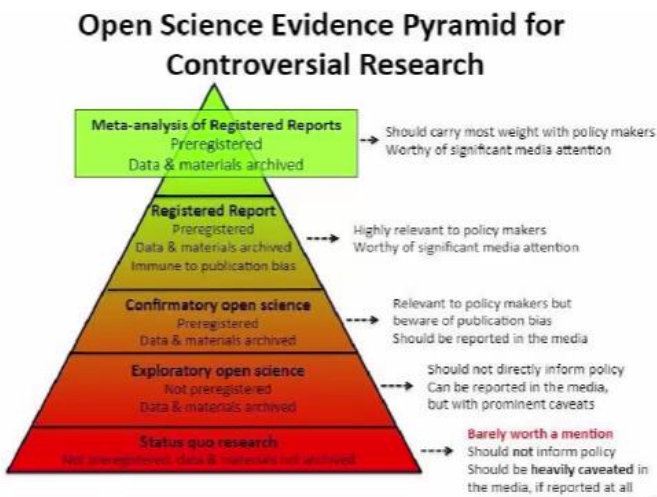
- **open science**

- ביקורת עמיתים מתומצמת **incentivised peer review** - לתמרץ אנשים (לשלם להם) לפרסם מאמר לפני שהוא נשלח לפרסום, לעודד אנשים להגיב ולחפש טעויות במחקר.

- **registered reports** - נותנים אפשרות לביקורת עמיתים נוספת, עוד לפני שאוספים ומנתחים את הנתונים, על המבוא והשיטה בלבד. אם זה עובר ומתקבל לפרסום ניתן להמשיך לאסוף נתונים ולנתח תוצאות. ככה התוצאות יתפרסמו גם אם הן לא מובהקות (נפתרת בעיית אפקט המגירה). מודל חדשני ועתידני.

יתרונות - הפרסום לא תלוי בתוצאות; מונע **harking**; מפחית דרגות חופש של חוקר. חסרונות - מתאים יותר למחקר אישושי (פחות למחקר גישושי).

אפשר גם לשלב את כל השיטות והפתרונות:



אתיקה במחקר

מחויבות אתית כלפי נבדקים

למה שנרצה להגדיר איך להתנהל מול נבדקים בצורה אתית? האם ההחלטה מה נכון ומה לא נכון היא לא פשוט החלטה טריוויאלית? מלחמת העולם השנייה הוכיחה אחרת. במלחמה ניסויים נרחבים של מדענים גרמניים בבני אדם כללו זוועות אמיתיות כולל כמה מהזוועות הכי גדולות שידע המין האנושי והם נעשו, כביכול, בשם המדע.

עד אז לא היו חוקים רשמיים שהיה אפשר לשפוט לפיהם, הבינו את זה כאשר רצו להעמיד לדין את אותם המדענים לאחר המלחמה וראו שאין שום חוקים רלוונטיים שמאפשרים העמדה לדין בגינם. במשפטי נירנברג לאחר המלחמה נוסח קוד נירנברג שמהווה עד היום חלק גדול מהבסיס לתקנות אתיקה.

הקוד מכיל בעיקר את שלושת העקרונות האלו:

1. הסכמה מודעת (ובהתנדבות) – נבדקים צריכים לדעת ולהתנדב למחקר מדעי.
2. יחס הולם של רווח מול סיכון – הרווח האישי לנבדקים יהיה מספיק גדול ושהסיכון יהיה קטן ככל הניתן.
3. הזכות להפסיק השתתפות ללא השלכות – אפשר להפסיק את הניסוי וזהו, זה לא ייצור להם בעיות.

בעיות אתיות לא התחילו עם מלחמת העולם השנייה, דוגמאות לבעיות אתיות אפשר למצוא למשל בעבודות של פרויד. למשל- איך שפרויד טיפל באמה אקשטיין בת ה-27 אמה באה לפרויד עם בעיית דיכאון קל. פרויד השתמש בה כדי לבחון כל מיני שיטות טיפול ניסיוניות, כולל צריבה של החלק הפנימי של האף שלה תחת הרדמה מקומית וקוקאין, טיפול שהקשר בינו לבין דיכאון מומצא. לא ברור למה פרויד עשה את הדברים האלו ומה הייתה המטרה שלו, וכמובן שאקשטיין לא ידעה שהוא עושה לה את הדברים האלה סתם כדי לבדוק.

