שיטות מחקר

פרק 4 - איכות כלי מדידה

תוכן העניינים

1. איכות כלי מדידה 1

איכות כלי מדידה:

הקדמה:

רקע:

כלי מדידה – הביטוי הכמותי של המשתנה התיאורטי. מאפשר למדוד כמותית את המשתנה התיאורטי.
איכות כלי המדידה – תלויה במידת ההלימה בין המשתנה התיאורטי, אותו הכלי נועד למדוד, למשתנה התצפיתי שהוא התוצאה המתקבלת מכלי המדידה.

איכות כלי המדידה מוערכת על ידי מדדים של: מהימנות ותוקף.

כלי מדידה הוא תמיד שילוב של שיטה אחת + תכונה אחת.

שיטה: הכוונה לאופן איסוף הנתונים (שאלון, ראיון, מבחן, תצפית).
תכונה: המשתנה התיאורטי שרוצים לבדוק (למשל רמת תוקפנות).

על מנת לבדוק את איכות כלי המדידה יש לבדוק את המהימנות והתוקף של הכלי:
מהימנות כלי המדידה – נשאלת השאלה עד כמה כלי המדידה שיטתי, עקבי.
במובן הזה שאם נחזור על המדידה פעם נוספת נקבל את אותן התוצאות.
תוקף כלי המדידה – נשאלת השאלה האם הכלי אכן בודק את מה שהוא אמור לבדוק, כלומר האם המשתנה התצפיתי אכן בודק את המשתנה התיאורטי ואך ורק משתנה זה.

מהימנות:

רקע:

עד כמה ניתן לסמוך על כלי המדידה שהוא מודד ערכו של המשתנה בדייקנות ועד כמה מושפע מטעויות מקריות.

טעות מדידה:

טעות שיטתית – הכלי מזייף בגודל קבוע ואחיד לאורך כל הנבדקים וכל המדידות: $\sigma_e^2 = 0$.

טעות מקרית – גודלה משתנה ממדידה למדידה ומנבדק לנבדק: $\sigma_e^2 \neq 0$.

מהימנות הכלי בודקת את העקביות של הכלי ולכן נפגעת רק מטעות מקרית (ולא נפגעת מטעות קבועה ששונותה = 0).

מקדם המהימנות - מבטא יחס בין השונות האמיתית לשונות הנצפית:

$$r_{tt} = \frac{\sigma_t^2}{\sigma_o^2} = r_{to}^2$$

המדידה מהימנה יותר ככל שהשונות בין האנשים מוסברת יותר על ידי הבדלים אמיתיים ופחות על ידי טעויות מקריות.

$$0 \leq r_{tt} \leq 1$$

מהימנות מושלמת	אין מהימנות
אין גורם טעות כלל	הכלי לא מבטא כלל את הערך האמיתי
שונות נצפית = שונות אמיתית	שונות נצפית = שונות הטעויות

אם למשל $r_{tt} = 0.55$ המשמעות היא כי המדידה משקפת 55% שונות אמיתית ו-45% שונות שמקורה בטעויות מקריות.

כלי נחשב מהימן אם $r_{tt} \geq 0.7$, תלוי באופן שבו המהימנות נמדדה.

שורש המהימנות ($\sqrt{r_{tt}} = r_{to}$) מהווה חסם עליון לכל מתאם בין כלי המדידה לכל כלי מדידה אחר: $r_{tt} \geq r_{xy}$.

אם מהימנותו של הכלי נמוכה זה אומר שהוא לא מתואם עם עצמו אז קל וחומר שלא יוכל להיות מתואם עם כלים אחרים.

המשמעות:

המהימנות משפיעה על יכולת הניבוי של הכלי. ככל שהמהימנות נמוכה כך גם יכולת הניבוי של הכלי תהיה נמוכה ולהיפך.

הגורמים לטעות המקרית של המדידה:

1) זמן המדידה – גורמים נסיבתיים העלולים להשפיע באופן שונה על כל נבדק:
   א. גורמים אקראיים – חיצוניים – קשורים במזל (כגון: מזג האוויר, תנאים פיזיים של המקום, טמפ' החדר, מידת הנוחיות, כמות הרעש, מצב התאורה וכו').
   ב. גורמים אישיים זמניים – קשורים בנבדקים עצמם (כגון: מצב הרוח, רמת ריכוז, מוטיבציה, עייפות, מצב בריאות וכו').
2) תוכן הפריטים – ייצוגיות הדגימה של עולם התוכן – האם יש דגימה טובה של פריטי המבחן כך שייצגו בצורה מהימנה את עולם התוכן של המבחן? האם ניסוחם ברור ומובן באותו אופן לכל הנבדקים?
3) סובייקטיביות הבודק – רלוונטיים למבחנים בהם הציון איננו אובייקטיבי אלא קשור לבודק מסוים. עד כמה הציון חשוף לשיקול דעתו ופרשנותו הייחודית של כל מעריך?

דרכים לבדיקת מהימנות הכלי:

בדיקות שונות הבאות לבדוק את רגישות הכלי למקורות הטעות המקרית.

1. מהימנות כיציבות.
2. מהימנות כאקוויוולנטיות.
3. מהימנות כעקיבות פנימית.
4. מהימנות בין שופטים.

מהימנות כיציבות:

מכונה מהימנות מבחן חוזר (Test-Retest): מקדם המהימנות מבטא את המתאם בין שתי העברות של אותו מבחן בהפרש של זמן: $r_{tt} = r_{t_1 t_2}$.
ההנחה היא שהמבחן מהימן כאשר תוצאותיו יציבות לאורך זמן.
בשיטה זו נבדקת רגישות המבחן לגורמים הקשורים לזמן המדידה – גורמים נסיבתיים – חיצוניים או פנימיים. כדי שהמבחן יחשב מהימן בשיטה זו: $t_{tt} \geq 0.8$.
ככל שהפער בזמן גדל, כך המהימנות שתתקבל תהיה נמוכה יותר.

מהימנות כאקוויוולנטיות:

מכונה מהימנות נוסחים מקבילים (Parallel Forms): מקדם המהימנות מבטא את המתאם בין שני נוסחים מקבילים של אותו המבחן: $r_{tt} = r_{AB}$.
ההנחה היא שהמבחן מהימן אם פריטי התוכן שבו מיוצגים בצורה טובה. מבחן זה רגיש לטעויות מדידה מקריות שנבעו מחוסר ייצוגיות הדגימה (תוכן, אופי וניסוח הפריטים).

