

Проблем ваљаности тестова у антропологији и социологији

МИХАЈЛО МИЈАНОВИЋ

Филозофски факултет
Никшић

Проблем ваљаности се поставља у везу са свим врстама тестирања, анкетирања и мјерења, посебно са тестирањем и мјерењем у антропологији и социологији. Ваљане селекције (одабира) било које врсте, затим незаобилазног процеса транзиције, па до повратне информације која је представљена у квантитативном или квалитативном изразу, нема без прецизних информација које се добивају на основу бројних тестирања и мјерења.

Тестирање, анкетирање и мјерење је присутно у свим наукама, посебно у бројној породици антрополошких наука, према томе саставни је дио емпириског истраживања у социологији.

Циљ науке је утврђивање истине или приближавање истини. На том тешком путу ка циљу користе се бројна средства и методе. С обзиром на то да су друштвене науке више упућене на бројна тестирања, анкетирања и мјерења у циљу објективизације извођењу одређених тврдњи и закључака, намеће се питање: да ли примјењени тестови, анкете, упитници и остали мјерни инструменти имају дозвољавајућу ваљаност? Свјedoци смо нелогичних закључака и тврдњи, који се често приписују непримјереним статистичким методама, а не промашеним (непримјереним) мјерним инструментима. Треба знати да примјерене статистичке методе неће изнедрити добре закључке ако су примјењени непримјерени мјерни инструменти (анкете, упитници и тестови). Ријеч је о ваги за тачно мјерење. За конструкцију ваге-теста потребно је посебно знање и умијеће. Сви они који се упуштају да конструишу неки тест,

анкету или упитник, морали би да имају велико искуство као истраживачи, да доста знају о социометрији имплицитно и статистици. Подразумијева се да су врсни познаваоци конкретне научне дисциплине. То није тако увијек, па се лакше одлучују за конструкцију тестова они који мало знају о елементарним правилима и методама како се конструишу тестови, анкете упитници и сл. О проблему ваљаности примјењених мјерних инструмената често се поклања мала пажња, а посљедице могу бити велике, посебно ако се на основу тестова сумњиве ваљаности врше селекције које често детерминишу судбину селекционисаних или неселекционисаних испитаника. Примјер бројних тестова приликом примања на посао, уписа у школу и факултет те за друге професионалне оријентације, опредјељења и усмјерења. На основу тестова у чију ваљаност нико не сумња или не смије да сумња, доносе се бројни планови, програми, не само у образовању, већ у економији, индустрији, саобраћају, урбанизму, здравству итд. Мало се пажње, и још мање одговорности посвећује томе што је неко дисквалификован зато што није постигао задовољавајуће резултате у тестовима који немају никакве везе са основном намјеном теста, или због тога што су промашене огромне инвестиције у неки пројекат који је на основу непримјерених тестова гарантовао профитабилност. Наиме ријеч је о тесту или тестовима који не мјере оно што треба да мјере. У природним наукама нормално питање је, да ли је вага исправна, те ако је некада била, да ли је у моменту употребе исправна, односно да ли има важећи атест. У друштвеним наукама, као да се не смије поставити питање о исправности ваге-теста, из простог разлога што вага-тест није никада ни била исправна. Тестови за разлику од вага су високо варијабилни тако да њихова ваљаност, упоредивост и употребљивост зависи од простора, времена, узорка, начина тестирања и низа других фактора. Све ово отежава ситуацију и додатно смањује ваљаност или је пак ставља под знаком питања. Неки тест високе ваљаности у одређеном времену и простору не подразумијева исту ваљаност у другом времену и простору. О проблему метриских ка-

рактеристика тестова понајвише су се бавили и дали велики допринос психолози, односно психометричари, тако да им емпиричари друштвених наука дугују велику захвалност.

Поставља се основно питање да ли тест има адекватну ваљаност тј. да ли тестом мјеримо оно што треба да мјеримо. Често није проблем примјенити мјерни инструмент, посебно ако је мјерни инструмент дефинисан физичким инструментима, гдје се бројеви користе за дефинисање појаве, нпр: масе, брзине, снаге, гравитације, отпора итд. У антрополошким наукама, као и у бројним другим наукама често се примјењују бесмислени тестови, гдје је корелација између теста и оног шта тест треба да мјери нула. Произилази да је тест сам себи сврха. Дакле ријеч је о недовољној стохастичкој вези између теста као предиктора и неког критерија.