בעיות אתיות קרו גם במחקרים יותר אמפיריים, למשל במחקר של ג'ון ווסטון והסטודנטית שלו רוזלי ריינר – הם רצו לבדוק אם אפשר להשתמש בהתניה קלאסית כדי לגרום לילדים לפחד מחפצים פרוותיים. הם גרמו לאלברט (הילד) לפחד מהדברים האלה באמצעות צימוד רעשים מאוד חזקים להופעה של עכברוש. בהמשך נראתה הכללה של הפחד הזה. כמובן שיש בעיות אחרות עם המחקר הזה- לא ברור אם ההורים של אלברט ידעו שהוא חלק מהמחקר, והסכנה של אלברט דיי גבוהה ביחס למה שהוא יכול להרוויח מהניסוי הזה וכו'.

זה שהסכימו על קוד נירנברג לא הפך את המחקרים לאתיים, גם אחרי הניסוח של הקוד היו לא מעט מחקרים בעייתיים. אחד המפורסמים שבהם הוא ניסוי הכלא של סטנפורד. בניסוי סטודנטים הוקצו לתפקידי סוהרים ואסירים, כיום אנחנו יודעים שהנסיינים שערכו את הניסוי עודדו את האסירים בצורות מסוימות, ואת הסוהרים במיוחד להתנהג בצורות מאוד מסוימות. אחרי זמן קצר הסוהרים התחילו להתעלם באסירים בצורה מילולית, פסיכולוגית ופיזית. כמובן שזה לא עונה לשום קריטריון של תועלת מול נזק- הנבדקים סבלו מאוד. יכול להיות שיש ערך לניסוי אבל לא הייתה שום שמירה על האתיקה. היו עוד לא מעט ניסויים אחרים שהיו לא אתיים.

בעיות קרו גם בניסויים שהובילו למסקנות חיוביות על העולם, כמו הניסוי של מוזפר שריף עם מערת השודדים. החוקר הוציא נערים למחנה קיץ ופיצל אותם ל 2 קבוצות מתחרות. הוא דאג שתהיה תחרות קשה בין הקבוצות גם באופן מלאכותי, וגרם להמון מתחים בין הקבוצות. הניסוי נגמר בפעילות לטובת מטרה משותפת מה שהראה שאפשר להתגבר על מחלוקות גם במקרים של חילוקי דעות קשים- המסקנה היא חיובית. הבעיה היא שהילדים וההורים שלהם לא ידעו שהם בניסוי, הורים שלחו את הילדים למחנה קיץ שחשבו שהוא תמים, ובפועל החוקר השתמש בילדים האלה ועשה עליהם ניסוי. מעבר לכך, התברר שהניסוי המתואר היה הניסוי השני של החוקר להוכיח משהו כזה, הניסוי הראשון לא הראה את התוצאות להן הוא קיווה. גם במקרה הזה לפעולות החוקר היו השפעות לא פשוטות על חיי הנבדקים.

אתיקה ונבדקים - סכנות במחקר

נהוג לחלק את הסכנות הספציפיות שיש במחקר לשלושה סוגים של סכנות:

1. אי נעימות פסיכולוגית או פיזית

העקרונות האלו נכונים לכל מחקר שהוא ולא רק במחקרים פסיכולוגיים. כיום, אנחנו מבינים שעלינו כחוקרים יש אחריות להגן על הנבדקים מהסכנות האלה.

אנחנו רוצים לוודא שאנחנו חושפים נבדקים לסכנות רק כשמתקיימים שני התנאים הבאים:

- כשהנזק האפשרי נראה קטן והפיך בקלות יחסית
- כשהתועלת למדע גדולה יותר מהסכנה לנזק לנבדקים

יש פיקוח על הדברים האלה על ידי וועדת אתיקה, בארה"ב נקראות IRB. מדובר על ועדות המאשרות מחקרים. בארה"ב ובמספר מדינות אחרות החוק אינו מתיק לערוך מחקר בלי אישור של ועדת אתיקה. התפקיד של ועדת אתיקה הוא לחשוב אם הסכנות שנבדקים יעמדו בהם בניסוי מוצדקות, לנוכח הפוטנציאל של הניסוי לתרום למדע ולאנושות.