קיימים שני סוגים של מהימנות כאקוויוולנטיות:

1. מבחן מקביל בו זמנית – שני הנוסחים מועברים באותו הזמן.
   נצפה ל: $r_{tt} \geq 0.8$.
2. מבחן מקביל לאחר זמן – שני הנוסחים מועברים בהפרש של זמן.
   נצפה ל: $r_{tt} \geq 0.7$.

המדד השני רגיש פרט לייצוגיות הדגימה גם לגורמים נסיבתיים - אקראיים שעלולים היו להשפיע על תוצאותיו (בגלל הפרש הזמנים).
אם נקבל מהימנות גבוהה נדע כי המבחן מתגבר על שתי הבעיות אולם במידה והמהימנות נמוכה לא ניתן לדעת האם הבעיה היא בייצוגיות הדגימה או בגורמים הנסיבתיים.

מהימנות כעקיבות פנימית:

קיימים שני סוגים של מהימנות כעקיבות פנימית:

1. מבחן חצוי.
2. קרונבך אלפא / קודר-ריצ'רדסון.

מבחן חצוי:

בדיקת מהימנות המבוססת על מתאם בין שני חלקים של המבחן. שלבי עבודה:

- חלוקה מקרית של המבחן לשני חלקים.
- חישוב מתאם בין שני החלקים.
- הצבת הערך שהתקבל בנוסחת ספירמן-בראון: $r_{tt} = \frac{2r_{1,2}}{1+r_{1,2}}$.

מהימנות המבחן כולו תהיה גדולה יותר מהמתאם בין שני חלקי המבחן (הנוסחה מבצעת תיקון כלפי מעלה של המתאם בין שני החלקים).

אלפא קרונבך:
מתבסס על ממוצע כל המתאמים הזוגיים האפשריים בין הפריטים (מדד קודר ריצ'רדסון מחושב עבור משתנים בסולמות דיכוטומיים). מדד זה מבטא את ממוצע המתאמים בין כל שני פריטים בשאלון. ככל שהמבחן ארוך יותר כך יהיה מהימן יותר כי דוגמים שוב ושוב את אותו המשתנה.
מדד זה בודק פרט לייצוגיות הדגימה גם את מידת ההומוגניות של המבחן.

מהימנות בין שופטים:
מדד מהימנות זה רלוונטי רק במדידות בהן ערכי המשתנה התצפיתי נקבעות באופן סובייקטיבי על ידי שופט. מקדם המהימנות מבוסס על המתאם או ממוצע המתאמים בין הערכות שופטי בלתי תלויים לביצוע מסוים: $r_{tt} = r_{ab}$.
הערך המצופה: $r_{tt} \geq 0.7$. הטעות נובעת מסובייקטיביות המעריך הכוללת כל דבר שקשור לבודק כמו מצב רוחו, עיוותי תפיסה, אפקט הילה, סטריאוטיפים וכו'.
2 השופטים חייבים להיות בעלי אותה הזיקה לקבוצת הנבדקים.

סוגי מבחנים ומדדי המהימנות המתאימים להם:

- מבחן הומוגני – פריטי המבחן מודדים גורם יחיד – ניתן להשתמש בכל סוגי המהימנויות.
- מבחן הטרוגני – פריטי המבחן מודדים היבטים שונים של תופעה מסוימת – לא ניתן להשתמש במדד אלפא של קרונבך.
- מבחן עוצמה – רמת קושי עולה – מדד קרונבך אלפא לא מתאים כי מצפים למתאמים נמוכים בין שאלות מרמות קושי שונות.
- מבחן מהירות – הגבלת זמן – לא מתאים עקיבות פנימית כי יש מספר שונה של פריטים שאנשים הצליחו לענות עליהם.
- מבחנים אובייקטיביים – בעלי מפתח ברור למתן ציונים שאינו תלוי בשיפוט סובייקטיבי של המעריך – מדד מהימנות בין שופטים איננו רלוונטי.
- סוג התכונה – אם התכונה אינה יציבה לא ניתן להשתמש במהימנות כיציבות ובמהימנות כאקוויוולנטיות מדד ב.

<u>הערה חשובה</u>: על מנת שנזהה שמדובר בבדיקה של מהימנות מסוג כלשהו, יש לזהות שמחושב מתאם בין שתי מדידות שמתבצעות על אותו כלי מדידה, כלומר שחייבים להיות שלושה מרכיבים בכל בדיקת מהימנות:

1. מדובר בקב' נבדקים אחת.
2. מדובר בבדיקת תכונה אחת.
3. באמצעות שיטה אחת.

אם אחד משלושת המרכיבים לא מתקיים, כבר לא מדובר בבדיקת מהימנות.

סיכום – מדדי המהימנות:

סוג מהימנות	מס' העברות	מה משתנה	גורם הטעות נובע מ-	המדד הסטטיסטי	מהימנות מצופה	חסרון
יציבות – מבחן חוזר	2	זמן ההעברה	גורמים נסיבתיים	מתאם בין שתי העברות של מבחן זהה בהפרשי זמן	$r_{tt} \geq 0.8$	1.זיכרון ולמידה. 2.לא מתאים לתכונות לא יציבות.
אקוויוולנטיות – מבחן מקביל בו זמנית	1	תוכן השאלות	חוסר ייצוגיות הדגימה	מתאם בין העברה של שני מבחנים מקבילים	$r_{tt} \geq 0.8$	קושי לבנות שני נוסחים באותה רמת קושי.
מבחן מקביל לאחר זמן	2	זמן העברה + תוכן השאלות	חוסר ייצוגיות הדגימה + גורמים נסיבתיים		$r_{tt} \geq 0.7$	במדד ב' – לא ניתן לדעת מהיכן הטעות.
עקיבות פנימית – α קרונבך וקודר ריצ'רדסון	1	תוכן השאלות	חוסר ייצוגיות הדגימה + הטרוגניות	נוסחת α קרונבך	$r_{tt} \geq 0.85$	לא מתאים למבחנים הטרוגניים.
מבחן חצוי			חוסר ייצוגיות הדגימה בלבד	חישוב מתאם בין שני חצאי מבחן והצבה בנוסחת ספירמן בראון.	$r_{tt} \geq 0.8$	מדויק פחות מ-α קרונבאך.
בין שופטים	1	המעריך	סובייקטיבית המעריך	מתאם בין הערכות השופטים.	$r_{tt} \geq 0.7$	1.השקעה בקריטריונים 2.אימון של בודקים.