Поставља се кључно питање, који је тест ваљан? Одговора постоји више, у најкраћем тест је ваљан ако мјери оно што треба да мјери, или тест је ваљан ако мјери истину. Ако се постави питање, да ли је тест ваљан? Одговор слиједи у следећем питању. У односу на шта? Да би неки тест или неки тестови били ваљани, неопходно је егзактно утврдити високу корелацију између дотичног теста и критерија.

Критериј у неким случајевима је прецизно дефинисан физичким димензијама и јединицама, као што су: секунде, килограми, метри, вати итд. У бројним примјерима критериј је изражен у доста недефинисаној димензији нпр: утврђивање социјалног статуса, или колико и да ли неко припада некој формалној или неформалној групи, затим колико је неко здрав или болестан, колико је неко интелигентан, аграсиван, депресиван, лијеп, ружан, подобан, неподобан и сл. У наведеним примјерима критерији је како видимо недефинисан и није га лако ваљано дефинисати. Ово су честа и озбиљна питања, а одговори на њих су изузетно тешки, из разлога што је тешко прецизно дефинисати критеријску варијаблу. Након дефинисања критеријске варијабле, слиједи проналажење таквих тестова који ће

бити у високој корелацији са критеријем. Тестови који систематски показују високу корелацију са критеријем смањавају се ваљаним тестовима. Постоје обимна истраживања како се утврђује ваљани критериј. Та истраживања су везана за бројне метричаре, који су дали ванредан допринос успостављању ваљаности композитних критерија. Утврђивање композитне критеријске варијабле као што је: квалитет неке појаве, социјални статус, здравствени статус, когнитивни и конативни ниво и сл., могуће је примјеном изузетно сложених статистичко-математичких метода, познатих као факторске анализе. Утврђивање ваљане критеријске варијабле уз помоћ факторске анализе и прве (главне) компоненте је велики проблем, због неизбјежне контаминираниости критеријске варијабле са фактором грешке. Када је ријеч о успостављању ваљаности критеријске варијабле нпр. ко је и на ком нивоу у погледу социјалне стратификације или ко је и на ком нивоу интелигенције или колика је припадност неког субјекта некој групи, проблем се рјешава методом факторске анализе, односно утврђивањем факторске ваљаности критеријске варијабле. Критеријска варијабла је композитна варијабла тј. сложена, а неопходно је у крајњем сажети на једну димензију, често названом првом (главном) компонентом. Поред проблема ваљаности критеријске варијабле, сусрећемо се са проблемом ваљаности предикторских варијабли. Разлика између критеријске и предикторских варијабли је у томе што се критеријска варијабла своди на једну димензију, док је предикторских варијабли у принципу, више. У циљу максималне ефикасности и предикторске варијабле се редукују на што је могуће мањи број, мада превелика парсимонија често иде на уштрб моћи предикције.

Предиктивна ваљаност

Скоро да нема области гдје није присутна селекција и потреба за селекцијом како у природи тако и у друштву. Проблем одабира односно селекције је посебно наглашено у спорту и то свакодневно. Селекција је присутна на сваком

тренингу, посебно на утакмици. Кад говоримо о селекцији обично мислимо на дугорочну селекцију, почев од најмлађег узраста. Ваљаност селекције условљена је ваљаношћу тестова и метода на основу којих се врши селекција. За неки тест можемо рећи да је ваљан ако резултат тог теста има високу корелацију са успјехом у конкретном занимању односно у спортској грани или спортској дисциплини. Наиме, тест је ваљан ако, уз помоћ теста, можемо вршити успјешна предвиђања у конкретной области или спортској дисциплини. Ваљан тест је онај тест са којим можемо предвидјети доброг ученика, студента, спортисту, радника, економисту, љекара, руководиоца (лидера) итд.

Критеријуми ваљаности

Озбиљан проблем је утврђивање критерија предиктивне ваљаности. Већ је речено да методе факторизације омогућавају успјешно долажење до ваљаног критерија. Након утврђивања критерија тј. критеријске варијабле тек тада има смисла приступити конструисању тестова тј. предиктивних варијабли које ће објашњавати критерије. односно изналажењу таквих тестова који ће бити у високој корелацији са критеријском варијаблом. Већ је речено да је критеријска варијабла врло сложена и она се често назива у факторској терминологији димензија или фактор. Примјер генерисаних варијабли имамо код утврђивања разних постигнућа (успјеха) у школи. студију. спорту, иначе у животу. На сва постигнућа поменута и бројна остала уутичу неке варијабле, више или мање те варијабле су у корелацији са димензијом постигнућа некада у већој или мањој мјери. Ваљани тестови су они који високо корелирају са критерије, односно који добро предвиђају постигнуће (успјех). Било би добро кад би имали такву варијаблу, односно такав тест са којим би без грешке могли предвидјети неко постигнуће. Нажалост за сада нема таквог теста нити таквих тестова помоћу којих у потпуности можемо предјети неко постигнуће, нпр. успјех на неком радном мјесту или факултету. Ипак има таквих тестова помоћу којих можемо доста

успјешно предвидјети постигнуће, односно успјех у некој активности, професији и сл. За такве тестове кажемо да су ваљани.