בישראל, כל זמן שלא מדובר בניסוי שמערב מניפולציות פסיכולוגיות או מדדים פיזיולוגיים חודרניים, אין חובה חוקית לקבל אישור אתי. אבל, בכל המחלקות לפסיכולוגיה באוניברסיטאות בארץ, החוקרים צריכים לקבל אישור אתי מאדם שאינו מעורב בפרויקט, בד"כ ועדת אתיקה של המחלקה.

הסכמה מדעת

כדי לצמצם את הסכנות, ההנחיות האתיות מחייבות חוקרים לקבל מנבדקים הסכמה מדעת באופן התנדבותי. כלומר, הסכמה מדעת זה מצב שבו אנחנו נותנים לנבדקים את האפשרות לבחור מראש אם להשתתף במחקר באופן התנדבותי. כדי שזו באמת תהיה בחירה התנדבותית צריך לתת לנבדקים את המידע הרלוונטי לגבי מה שיקרה בניסוי לפני שהם משתתפים בו, כדי שיוכלו להחליט אם הם רוצים להשתתף או לא. המידע המפורט של המחקר צריך להכיל גם פירוט של התגמול שהנבדקים מקבלים ממנו ומשך זמן המחקר.

יש מעט מחקרי שדה שבהם נבדקים מגויסים למחקר בלי שמודיעים להם על כך. עושים את זה רק במקרים ברורים שבהם אין סכנה לנבדקים. את המידע המקדים לנבדקים מעבירים בטופס הסכמה מדעת והנבדקים צריכים לקרוא ולחתום על הטופס. ילדים זקוקים (גם) לאישור הוריהם. בנוסף, יש לאפשר לנבדקים לפרוש מהמחקר בכל שלב מבלי שתהיה לכך השלכות שליליות לגביו.

פלישה לפרטיות

טופס ההסכמה צריך לפרט גם איך הפרטיות של הנבדק תישמר ואיך הנתונים שיאספו עליו ישמרו באופן סודי. ברוב הניסויים במצב הוא שלא ניתן לקשר בין הנתונים שנאספו לבין זהות הנבדק. לפעמים, אנחנו כן רוצים לשמור על קישור בין הנתונים לבין זהות הנבדקים (למשל ניסוי שצריך לבוא אליו מספר פעמים וכדומה), במקרים אלו יש להגביל את הגישה לנתונים לחוקרים ולעוזריהם בלבד.

בנוסף, יש לצמצם את הפגיעה בפרטיות במידת האפשרי, ורק במידה ההכרחי (שמות הנבדקים לרוב לא רלוונטים, נטיה מינית וכו'). לעומת זאת, אם אנחנו רוצים לצלם נבדקים למחקר, חשוב לוודא שהמידע נשמר בפרטיות והוא לא מועלה לשירותי ענן, נשמר בצורה מוצפנת שנגישה רק לעורכי המחקר. אם אנחנו מצלמים נבדקים מתוך מטרה לשתף את הקהילה המדעית, חשוב לציין את זה בטופס ההסכמה בצורה מובנת.

הונאה

- לפעמים אין אפשרות לחשוף את כל האמת על המחקר לנבדקים לפני השלמת ההשתתפות בו. אי אמירת כל האמת על השערות הניסוי זו הנורמה בניסויים בפסיכולוגיה וזה נובע מחשש שאם הנבדקים ידעו מה השערת המחקר הם יתנהגו באופן שונה ומותאם להעשרת הניסוי (רציית ניסוי למשל) ואז יפגע התוקף החיצוני של הניסוי. יכולה להיות גם פגיעה בתוקף הפנימי של הניסוי- לא נדע אם המניפולציה שעשינו הפעילה רק את המשתנה הבלתי תלוי או אולי גם את ההשערות של הנבדק לגבי ההשפעה של המשתנה הבלתי תלוי. במקרים כאלה יש חובה אתית לספר לנבדקים בסוף הניסוי מה הייתה השערת הניסוי.
- הונאה אחרת היא מתן מסר שקרי לנבדקים, זה פחות שכיח כי שקר עשוי לפגוע בנבדקים וכי חוקרים לא רוצים שנבדקים בפסיכולוגיה תמיד יחשדו שמשקרים להם. משתמשים בשקרים רק כמוצא אחרון ולפעמים יש צורך להשתמש בשקרים

למשל- שאלת מחקר שקרית, מידע שקרי על המציאות, שימוש במשתפי פעולה, מידע שקרי על הנבדקים.