תרגול:

1) חוקר בנה שלושה שאלונים דומים (X, Y, Z) לבדיקת בטחון עצמי בעבודה טיפולית. כל אחד מהשאלונים כלל 40 פריטים. שלושת השאלונים הועברו לקבוצה בת 100 נבדקים, גברים ונשים, באפריל 85 (מועד א׳) ושוב במאי 85 (מועד ב׳).
עבור כל אחד מהמצבים המתוארים להלן, ציינו והסבירו מהי הדרך שנבחרה למדידת המהימנות ומהו האחוז בשונות הציונים הנצפים שניתן לייחס למקור הטעות הרלוונטי:

א. התקבל מתאם של 0.9 בין X ל Y שהועברו במועד הראשון.

ב. התקבל מתאם של 0.86 בין X של המועד הראשון לבין Y של המועד השני.

ג. התקבל מתאם של 0.95 בין Y שהועבר במועד הראשון לבין Y שהועבר במועד השני.

ד. שאלון X של מועד א חולק לשניים. נתקבל מתאם של 0.85 בין שני החלקים.

ה. ממוצע המתאמים של כל החלוקות האפשריות של פריטי שאלון Z היא 0.89.

2) יש לסמן נכון או לא נכון לגבי כל אחת מהטענות הבאות ולנמק:

א. כאשר מתקבל מקדם אלפא נמוך ניתן לדעת בוודאות שהמבחן אינו מהימן.

ב. נבדקה מהימנותו של מבחן חדש למדידת אסרטיביות באמצעות שיטת הנוסחים המקבילים בזמנים שונים. נמצא כי מקדם המתאם בין שני הנוסחים הינו 0.87 מכאן משמעות הדבר היא ש-13% משונות הנצפית ניתן לייחס לטעויות שמקורן בגורמים נסיבתיים בלבד.

ג. נמצא כי ממוצע inter-item correlation שהתקבל לשאלון דימוי עצמי הינו 0.82, מכאן ניתן לומר שהשאלון מהימן והומוגני.

3) חוקרת בנתה מבחן לבדיקת כושר מכני. המבחן כלל שאלות שונות כגון: ״בנסיעה ברכב בזמן בלימה לאן יטולטל קודם גוף הנוסעים, קדימה או אחורה?״. השאלון כלל 25 פריטים, לכל פריט היו שתי תשובות אפשריות. אחת נכונה ואחת לא נכונה. המבחן הועבר ל 50 נקידים. מדד קודר ריצ׳רדסון שהתקבל הינו 0.92.
על מנת לבחון את מהימנות מבחן 1 בנתה החוקרת נוסח דומה (להלן מבחן 2) והעבירה אותו לאותם אנשים לאחר שבוע. היא בחנה את הקשר בין הציונים שנתקבלו ממבחן 1 לבין הציונים שנתקבלו ממבחן 2.

א. איזו בדיקת מהימנות ערכה החוקרת?

ב. המתאם שהתקבל היה 0.7. האם החוקרת צריכה להיות מופתעת שלא קיבלה ערך זהה למדד קודר ריצ׳רדסון?

ג. עוזר המחקר הציע לחוקרת להעביר את מבחן 1 לקבוצה אחרת במועד אחר, בנוסף להעברה הראשונה של מבחן 1. הוא בדק מהימנות מסוג:
   i. יציבות.
   ii. אקוויוולנטיות.
   iii. שילוב של יציבות ועקיבות.
   iv. לא ניתן לדעת.
   v. זו לא בדיקת מהימנות.

4) נבחנה מהימנות מבחן משכל חדשני. לשם כך נאספו נתוני הביצוע של נבדקים מחוננים במבחן בנוסח יחיד של המבחן, לצורך חישוב מהימנות המבחן החצוי. מהימנות המבחן נמצאה נמוכה. הסבר אפשרי לכך הינו:
   א. מדובר במבחן הטרוגני מבחינת תחומי התוכן הנחקרים בו.
   ב. מדובר במבחן הומוגני מבחינת תחומי התוכן הנחקרים בו.
   ג. מדובר במדגם נבדקים הומוגני מבחינת התכונה הנחקרת.
   ד. תשובות ב' ו-ג' נכונות.

5) שני נוסחים מקבילים של מבחן לבדיקת מנת משכל הועברו למדגם של 100 נבדקים. לאחר שבוע הועבר הנוסח השני של המבחן שנית. להלן תוצאות חישובי המהימנות:
התקבל מתאם של 0.90 בין שני הנוסחים שהועברו באותו הזמן.
התקבל מתאם של 0.85 בין נוסח א' לבין נוסח ב' לאחר שבוע.
התקבל ערך קרונבך של אלפא 0.88 עבור נוסח א'.
מה אחוז השונות בציונים נצפים שאפשר לייחס לטעויות שמקורן:
   א. בייצוגיות דגימת התוכן בלבד.
   ב. בגורמים נסיבתיים/אקראיים בלבד.
   ג. בייצוגיות דגימת התוכן ובהטרוגניות המבחן.

6) חוקר בנה מבחן לרצייה חברתית והעביר אותו פעמיים בהפרש של חודשיים ולאחר מכן חישב את המתאם בין שתי ההעברות.
   א. איזה סוג מהימנות נבדק?
   ב. מה בודקים בסוג מהימנות זה?
   ג. המתאם שהתקבל היה 0.74, מה ניתן לומר על מהימנות המבחן?
   ד. החוקר ביקש לבדוק האם המבחן מושפע מבחירת הפריטים, ניסוחם וסוג הפריטים. באילו סוגי מהימנות הוא יכול להשתמש?
   ה. במה עדיף להשתמש?
   ו. לתכונת חרדה יש מספר ממדים, איזה סוג מהימנות יחשב החוקר?
   ז. אם בכל זאת יחושב סוג המהימנות השני, מה צפוי להיות ערכו?
   ח. מהימנות מבחן חצוי=0.82. החוקר טען שהתגבר על גורמי הטעות. האם צדק?