Теорија фактора и логика ваљаности

Ваљаност неког теста који по својој структури може бити једноставнији или сложенији, посматрамо помоћу укупне варијансе. Укупна варијанса је компонована од три варијансе и то:

1. Варијанса која је објашњена једним или више заједничких фактора (заједничка варијанса),
2. Варијанса која је објашњена само једним тестом (специфичној варијанси)
3. Варијанса грешке (необјашњена варијанса)

Предпоставља се да заједничка варијанса фактора и специфична варијанса неког тест нису у међусобној корелацији.

Теорема ваљаности може се представити у виду следеће једначине

$$\sigma_t^2 = \sigma_{fa}^2 + \sigma_{fb}^2 + \dots + \sigma_n^2 + \sigma_s^2 + \sigma_e^2$$

σ_t^2 = тотална варијанса (укупна варијанса теста)

$\sigma_{fa}^2, \sigma_{fb}^2, \dots, \sigma_n^2$ = варијанса заједничких фактора (заједничка варијанса)

σ_s^2 = варијанса која је специфична за сваки тест (специфична варијанса)

σ_e^2 = варијанса грешке

Ако подијелимо једначину са тоталном варијансом добијемо коефицијент један. Овом операцијом добили смо пропорције варијанси, чија је сума увијек једнака 1.

$$\frac{\sigma_t^2}{\sigma_t^2} = \frac{\sigma_{fa}^2}{\sigma_t^2} + \frac{\sigma_{fb}^2}{\sigma_t^2} + \dots + \frac{\sigma_n^2}{\sigma_t^2} + \frac{\sigma_s^2}{\sigma_t^2} + \frac{\sigma_e^2}{\sigma_t^2} = 1$$

Комуналитет неког теста представљен је сумом пропорција заједничких варијанси. Комуналитети су заједничка варијанса свих тестова, другим ријечима то су корелације између тестова. Величину комуналитета можемо израчунати ако од укупне варијансе одузмемо пропорцију варијансе грешке и пропорцију специфичне варијансе.

$$h^2 = 1 - \left[\frac{\sigma_s^2}{\sigma_t^2} + \frac{\sigma_e^2}{\sigma_t^2} \right]$$

У следећем примјеру дата су три теста један критериј и два фактора. У табели 1. су приказане пропорције заједничке варијансе, пропорције специфичне варијансе, пропорције варијансе грешке и сума пропорција заједничких фактора тј комуналитет.

Табела 1.

тестови	Fa	Fb	специфична варијанса	варијанса грешке	комуналитет x^2	укупна варијанса
1. тест	0.30	0.45	0.15	0.10	0.75	1
2. тест	0.60	0.20	0.05	0.15	0.80	1
3. тест	0.40	0.10	0.20	0.30	0.50	1
критерији Z	0.80	0.10	0.00	0.10	0.90	1

Први тест има 30% варијансе за коју је одговоран фактор Fa и 45% варијансе за коју је одговоран други фактор Fb. Сума варијанси је комуналитет овог теста који износ 0.75 или 75%. Ако комуналитету додамо пропорцију специфичне варијансе и пропорцију варијансе грешке, добивамо укупно варијансу која увијек износи 1.

Други тест има 60% варијансе за коју је одговоран фактор Fa и 20% варијансе за коју је одговоран фактор Fb, комуналитет је 0.80 или 80%., остатак до 1 или 100% односи се на специфичну варијансу и варијансу грешке.

Трећи тест како видимо има 50% варијансе за коју је одговоран фактор F_a и F_b , а 50% односи се на специфичну варијансу и варијансу грешке. Критеријум Z може се интерпретирати на сличан начин.

Табела 2. је редукована табела 1. Коефицијенти у табели 2 су корелације између два фактора три теста и критерија Z . Корелациони коефицијенти су израчунати тако што су израчунати квадратни корјенови из коефицијената у табели 1.

Корелациони коефицијенти у табели 2 упућују на степен засићења тестова и критерија са факторима F_a и F_b . Ово засићење често се назива сатурација факторима.