בכל מקרה של הונאה, חובה לחשוף את ההונאה בהקדם האפשרי. חשוב לעשות זאת בעדינות, ובדרך שלא תגרום לנבדקים להרגיש מטומטמים ושעבדו עליהם. כדאי להציג את ההונאה כמשהו קטן שלא הייתה שום ברירה אלא לעשות אותו. עדיין, צריכים להיות קשובים לתגובות שליליות מצד הנבדקים, ולהבהיר שלא הייתה ברירה אלא לשקר, ולגלות הבנה לתחושות הנבדקים. לפעמים יש חשש שעם גילוי השקר, הנבדק יספר לחברים שלו שיכולים להיות נבדקים פוטנציאליים אחרים, ואז זה יתגלה לנבדקים לפני שיושלם הניסוי כולו. לכן, מקובל לבקש מנבדקים לא לשתף את המידע על ההונאה עם חברים שלהם. במקרים מאוד יוצאי דופן כשההונאה היא לא פוגענית והחשש שהשקר יתגלה לנבדקים פוטנציאליים ויפגע בניסוי, אפשר לחכות עם חשיפת השקר עד שהשלמנו את כל תנאי הניסוי, ואז בסוף הרצת הניסוי צריך ליצור קשר עם נבדקים ולהסביר להם את מטרת הניסוי ולחשוף את ההונאה.

- ההונאה האחרונה שאפשר לעשות היא מתן פידבק שלילי, שזו ההונאה המסוכנת ביותר. בניסויים שעשו על הונאות מהסוג הזה, מצאו שגם אחרי שנבדקים הבינו שהונו אותם, הם עדיין חשבו שהפידבק השלילי נכון במידה מסוימת. יש כמה הסברים לתופעה הזאת- אחד מהם הוא, שפידבק שלילי גורם לנבדקים לחשוב על עוד מקרים בהם הם חשו שהיכולת שלהם מוגבלת וזה עוזר בגיבוש האמונה שהפידבק השלילי הולם את הנבדקים. לנבדקים יש מוטיבציה לאשש את הפידבק השלילי. כדי להפחית את הסכנה הזו אפשר לבקש מהנבדקים לחשוב איך היו מרגישים אם היו מקבלים פידבק חיובי, פעולות אלו נמצאו כמפחיתות את השפעת הפידבק השלילי. שיטה אחרת, בעיקר עם ילדים, היא לסיים את הניסוי עם היפוך הפידבק השלילי לחיובי, למשל לספק לילדים הרבה פידבקים חיוביים בסוף ואז לספק שהפידבק השלילי היה שקרי. עדיין דורש לשקר.

מחויבות אתית בעריכת המחקר ובדיווח עליו

על המדען מוטלת האחריות להפעיל שיקול דעת כיצד לתרום באמצעות מחקר מדעי למדע ולאנושות. הרבה פעמים יכולות להיות מוטיבציות אחרות לעריכת מחקר כגון קידום אישי, קידום דעות קדומות אישיות וכו'. עלינו כמי שעוסקים במחקר מוטלת האחריות למנוע מהמטרות האלה לפגוע במטרה שלשמה נערך מחקר מדעי- הוספת ידע אמיתי על העולם.