7) כשמשקל א' מראה שהנשקל שוקל 60 ק"ג ומשקל ב' מראה שהוא שוקל 58 ק"ג אפשר לומר:

א. מהימנותו של משקל א נמוכה.

ב. מהימנותו של משקל ב נמוכה.

ג. מהימנותם של שני המשקלים נמוכה.

ד. מהימנותם של שני המשקלים גבוהה.

ה. לא ניתן להסיק מהבדיקה שבוצעה על מהימנות המשקלים.

8) חוקר בנה 3 שאלונים שבודקים דימוי עצמי:
C=היבט חברתי, S=היבט רגשי, G=היבט גופני.
כל אחד מהשאלונים כלל 40 פריטים שהועברו לקבוצה בת 100 נבדקים, גברים ונשים באפריל 2015 (מועד 1) ושוב במאי 2015 (מועד 2).
עבור כל אחד מהמצבים המתוארים להלן, ציין והסבר מהי הדרך שנבחרה לבדיקת המהימנות. אם מדובר בעקיבות פנימית ציין מהי הטכניקה ומהו המקדם שחושב.
ציין מהו חלקה של השונות האמיתית וממה היא משוחררת.

א. $r_{G_1,G_2}=0.8$.

ב. שאלון S של מועד 1 חולק לשניים. נתקבל מתאם של 0.81 בין שני החלקים.

ג. ממוצע המתאמים של כל החלוקות האפשריות של פריטי שאלון S היה 0.89.

ד. עוזר מחקר רצה להעלות את הערך שהתקבל בסעיף הקודם. לצורך זה הוא הוסיף קבוע חיובי לכל ציוני הנבדקים בשאלון S. האם הוספת קבוע זה תשפיע על ערך מקדם המהימנות? הסבר.

ה. מבחן C שהועבר במועד 2 ניתן לקידוד של שני מעריכים בלתי תלויים. המתאם שחושב בין שני המעריכים היה שווה ל-1.

תשובות סופיות:

1) א. אקוויוולנטיות, 10%. ב. אקוויוולנטיות, 14%. ג. מבחן חוזר, 5%.
ד. עקיבות פנימית, 15%. ה. עקיבות פנימית, 11%.

2) א. לא נכון. ב. לא נכון. ג. נכון.

3) א. אקוויוולנטית. ב. לא. ג. v.

4) ג'.

5) א. 10%. ב. 5%. ג. 12%.

6) א. יציבות, מבחן חוזר. ב. רגישות המבחן לטעויות שמקורן בגורמים נסיבתיים.
ג. נמוכה. ד. אקוויוולנטיות או עקיבות פנימית. ה. עקיבות פנימית.
ו. עקיבות פנימית. ז. נמוך. ח. לא.

7) ה.

8) א. יציבות, 80%. ב. עקיבות פנימית-מבחן חצוי, 90%.
ג. עקיבות פנימית, 89%. ד. לא. ה. בין שופטים.

תוקף:

רקע:

תוקף מתאר את המידה שבה כלי המדידה מודד את התכונה שהוא אמור למדוד ולא תכונות אחרות.
מאחר ולא נתון בידינו המשתנה התיאורטי, אנו נבדוק את מידת ההתאמה של הכלי שלנו למשתנה התיאורטי בדרכים עקיפות.

לתוקף שלוש הגדרות אופרטיביות:
תוקף תלוי קריטריון (תוקף ניבוי) – המידה שתוצאות המבחן מאפשרות חיזוי מדויק של מדידה אחרת (קריטריון).
תוקף תוכן – המידה בה פריטי המבחן בודקים את עולם התוכן של המשתנה.
תוקף מבנה – מידת ההתאמה בין המשתנה הנמדד (התצפיתי) למשתנה התיאורטי (התכונה).

תוקף תלוי קריטריון:

מתייחס למצב בו החוקר מעוניין לפתח כלי שניתן יהיה לנבא באמצעותו משתנה מסוים הקרוי קריטריון.
למשל, הפסיכומטרי הוא כלי שנועד לנבא הצלחה בלימודים -הקריטריון.
הכלי (מבחן פסיכומטרי) יחשב בעל תוקף תלוי קריטריון גבוה אם מידת ההתאמה בינו לבין הקריטריון (ציונים בתואר הראשון) תהיה גבוהה.

בדיקת תוקף תלוי קריטריון:

בדיקת התוקף תלויה באופי המשתנים (איכותיים או כמותיים):

1. חישוב מתאם בין מנבא לקריטריון - r_{xy} - מקדם תוקף תלוי קריטריון (כלי וקריטריון כמותיים).
2. טבלת ציפיות (קריטריון איכותי).
3. טבלת שכיחויות (כלי וקריטריון איכותיים).

שיטות תיקוף תלויות קריטריון:

ישנן 3 שיטות תיקוף תלויות קריטריון. שיטות אלו נבדלות בזמני מדידת הציון בכלי (מנבא) ביחס לציון בקריטריון (מנובא):

1. תוקף לאחר מעשה – קריטריון נמדד לפני הכלי.
2. תוקף בו זמנית – הקריטריון והכלי נמדדים בו זמנית.
3. תוקף הניבוי – הקריטריון נמדד אחרי הכלי.

תרגול:

1) מהו סוג תוקף תלוי קריטריון במקרים הבאים:

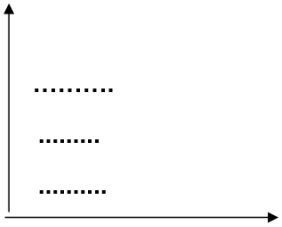
א. תלמידים לקראת קבלה בבית ספר למשפטים נדרשים לעבור מבחן באנגלית. חושב מתאם בין ציוני המבחן לבין ממוצע הציוני שנה א' של סטודנטים שהתקבלו באותה שנה ללימודים.

ב. קבוצה של חולי לב וקבוצה של אנשים בריאים התבקשו למלא שאלון לבדיקת סיכון ללקות במחלת לב. חושב הקשר בין ציוני השאלון לבין שתי הקבוצות.