Табела 2. засићење-сатурација факторима

тестови	F_a	F_b
1. тест	0.55	0.67
2. тест	0.77	0.45
3. тест	0.63	0.32
критерији Z	0.89	0.32

Како видимо највећа засићеност са фактором F_a испољава критериј Z и износи 0.89, а најмања засићеност са фактором F_a испољава први тест и износи 0.55. Највећа засићеност фактором F_b има први тест и износи 0.67, а најмања засићеност фактором F_b има критериј Z и износи 0.32.

Корелација између два теста једнака је збиру унакрсних производа њихове засићености заједничким фактором. То је приказано помоћу следеће једначине

$$r_{ij} = a_i a_j + b_i b_j + \dots + n_i w$$

Наведена једначина објашњава предиктивну ваљаност тестова. У примјеру имамо да највећу предиктивну ваљаност за критериј Z је показао други тест, $r_{z2}=0.83$., а најмању предиктивну ваљаност за критерији Z је показао трећи тест, $r_{z3}=0.66$

$$r_{z1} = (0.89)(0.55) + (0.32)(0.67) = 0.70$$

$$r_{z2} = (0.89)(0.77) + (0.32)(0.45) = 0.83$$

$$r_{z3} = (0.89)(0.63) + (0.32)(0.32) = 0.66$$

Ваљаност предвиђања критерија

Ваљаност предиктивних тестова изражава се помоћу корелационог коефицијента и регресионих коефицијената. Основно питање је колико треба да износи корелациони коефицијент и регресиони коефицијент, да би се неки тест могао употријебити као ваљан предиктор.

Ваљаност тестова и примјена одређених метода за утврђивање ваљаности зависи од два момента и то:

1. Да ли се тест користи у сврху предикције резултата једног испитаника (субјекта)?
2. Да ли се тест користи као средство селекције?

Предвиђање критерија на основу регресионе једначине

Често се критериј дефинише као резултат (постигнуће). Познато је да су неки резултати или постигнућа лако мјерљиви, захваљујући утврђеним међународним физичким јединицама и мјерама. Дакле, познате су општеважеће јединице и мјере за: убрзање, силу, снагу, масу, рад, отпор итд.

Поред наведених примјера и јединица које су егзактне и углавном примјенљиве у природним наукама, у друштвеним наукама упућени смо на мјерења гдје су мјерне јединице и мјерни инструменти непрецизно дефинисани, тако смо често у ситуацији да неку појаву оцјењујемо и процјењујемо уз помоћ неких тестова који немају гаранцију ваљаности, поузданости ни дискриминативне осјетљивости. Дакле, критериј је композитан и сигурно се до њега долази уз помоћ више судија, који добро познају конкретан проблем. Примјер како се утврђује добро вино или добар мед. Познато је да то не може утврђивати било ко, у принципу не може један дегустатор (судија).

Даља процедура је статистичко-математичка, а своди се на израчунавање корелација између оцјењивача (судија) и извођење критеријске варијабле уз помоћ факторске анализе, тј. прве (главне) компоненте. Када смо ваљано дефинисали критеријску варијаблу, тада можемо приступити утврђивању ваљаности предиктивних варијабли тј. тестова. Уколико имамо само једну предиктивну варијаблу проблем

се ријешава примјеном линеарне једначине са двије непознате:

$$Y = a + bX \text{ гдје је: } Y = f(X)$$

Y – је зависна (критеријска) варијабла

X – је независна (предикторска) варијабла

a – регресиони параметар који у графичком смислу представља исходите функције у тачци $X = 0$.

b – регресиони параметар који одређује величину промјене зависно промјенљиве уколико се независно промјенљива промјени за једну јединицу.

Регресиони параметри a и b могу се израчунати на више начина. У принципу се ради о систему једначина са двије непознате:

$$\sum_{i=1}^n y_i = na + b \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2$$

Уколико имамо више предиктивних варијабли, проблем се рјешава примјеном линеарних једначина са више непознатих. Како је реалност доста комплексна, објективно смо упућени на примјену сложених (мултиплих) регресионих, корелационих и факторских модела.

Мултипли регресиони модел може се дефинисати на следећи начин:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Y – је зависно промјенљива (критеријска) варијабла

$X_{1i}, X_{2i}, \dots, X_{ki}$ – су независно промјенљиве (предикторске) варијабле

β – су непознате константе (параметри) регресионог модела

ε_i – стохастичка (случајна) грешка модела

Примјер:

Висина плате десет радника изражена у бодовима приказана је у (колони у). Важност радног мјеста радника вреднована је у распону од 5 до 10 (колона x_1). Квалификација радника вреднована је у распону од 5 до 10 (колона x_2). У примјеру варијабла Y је зависна-критеријска, а варијабле (X_1 и X_2) су независне-предикторске. Емпиријске вриједности приказане су у табели 3.