זה בא לידי ביטוי בכמה אופנים:

- אמינות הנתונים - הכלל האתי הברור ביותר בו מדענים צריכים לעמוד הוא שהמדע שלהם יהיה אמין, ומטרת המדע היא להוסיף לחקר האמת, זה אומר למשל שהמצאת נתונים או עיוות שלהם זו עבירה אתית מאוד חמורה שעלולה להוביל לקטיעת קריירות.
- ניגוד עניינים - על החוקרים מוטלת האחריות לבצע את מחקרם ללא פניות והטיות. אסור ששיקולים זרים ישפיעו על אופן ביצוע המחקר, שיקולים זרים יכול להיות למשל ניגוד עניינים כספי, ניגוד עניינים תעסוקתי (מחקרים שיכולים לספק עבודה לעתיד למשל וכו'). בפרסומי המאמרים, חוקרים צריכים לחשוף כל עניין שיש בו, או שעשוי להיראות כי יש בו, משהו שיכול להטות את המחקר או להשפיע באופן לא ראוי על החלטות הנוגעות לניהול המחקר ותוצאותיו. בכל מקרה, השמטת מידע הינה עבירה משמעותית חמורה.
- מקור אחר לניגוד עניינים הוא מבנה התמריצים באקדמיה. לחוקרים באקדמיה יש אינטרס לפרסם כמה שיותר מאמרים כדי לקבל קידומים, לקבל יותר יוקרה ויותר מעמד. יש לא מעט אנשים היום שקושרים את משבר הרפליקציה למבנה התמריצים הזה. כשמגיעים מהפרספקטיבה הזאת אפשר להבין אולי מה הגורמים שנותנים לחוקרים מוטיבציה לבצע פעולות שהזכרנו כשדיברנו על תוקף המסקנה הסטטיסטית ועל הקהילה המדעית (למשל פרסום מאמרים שהקשר ביניהם לבין המציאות רופף). לסיכום, ניגוד עניינים עלול להוביל להרבה התנהגויות שעלולות לפגוע בהתקדמות המדעית.
- אתיקה בכתיבה - הנושא הזה נכון לא רק למאמרים מדעיים אלא לכתיבה בכל הקשר אקדמי. כשאנחנו כותבים אנחנו רוצים תחילה לקרוא חומרים אחרים, לחשוב בעצמנו ואז לכתוב. אנחנו תמיד רוצים לכתוב את המחשבות שלנו במילים שלנו, ובהתאם אנחנו צריכים לתת קרדיט לאחרים כשאנחנו מציגים את המחשבות שלהם בכתיבה שלנו. זה נכון בדברים כתובים וזה נכון בחומר ויזואלי שאנחנו מציגים. במילים אחרות- אסור לרמות, להעתיק ממקורות אחרים ללא קרדיט, להעתיק ולהדביק טקסטים שנכתבו על ידי אחרים או תרגומים שלהם ללא

- Garden of forking paths
- File drawer problem
- HARKing (hypothesizing after results are known)

מחויבות אתיות בשימושים במחקר

• Plagiarism: "the practice of taking someone else's work or ideas and passing them off as one's own" (Oxford Dictionaries)

לאחר ביצוע המחקר, כתיבתו ופרסומו לא נגמרת אחריותו של החוקר. גם איך מה שחוקרים כותבים במאמר שלאחר הניסוי וגם איך שהם כותבים אותו יכול למשל לפגוע באוכלוסיות מסוימות ולכן הם צריכים לקחת על זה אחריות. חלק מהאחריות האתית למחקר קשורה גם לאיך שאנשים אחרים מפרשים את המדע שאנחנו מייצרים ואיך משתמשים בו. מעורבות ומודעות יותר עמוקה ומקיפה של מדענים, גם אחרי פרסום המאמר, אמורה להבטיח התנהלות יותר אתית ופגיעה בפחות אנשים בעולם.

מחויבות אתית בשימוש במחקר יישומי

מחקר שנעשה לא בכותלי האקדמיה אלא בעולם האמיתי- למשל התערבויות חינוכיות. יש פה שטח שהוא לא ממופה, יש מעט מאוד עבודה על האתיקה של מחקר יישומי ושל השימושים במחקר באופן כללי.

- אפשר להשתמש בידע מחקר לעשות טוב (או רע) לעולם
- אפשר להשתמש בידע מחקר כדי להגדיל מכירות למשל
- יכולות להיות גם סוגיות של פגיעה בנבדקים- במחקר יישומי אפשר לפגוע למשל בפרטיות, פעולות שמנוגדות לרווחת הנבדקים, או השתתפות במחקר שלא מרצון ועוד.