ג. עובדים בחברת הייטק התבקשו לענות על שאלון לבדיקת יכולת מנהיגות והממונים עליהם נתבקשו למלא שאלוני הערכה על ביצועי העובדים. מתאם חושב בין ציוני השאלונים.

ד. חוקר בנה מבחן שמטרתו לאתר את המתאימים להכשרה לתפקיד דיפלומטי במשרד החוץ. קבוצה של 90 מועמדים נבחנה במבחן. כל המועמדים התקבלו לקורס בלי קשר להישגיהם במבחן. בתום הקורס נבחן הקשר בין תוצאות במבחן (מתאים/לא מתאים) לבין ציון סוף הקורס (עבר/נכשל).

2) הקשר בין ציוני מבחן להכשרה לתפקיד דיפלומטי לבין ציון סוף קורס מוצג בתרשים:



האם יש למבחן תוקף ניבוי?

תשובות סופיות:

1) א. בו זמני. ב. לאחר מעשה. ג. בו זמני. ד. ניבוי.

2) לא.

רקע:

תוקף תוכן:

מבטא את המידה שבה פריטי המבחן בודקים את עולם התוכן של המשתנה הנמדד. האם פריטי המבחן רלוונטיים לעולם התוכן הנמדד והאם מייצגים את כל היבטיו? כדי להבטיח תוקף תוכן אנחנו צריכים להקפיד על ניתוח טוב של עולם התוכן בו אנו מעוניינים, צריכים לבחור אוסף מייצג של פריטים מעולם התוכן הזה ובנייה טובה (סדר השאלות, למשל) של מכשיר המדידה (השאלון / מבחן).

תהליך התיקוף של תוקף תוכן:
תהליך התיקוף הוא התרשמותי-שיפוטי ומתרחש בשלב הבנייה של הכלי ואמורים להשתתף בו מומחים ויועצים בתחום.
תהליך התיקוף יכלול את השלבים הבאים:

1. ניתוח שיטתי של עולם התוכן בו מעוניינים.
2. בניית אוסף מייצג של הפריטים.
3. שיפוץ הפריטים בתהליך דיון והסכמה.

תוקף מבנה:

מתייחס למידה שבה המבחן מודד את התכונה / המבנה התיאורטי. כלומר, למידת ההתאמה בין המשתנה התצפיתי למשתנה התיאורטי.

למשל:

משתנה תיאורטי: רמת תוקפנות

↓

הגדרה נומינאלית: התנהגות אלימה של אדם

↓

הגדרה אופרציונלית: מדידת כמות המילים השליליות בשיחה

↓

משתנה תצפיתי: מספר המילים השליליות בשיחה.

תהליך התיקוף של תוקף מבנה:
תהליך ארוך יותר מהקודמים וכולל את כל סוגי הבדיקות של התוקף והמהימנות שהוזכרו.

הדרכים לבחינת תוקף מבנה:

1. קורלציות עם משתנים קשורים: בדיקת מתאמים של אותו מושג עם מושגים דומים.
2. הבדלים בין קבוצות: החוקר יבחן את המידה שבה מתגלים הבדלים בין קבוצות שההבדל הבולט ביניהן הוא בתכונה הנמדדת ולהשוות את הציונים שלהן במכשיר המדידה.
3. תוקף הניבוי כחלק מתוקף מבנה.
4. בדיקת מהימנות כחלק מתוקף מבנה.

דרכים אלו לתוקף מבנה מתבססים על שיטתם של **קרונבך ומיל (דרכים לפיתוח תיאוריה).**

5. תוקף מתכנס: נמדוד את אותו המושג בדרכים שונות (שיטות שונות). אם המתאמים המתקבלים הם דומים, נסיק על תוקף מבנה.
6. תוקף מבחין: בדיקת מתאמים עם מושגים שאינם קשורים.
אם המתאמים נמוכים, נסיק שיש אפשרות שמדובר במושג חדש.

שתי הדרכים הללו לתוקף מבנה מתבססים על שיטתם של **קמבל ופיסק (מטריצת תכונות שיטות).**

מטריצת MMM (מטריצת שיטות תכונות):

כלי מדידה = תכונה + שיטה.
הנתונים במטריצה מתייחסים למקדמי המהימנות והתוקף (מבחין ומתכנס) של שני כלים שונים:

		שיטה 1 (שאלון)		שיטה 2 (ריאיון)	
		תכונה 1 (חרדה)	**תכונה 2 (פחד)**	**תכונה 1 (חרדה)**	**תכונה 2 (פחד)**
שיטה1 (שאלון)	**תכונה 1 (חרדה)**	מהימנות			
	תכונה 2 (פחד)	מבחין א	מהימנות		
שיטה 2 (ריאיון)	**תכונה 1 (חרדה)**	מתכנס	מבחין ב	מהימנות	
	תכונה 2 (פחד)	מבחין ב	מתכנס	מבחין א	מהימנות

כללים:

1. ערכי המהימנות – אותה תכונה הנבדקת באותה שיטה – צריכים להיות הערכים הגבוהים בטבלה.
2. ערכי התוקף המתכנס – אותה תכונה הנבדקת בשיטות שונות – צריכים להיות גבוהים.
3. ערכי תוקף מבחין א' – תכונות שונות הנבדקות באותה השיטה – נמוכים מהמתכנס.
4. ערכי תוקף מבחין ב' – תכונות שונות הנבדקות בשיטות שונות – הנמוכים ביותר בטבלה.

ניתוח המטריצה צריך להתייחס בראש ובראשונה למהימנות. אם שיטה איננה מהימנה אז כל מדדי התוקף שייבחנו עם אותה השיטה יהיו מוטלים בספק.

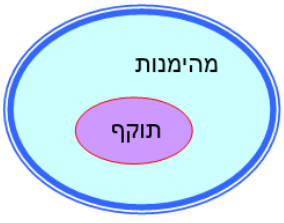
דרך להבחין בין בדיקת תוקף לבדיקת מהימנות:
מהימנות: אותה שיטה ואותה תכונה.
תוקף: שיטות שונות ו/או תכונות שונות.