יש כל מיני פיתוחים של כלים שעשויים לעזור לשמור על אתיקה במחקרים יישומיים. לדוגמא - the behavioral scientists ethics checklist.

סיכום הקורס

על מה דיברנו:

יתקפים

- מערכי מחקר שגויים
- מערכי מחקר תקינים
- מערכי מחקר קוואזי-ניסויים
- מערכי מחקר עם נבדקת אחת
- יאומים נפוצים על תקפות הניסוי
- ימדידה (מהימנות + תוקף)
- איך פועלת קהילת המדע בפסיכולוגיה
- אתיקה

המטרה של הקורס: ושל מדע באופן כללי, לאפשר לנו להשפיע על עצמנו, על אחרים, על הסביבה. נרצה לדעת מה משפיע על מה - השונות של איזה משתנה משפיעה על איזה משתנה שמעניין אותנו. איך משתנה בלתי תלוי משפיע על משנה תלוי. נרצה לדעת מה מידת ההשפעה של משתנה בלתי תלוי כלשהו, על משתנים נוספים.

אם יש משתנה שמעניין אותנו לדעת אם הוא משפיע על משתנים אחרים, נרצה לדעת האם הוא אכן משפיע וכיצד - כלומר, האם שונות בערכי המשתנה הבלתי תלוי X תגרום לשונות בערכי המשתנה התלוי Y.

רצינו לבדוק את טיוב אומדן ההשפעה ב 2 דרכים:

- שליטה בשונות המשתנה הבלתי תלוי
- שליטה במקורות השונות של המשתנה התלוי

על ידי:

- מיקסום השונות של המשתנה הבלתי תלוי
- בקרה על שונות משתנים חיצוניים לא רצויים
- הקטנת השפעה מקרית.

שליטה בשונות המשתנה הבלתי תלוי

מגדילה את היכולת להסיק סיבתיות

- האם שונות במשתנה X קשורה לשונות במשתנה Y?
- האם שונות במשתנה X משפיעה על שונות במשתנה Y?

יש הבדל גדול בין השאלות הללו - השפעה זה סוג אחד של קשר, והוא קשר חשוב ומהותי עבורנו. לכן עלינו לשלוט בשונות, כדי לדעת שזה קרה לא במקרה, לא בגלל משתנה שלישי.

הדרך הנפוצה ביותר במדע להסיק סיבתיות היא על ידי ניסוי:

- מחקרי שדה הרבה פעמים לא מאפשרים בחירה של מסגרת הבחינה, של המניפולציות או של המדדים
- מחקרים מתאמיים מאפשרים שליטה על כלי המדידה אבל לא על המשתנים הבלתי תלויים
- ניסויים הם כלי שמאפשר שליטה ובידוד של משתנים ומאפשרים הסקה סיבתית, על ידי שליטה בשונות המשתנה הבלתי תלוי.

המניפולציה שנעשית על המשתנה הבלתי תלוי היא מה שעוזר לנו להסיק שמדובר בסיבתיות. דיברנו על איך אפשר להסיק שהקשר שנמצא בין המשתנים נבע מסיבתיות (ולא מארטיפקט בלבד).

האיום המרכזי על הסקה סיבתית (ברמה האופרציונלית) הוא משתנים חיצוניים .

בחיים האמיתיים יש המון ארטיפקטים. יש המון אפשרויות להסברים חלופיים - משתנה חיצוני שמוצמד למשתנה שלנו ומקשה על הסקת סיבתיות.

איומים נוספים שדיברנו עליהם: רציית נסיין, הטיין נסיין ועוד...

האיומים המרכזיים על הסקה סיבתית ברמה התיאורטית הם: איום על תוקף חיצוני, איום על תוקף מבנה (הגדרת משתנים+תוקף מדידה מפוקפק).

המסר הוא - כדאי לא למחר ולהסיק מסקנות סיבתיות מהר מידי.

מחקר מתאמי לעומת מחקר ניסוי:

מערך ניסוי	מערך מתאמי
משתנה בלתי תלוי מתופעל	ייחוס / נמדד
משתנה תלוי נמדד	נמדד
משתנים מתערבים בקרה ניסויית	בקרה סטטיסטית
ניתן לטעון לסיבתיות	קשה לטעון לסיבתיות

ככל שיש לנו פחות שליטה על כל אחת מהנקודות הללו, ככה היכולת שלנו להסיק סיבתיות הולכת ופוחתת.

שליטה בשונות המשתנה התלוי

מגבירה את הדיוק במדידת הקשר (ואת הסיכוי שלנו לגלות את הקשר).

זה שניסוי יכול להעיד לנו על סיבתיות, לא אומר שהניסוי טוב בזיהוי הקשר - נרצה לזהות ולא להחמיץ קשרים בין משתנים.

לכן נצטרך למדוד את המשתנה התלוי בצורה טובה.

במצב אופטימלי היינו רוצים ששונות המשתנה התלוי תנבע רק משונות המשתנה הבלתי תלוי, כך גדלים הסיכויים שאם נמצא קשר ביניהם אז הוא קיים במציאות, וגם שאם לא נמצא קשר ביניהם - סימן שהוא לא קיים במציאות.

רוב המבחנים הסטטיסטיים משווים את השונות בין תנאי הניסוי השונים, לשונות הצפויה בין תנאי הניסוי בהינתן הידוע לנו על המשתנה התלוי (=אמדן שונות הדגימה).

$$Cohen's d = \frac{\bar{x}_1 - \bar{x}_2}{\hat{S}}$$

האפקט
גורם הטעות (הבדלים בינאישיים + טעויות מדידה...)

גורם הטעות יכול לנבוע מהבדלים בין אנשים שונים בתנאי הניסוי.

נרצה שהשונות של הגורם הבין אישי הזה תהיה חלשה, והשונות שנגרמת מהמניפולציה (שהיא הקשר בין המשתנה התלוי והבלתי תלוי) תהיה גדולה ככל האפשר.

יש 3 דרכים אפקטיביות לשלוט על השונות של המשתנה התלוי:

1. מקסימיזציה של שונות המשתנה הבלתי תלוי - זה יגרום להגדלת ההשפעה של המונה בלי להגדיל את המכנה.
2. בקרה על שונות משתנים חיצוניים לא רצויים - כדי להקטין את השונות המקרית, במכנה.
3. הקטנת השפעה מקרית - גם כן יקטין את השונות המקרית.

מקסימיזציה של שונות המשתנה הבלתי תלוי

שונות שיטתית = שונות במשתנה התלוי הנובעת מהמשתנה הבלתי תלוי.

הדרך לעשות זאת - שימוש בערכים קיצוניים של המשתנה הבלתי תלוי.

הגדלת השונות של המשתנה הבלתי תלוי למעשה מגדילה את חוזק המניפולציה, הרמות של המשתנה הבלתי תלוי שונות יותר אחת מהשניה, מה שישפיע על תוקף המסקנה הסטטיסטית.

בקרה על שונות משתנים חיצוניים

משתנים חיצוניים = משתנים בלתי תלויים שלא מעניינים את החוקר.

התייחסנו למשתנים חיצוניים כמשתנה שמוצמד למשתנה שלנו ומקשה על היכולת להסיק מה באמת גרם להבדלים שמצאנו.

זו לא הבעיה היחידה שמציבים בפנינו משתנים חיצוניים - הם גורמים לנו ליותר שונות, גם אם עשינו הקצאה מקרית, הם מגדילים לנו את מה שאנחנו מתייחסים אליו כהבדלים בין אישיים וטעויות מדידה. כלומר, הם עשויים להגדיל את אומדן השונות הצפויה בתוך הקבוצות, מה שפוגע ביכולת שלנו לאתר את השונות שנובעת מהמשתנה הבלתי תלוי. פגיעה בתוקף הפנימי.

הפתרונות:

1. א. הקצאה מקרית - מגדילה את הסיכוי שלא יהיה הבדל בין הנבדקים במשתנים החיצוניים. אם נמצא הבדלים בין הקבוצות יהיה קשה לשייך אותם למשתנה החיצוני, כי לא יהיה הבדל בין הקבוצות במשתנה החיצוני. יש חשיבות לגודל המדגם.