סיכום תוקף – שיטות ותכונות:

סוג התוקף	תכונות	שיטות	קבוצות	
תוקף תלוי קריטריון	1	2	1	מנבא וקריטריון
חקר הבדלים בין קבוצות	1	2	2	מכשיר המדידה והשתייכות קבוצתית
חקר מתאמים	2 שונות	1 או 2	1	
תוקף מתכנס	1	2	1	
תוקף מבחין	2 דומות	1 או 2	1	

קשר בין מהימנות לתוקף:
אם במהימנות אנו בודקים את רגישות הכלי לטעויות מקריות, בתוקף אנו בודקים את רגישות הכלי לטעויות קבועות ושיטתיות.
לדוגמא: משקל שמראה תמיד קילו נוסף – מהימן אך לא תקף.

1. **מהימנות היא תנאי הכרחי אך לא מספיק לתוקף.**
כדי שמבחן יהיה תקף הוא חייב להיות מהימן אולם מבחן מהימן הוא לא בהכרח תקף.



2. **מהימנות מהווה חסם עליון לתוקף** – שורש המהימנות מהווה חסם עליון לקשר שבין כלי המדידה לכל כלי מדידה אחר. ערך התוקף יגיע לכל היותר לערך שורש המהימנות שחושבה.
3. **הקשר בין מהימנות לתוקף תוכן** – אם העקיבות הפנימית גבוהה יש אישוש לתוקף תוכן (אלא אם כן המבחן הטרוגני).
4. **הקשר בין מהימנות ותוקף מבנה** – התיאוריה מכתיבה את מידת העקיבות הפנימית ומידת היציבות של התכונה לאורך זמן.

תרגול:

מטריצת MMM:

1) לפניך טבלה המסכמת מתאמים בין ציונים שהתקבלו בארבעה כלי מדידה שונים, וכן נתונים לגבי מהימנותם:

		שאלון		ראיון	
		אסרטיביות	**מנהיגות**	**אסרטיביות**	**מנהיגות**
שאלון	**אסרטיביות**	0.85			
	מנהיגות	0.46	0.86		
ראיון	**אסרטיביות**	0.62	0.28	0.84	
	מנהיגות	0.28	0.40	0.65	0.89

התוקף המתכנס, התוקף המבחין החד-שיטתי והתוקף המבחין הרב-שיטתי של ראיון אסרטיביות הם (לפי סדר זה):

א. 0.62 ; 0.46 ; 0.28.
ב. 0.62 ; 0.65 ; 0.28.
ג. 0.85 ; 0.62 ; 0.21.
ד. 0.46 ; 0.62 ; 0.21.

2) פותח ראיון פסיכיאטרי חצי-מובנה לאבחון דיכאון. במחקר המשך, נבחנה מהימנותו, וכן קשרים בין ההערכות המתקבלות באמצעותו לבין כלי מדידה אחרים. הממצאים מסוכמים במטריצה שלפניך (הערכים בטבלה הם תוצאות חישוב המתאמים הרלוונטיים):

		דיכאון		סכיזופרניה	
		סולם MMPI	ראיון פסיכיאטרי	סולם MMPI	ראיון פסיכיאטרי
דיכאון	סולם MMPI	0.86			
	ראיון פסיכיאטרי	0.25	0.87		
סכיזופרניה	סולם MMPI	0.54	0.18	0.84	
	ראיון פסיכיאטרי	0.36	0.23	0.26	0.88

מה ניתן ללמוד מן הנתונים שהתקבלו במחקר (המפורטים בטבלה דלעיל) על תוקפו של כלי המדידה שפותח? התייחס לכלל הנתונים הרלוונטיים. נסה להציע הסבר לממצאים.

תוקף:

3) במטרה לברור מועמדים מתאימים לביה"ס לפקידות, הועבר למועמדים מבחן מטלות פקידותיות X1 שבוחן את מידת הדיוק בהעתקת מסמך. כל המועמדים התקבלו לביה"ס. לאחר סיום הלימודים ניתן לכל נבדק ציון סופי Y שביטא את מידת ההצלחה בלימודים. בדיקת המתאם בין X1 ל-Y היא בדיקת:

א. תוקף מתכנס של X1.
ב. תוקף ניבוי של X1.
ג. תוקף מבחין של X1.
ד. תוקף תוכן של Y.

4) כשפותחו מבחני האינטליגנציה, על מנת לתקפם, בחנו ילדים במבחן אינטליגנציה וביקשו גם מהמורה שתיתן לכל ילד ציון אינטליגנציה על פי היכרותה עימו. החוקרים חישבו מתאם בין שתי מערכות ציונים אלו. במקרה זה ניתן לומר כי:

א. נבדק תוקף התוכן של מבחן האינטליגנציה כחלק מתוקף המבנה של המבחן.

ב. כחלק מבדיקת תוקף המבנה של מבחן האינטליגנציה נבדק תוקף הניבוי של המבחן, הקריטריון: הערכת המורה.

ג. נבדק תוקף הניבוי של הערכת המורה. הקריטריון: ציון במבחן.

ד. נבדק התוקף המבחין של מבחן האינטליגנציה כחלק מבדיקת תוקף המבנה של המבחן.

ה. אף תשובה אינה נכונה.

5) חוקר בנה שאלון למדידת כשרון מוזיקלי. כדי לבדוק את תוקפו המתכנס נתן לכל נבדק ציון בכישרון מוזיקלי על ידי מומחה שהאזין לקטע שהנבדק ניגן. המהימנות שחישב החוקר לשאלון הייתה קרובה ל-0. מכאן ניתן להסיק:

א. התוקף המתכנס (מתאם בין השאלון לציון על הקטע המוזיקלי) יהיה גם הוא קרוב לאפס.

ב. התוקף המתכנס יהיה קרוב ל-1.

ג. על סמך הנתונים אי אפשר לדעת מה יהיה ערכו של התוקף המתכנס.

ד. מקדם התוקף המתכנס יכול להיות כל מספר.

6) פסיכולוג בנה שאלון למדידת גבריות, והעביר אותו בנפרד לקבוצת גברים ולקבוצת נשים, לצורך חישוב ההבדלים בין הקבוצות. השיטה בה השתמש הפסיכולוג לתיקוף השאלון מקבילה להשוואת הביצוע:

א. של גברים ושל נשים במבחן אינטליגנציה.

ב. של מדריכים ושל חניכים בקורס טייס, במבחן אינטליגנציה.

ג. של נשים ושל גברים במבחן טייס.