ב. התאמה (matching) - עדיין, יש גבול להתאים בכל משתנה חיצוני אפשרי, היכולת להשתמש בהתאמה מוגבלת. מבחינה מעשית כדי להתאים במשתנים חיצוניים יש למדוד אותם מראש ולהספיק לחלק את כל הנבדקים מראש, לכן מעדיפים לעשות הקצאה מקרית.

חולשות של פתרונות אלו: אמנם שונות המשתנים החיצוניים הופכת לשונות מקרית בתוך כל קבוצה (באופן מאוזן פחות או יותר) אבל עדיין מקטינות את היחס בין האפקט לשונות, עדיין יגדיל את השונות הצפויה בתוך קבוצות הניסוי גם תחת השערת האפס.

2. הפיכת המשתנה החיצוני למשתנה בלתי תלוי - להכניס אותו כגורם לניסוי. אפשר גם להוסיף משתנים לניתוח הסטטיסטי בלי לתפעל אותם אקטיבית, שליטה ב covariant.

בניסוי בין נבדקי (ניסוי בו בכל שילוב של רמות משתנים בלתי תלויים יש נבדקים שונים), תוספת משתנים בלתי תלויים תשפיע על תוקף המסקנה הסטטיסטית - תוספת משתנים בלתי תלויים בין נבדקים תגרום למצב של פחות נבדקים בכל שילוב תנאים בניסוי, מה שיחליש את העוצמה הסטטיסטית (בפרט של האינטראקציות) ויפגע בתוקף המסקנה הסטטיסטית.

חולשות של הפתרון הזה:

- אי אפשר לחשוב על כל המשתנים החיצוניים האפשריים
- צריך למדוד את כל המשתנים החיצוניים האפשריים - תוספת משתנים מצריכה מדידה שלהם, המחקר יהפוך מסובך וצריך לבדוק שהם לא משפיעים על התהליך.
- סכנת ממצאים מקריים (אינטראקציות, אפקטים פשוטים) - ניפוח אלפא, טעות מסוג אחד... חשוב לעשות רפליקציה כדי לראות שהתוצאה לא יצאה מובהקת באופן מקרי.

3. שמירת המשתנים החיצוניים קבועים - לדוגמה להריץ ניסוי על ימנים בלבד כדי להימנע מהרעש שנגרם מהשונות של שמאלים.

חולשות של הפתרון:

- יש גבול כמה משתנים חיצוניים אפשר להשאיר קבועים - חשוב להחליט מראש מי המשתנים החשובים ביותר.
- פגיעה בתוקף החיצוני - מאפשר להסיק רק לגבי אנשים ימנים.

הקטנת השפעה מקרית

שונות מקרית = שונות הטעות, נובעת ממשתנים חיצוניים לא ידועים.

איך אפשר לשלוט על גורמים לא ידועים:

1. מניפולציה יציבה - לדאוג שכל הנבדקים בניסוי יחוו חוויה דומה ככל האפשר, להתאמץ להגדיל את הדמיון בין התנאים השונים של הנבדקים השונים. במקרה זה אנחנו לא רוצים לחשוב על כל המשתנים החיצוניים שיכלו להשפיע כדי למנוע אותם (האור בחדר, השעה, סוג המחשב, מין הנסיין וכו'..). אלא פשוט לנסות מראש להשוות את התנאים
2. שיפור מהימנות המדידה - לבחור מדדים מהימנים ככל האפשר, יקטין את שונות המקרית.
3. מערך תוך נבדקי (repeated measures) - זה מקרה קיצוני של התאמה. בהתאמה לוקחים נבדקים דומים במשתנים חיצוניים ומקצים אותם באופן מקרי לקבוצות השינוי, במערך תוך נבדקי לוקחים את אותו נבדק לכל תנאי הניסוי השונים. במצב זה אנחנו לא חשופים לאיומים הנובעים מהשונות בין הנבדקים, מקטין לאפס את התרומה של ההבדלים הבין אישיים.

חולשות של מערך תוך נבדקי (דיברנו עליהן):

- איום אפקט סדר
- איום carry over
- איום מדידה תגובתית (אינטראקציה מדידה-מניפולציה)

ה-סוף.