ד. של מדריכים ושל חניכים בקורס טייס, במבחן טייס.

7) אילו מבחן הנהיגה בארץ היה כולל שאלות במתמטיקה דיפרנציאלית ואינטגראלית, אילו מהטענות הבאות הייתה נכונה:

א. המבחן אינו אמין.

ב. המבחן אינו תקף.

ג. המבחן אינו מהימן.

ד. המבחן אינו מבחן הישג.

8) הקשר שבין מבחן הכניסה הפסיכומטרי לבין הישגי התלמידים בסיום תואר ראשון, מודדת את ה _____ של המבחן הפסיכומטרי :

א. מהימנות.

ב. תקיפות.

ג. מהימנות ותקיפות.

ד. אף אחד מהנ"ל.

9) מרפאה לבריאות הנפש סובלת מבעיה של חוסר בכוח אדם, ולכן היא מעוניינת בסלקציה בפונים לפי הסיכוי שלהם להצליח בטיפול. נמצא מתאם של 0.2 בין מבחן שבדק את מידת ההתאמה של נבדק לטיפול לבין נשירת הנבדקים מהטיפול. סמן/י את המשפט הנכון ביותר :

א. למבחן יש תוקף ניבוי.

ב. למבחן אין תוקף ניבוי.

ג. למבחן יש מהימנות גבוהה.

ד. תוקף הניבוי גבוה אך לא מהימנותו של המבחן.

ה. אף אחד מהנ"ל.

10) חוקר רצה להחליט האם להשתמש בשאלון רגישות (X1) ובשאלון יכולת אמפאטית (X2) לקבלת מועמדים לתואר שני בפסיכולוגיה קלינית. השאלונים הועברו לכל המועמדים. כולם התקבלו ללימודים. לאחר שנת הלימודים הראשונה נבחן כל מועמד וקיבל ציון על הישגיו בתום שנה זו (Y1). כמו כן בתום השנה העריך המורה את היכולת של כל מועמד להיות פסיכולוג קליני באמצעות ריאיון (Y2). הנח כי כל המדדים היו מהימנים.

שימו לב כי :

X1 – ככל שהציון בשאלון גבוה יותר הרגישות נמוכה יותר.

X2 – ככל שהציון בשאלון גבוה יותר היכולת האמפאטית גבוהה יותר.

Y1 – ככל שהציון גבוה יותר ההישגים גבוהים יותר.

Y2 – ככל שהציון גבוה יותר הערכת המועמד כמתאים להיות פסיכולוג גבוהה יותר.

להלן התוצאות שהתקבלו :

המתאם בין X1 ל-Y1 : 0.4-.

המתאם בין X1 ל-Y2 : 0.3.

המתאם בין X2 ל-Y1 : 0.03.

המתאם בין X2 ל-Y2 : 0.4.

א. לגבי כל מתאם ציין באיזו בדיקת תוקף מדובר.

ב. מה תוכלו לומר על ערכו של שאלון הרגישות לאור נתוני התוקף שלעיל? הסבירו.

ג. מה תוכלו לומר על ערכו של שאלון היכולת האמפאטית לאור נתוני התוקף שלעיל? הסבירו.

תשובות סופיות:

1) ב'.
2) מהימנות ותוקף מבחין טובים, תוקף מתכנס חלש.
3) ב'.
4) ב'.
5) א'.
6) ד'.
7) ב'.
8) ב'.
9) ב'.
10) ראו סרטון.

תרגול מסכם – תוקף ומהימנות:

תרגול:

1) חוקר בנה מבחן לבדיקת כושר מכני (להלן מבחן 1), הכולל 25 פריטים, לכל פריט שתי תשובות אפשריות, אחת נכונה והשנייה לא נכונה.

א. על מנת לבדוק את מהימנות המבחן בנה החוקר נוסח נוסף (להלן מבחן 2) והעביר אותו לאותם אנשים באותו הזמן. החוקר בחן את הקשר בין הציונים שהתקבלו במבחן 1 לבין הציונים שהתקבלו במבחן 2. איזו בדיקה ביצע החוקר ומה מטרתה? הסבר בקצרה.

ב. עוזר המבחן הציע להעביר את מבחן 1 (בנוסף להעברתו הראשונה) לקבוצה אחרת של נבדקים ובמועד אחר.
   i. ציינו איזו בדיקת מהימנות התכוון עוזר המחקר לבדוק?
   ii. האם אמנם ניתן לבדוק מהימנות זו בדרך שהוצעה? הסבירו בקצרה.

ג. החוקר העביר את מבחן 1 לשני סוגי נבדקים: מדגם של מכונאי רכב ומדגם של פקידים.
   i. איזה בדיקה ערך החוקר? בדיקת תוקף או מהימנות?
   ii. מאיזה סוג? הסבר באמצעות מונחי שיטות ותכונות.
   iii. איזה תוצאה תהיה רצויה לו ואיזה לא? הסבר בקצרה.
   iv. מהו סוג המשתנה הנבדק? הסבר בקצרה.

ד. 30 נבדקים שענו על מבחן 1, התקבלו לעבודה כמכונאים ללא התייחסות לציונם במבחן. החוקר בדק האם יש קשר בין הצלחתם במבחן להערכת עבודתם כמכונאים על ידי הממונה הישיר עליהם. להלן התוצאות (בתאים מספר האנשים):

הערכת הממונה	הישגים במבחן		סה"כ
	הצלחה	כישלון	
גרוע	5	5	10
טוב	10	10	20

האם למבחן יש תוקף ניבוי? (על מנת לענות על נשאלה אין צורך בחישוב מעבר לחישוב אחוזים).

2) כחלק ממיון מועמדים להוראה בסמינר למורים, הועבר ל-80 מועמדים מבחן יכולת עמידה מול כיתה. המועמדים נתבקשו לתת שיעור לכיתה למשך חצי שעה, כשבאותה עת צפו בהם שני מורים ותיקים (להלן צופה א' וצופה ב'). כל מורה ותיק נתן באופן בלתי תלוי הערכתו לגבי רמת המועמד על פני ממדים שונים, כמו גם ציון סופי של מידת ההתאמה להוראה על סולם שבין 1 (כלל לא מתאים להוראה) לבין 5 (מאוד מתאים להוראה).

א. החוקר בדק את המתאם בין שני המורים הוותיקים. עבור 85% מהנבדקים התקבלה הסכמה בין שני הבודקים.
   i. האם זו בדיקת תוקף או מהימנות?
   ii. מאיזה סוג? הסבירו בקצרה.
   iii. האם ניתן לומר ש-85% מהשונות האמיתית משוחררת מבעיות הקשורות בזמן? הסבירו.

ב. הנבדקים התבקשו לאחר כחודשיים לתת שיעור נוסף ששוב צפו בהם שני המורים הוותיקים (צופה א' וצופה ב') ונתנו את הערכתם לגבי מידת ההתאמה להוראה. עבור צופה א' נתקבל מתאם של 0.8 בין הערכתו את השיעור הראשון לבין הערכתו את השיעור השני.
   i. האם זוהי בדיקת תוקף או מהימנות?
   ii. מאיזה סוג? הסבירו במונחי תכונות שיטות.

ג. כל המועמדים התקבלו לסמינר בלא התחשבות בהערכת שני הצופים. החוקר בדק את הקשר בין הציון שהתקבל על ידי צופה ב' לבין ממוצע הציונים של המועמד שהתקבל בסיום הסמינר. התקבל מתאם של 0.34.
   i. איזו בדיקה ערך החוקר? בדיקת תוקף או מהימנות?
   ii. מאיזה סוג? הסבירו באמצעות שיטות ותכונות.

ד. החוקר המשיך לעקוב אחר תפקודם של המועמדים בעבודה כמורים. כל מועמד הוערך על תפקודו כמורה על ידי מנהל בית ספרו (בסולם שבין 1-גרוע מאוד ל-7-מצויין). נתקבל מתאם של 0.5 בין ממוצע הציונים של המועמד בסמינר לבין הערכת המנהל (הנח כי כל 80 המועמדים התקבלו לעבודה כמורים).
   i. האם זוהי בדיקת תוקף או מהימנות?
   ii. מאיזה סוג? הסבירו באמצעות שיטות תכונות.
   iii. באיזה סולם מדידה נתונה מידת ההערכה להתאמה להוראה?
   iv. האם ניתן לערוך טרנספורמציה על ציוני סולם זה של הוספת קבוע חיובי? הסבירו בקצרה.

3) חוקר בנה שאלון חדש שמטרתו הייתה לבדוק את אהבת קריאת הספרים של האדם. השאלון כלל 25 פריטים, כשסולם התשובות נע בין 1 ל-5 לכל פריט (1 - חוסר עניין מוחלט בקריאת ספרים ו-5 - אהבה רבה של קריאת ספרים). השאלון הועבר למדגם של 100 נבדקים.

א. בבדיקה שערך החוקר התקבלה α של קרונבך 0.8.
   i. איזה בדיקה ערך החוקר?
   ii. על סמך הבדיקה: מה אחוז השונות הנובע מטעויות?
   iii. מאיזה מקור לטעויות משוחררת בדיקה זו?

ב. מהו המקור האפשרי לשונות אמיתית בציוני הנבדקים בשאלון אהבת הקריאה?

ג. החוקר העביר את השאלון למדגם של נשים ולמדגם של גברים. לא נמצא הבדל בציוני שתי הקבוצות.
   i. איזו בדיקה ערך החוקר?
   ii. מה מעידה התוצאה על טיבו של השאלון?

ד. החוקר העביר את השאלון לאותם נבדקים לאחר חמישה חודשים. נמצא מתאם של 0.9 בין הציונים של ההעברה הראשונה לציונים של ההעברה השנייה.
   i. איזו בדיקה ערך החוקר? תוקף או מהימנות?
   ii. מאיזה סוג? הסבר במונחי תכונות שיטות.
   iii. האם ניתן לומר על סמך התוצאות ש-90% מהשונות היא שונות אמיתית המשוחררת בתלות בתוכן?

ה. בבדיקה נוספת נבדק המתאם בין ציון בשאלון הבודק את מס׳ הספרים שקרא הנבדק בשנה האחרונה לבין ציונו בשאלון אהבת הקריאה.
   i. איזו בדיקה ערך החוקר-תוקף או מהימנות?
   ii. מאיזה סוג. הסבר במונחי תכונות שיטות.

ו. בבדיקה נוספת הועבר השאלון אהבת הקריאה לקבוצת נבדקים הפוקדים מדי שנה את אירועי שבוע הספר לבין קבוצת נבדקים שמעולם לא ביקרה בשבוע הספר. לא נמצא הבדל מובהק בין שתי הקבוצות בציון בשאלון.
   i. איזו בדיקה בוצעה-תוקף או מהימנות?
   ii. מאיזה סוג? הסבר במונחי תכונות שיטות.
   iii. מה מעידה תוצאת הבדיקה על טיבו של השאלון?

ז. נמצא מתאם של 0.5 בין ציון במבחן הבוחן את הרמה המילולית של הנבדק לבין ציון בשאלון אהבת הקריאה.
   i. איזה בדיקה נערכה-תוקף או מהימנות?
   ii. מאיזה סוג? הסבר במונחי תכונות שיטות.
   iii. מה מעידה התוצאה על טיבו של השאלון?

תשובות סופיות:

1) א. אקוויולנטיות, מטרה: זיהוי רגישות הכלי לייצוגיות הדגימה של פרטי התוכן.
ב.i. מבחן חוזר. ii. לא. ג.i. תוקף. ii. מבנה.
iii. ראו סרטון. iv. מנה. ד. לא.

2) א.i. מהימנות. ii. בין שופטים. iii. לא. ב.i. מהימנות.
ii. מבחן חוזר. ג.i. תוקף. ii. תלוי קריטריון – ניבוי.
ד.i. תוקף. ii. תלוי קריטריון – ניבוי. iii. סדר. iv. כן.

3) א.i. עקיבות פנימית. ii. 20%. iii. ייצוגיות הדגימה של פרטי התוכן.
ב. ההבדל בין האנשים בתכונה הנמדדת. ג.i. תוקף מבנה. ii. תמיכה.
ד.i. מהימנות. ii. מבחן חוזר. iii. לא. ה.i. תוקף.
ii. מתכנס. ו.i תוקף. ii. מבנה. iii. לא תומכת.
ז.i. תוקף. ii. מבחין. iii. נמוך.