Табела 3.

n	X_1	X_2	бодови Y
1.	6	5	120
2.	7	6	145
3.	5	7	115
4.	8	10	160
5.	9	7	170
6.	10	9	200
7.	5	7	110
8.	9	8	180
9.	6	7	125
10.	5	6	130

Дакле имамо сложени модел (двije независне-предикторске варијабле и зависну у примјеру висина плате радника).

- а) Израчунај регресиону праву гдје је $Y=f(x_1, x_2)$
- б) Израчунај стандардну грешку регресије
- ц) Тестирај ваљаност регресионих коефицијента β_1, β_2
- д) Израчунај коефицијент мултипле корелације и коефицијент детерминације
- е) Тестирај значајност коефицијента корелације и коефицијента детерминације
- ф) Израчунај коефицијенте интеркорелације
- г) Израчунај коефицијент алинеације и индекс ефикасности предвиђања зависне варијабле

Регресиона права је:

$$Y_{(X_1, X_2)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 = 29.64 + 15.14 X_1 + 1.37 X_2$$

Коефицијенти: мултипле корелације, коефицијент детерминације, кориговани коефицијент мултипле корелације и стандардна грешка регресије приказани су у табели 4.

Табела 4.

R	R ²	Кориговано R	Стан.грешка регресије
0.97	0.94	0.93	8.27

Просјек квадрата одсујања и значајности регресије приказана је у табели 5.

Табела 5.

-	Сума квадрата	df	Просјек квадрата	F	Значајност
Регресија	7993.3	2	3996.6	58.38	0.00
Резидуал	479.21	7	68.46		
Тотал	8472.5	9	-		

Стандардизовани и нестандардизовани регресијски параметри и њихова значајности приказана је у табели 6

Табела 6.

-	нестад. β	станд.гр β	станд. β	t	Значајност
константа	29.64	13.98	-	2.12	0.07
X ₁	15.14	1.83	0.93	8.29	0.00
X ₂	1.37	2.33	0.06	0.59	0.57

Предиктивна вриједност прве варијабле X₁ тестирана преко т-теста је: $t = 8.29$. Грешка је мања од 0,00 што упућује на закључак да постоји изузетно висока предиктивна вриједност варијабле X₁ на критеријску варијаблу Y. Предиктивна вриједност друге варијабле X₂ тестирана преко т-теста је $t = 0.59$. Грешка је већа од 0.57, што упућује на закључак да главну предиктивност преузима варијабла X₁.

Коефицијент мултипле корелације је: $R = 0,97$, а коефицијент детерминације је: $R^2 = 0,94$.

Коефицијенти R и R² су врло високи и статистички значајни, грешка је мања од 0,00.

Коефицијент алинације је:

$$k_a = \sqrt{1 - r_{xy}^2} = \sqrt{1 - 0.94} = 0.24$$

Коефицијент ефикасности предвиђања је:

$$I_e = 100 \left[1 - \sqrt{1 - r_{xy}^2} \right] = 100 \left[1 - \sqrt{1 - 0.94} \right] = 76\%$$

Коефицијент ефикасности износи 0.76 или 76% на основу којег можемо закључити да је предиктивна вриједност варијабли (X_1 и X_2) задовољавајућа.

Матрица интеркорелације табела 7.

Табела 7.

-	x_1	x_2	y
x_1	1,00	0.60* ¹	0.97** ²
x_2		1,00	0.62*
y			1,00

Вриједности коефицијената интеркорелације упућују на закључак да је интеркорелација прве предикторске варијабле са критеријем изузетно висока, позитивна и статистички врло значајна. То потврђује и регресиони коефицијент предикције.

Корелација између критеријске варијабле и друге предикторске варијабле је знатно нижа, што потврђује и коефицијент предикције.

На основу свих релевантних регресионих и корелационих показатеља намеће се закључак да укључивањем друге предикторске варијабле није битно повећана укупна предиктивна вриједност, наиме прва предиктивна варијабла је високо сатурирана другом. У конкретном примјеру можемо говорити о примарној ваљаности варијабле X_1 . Дакле, положај на радном мјесту значајније утиче на висину плате

¹ Статистичка значајност гдје је грешка мања од 0.05

² Статистичка значајност гдје је грешка мања од 0.01

у односу на квалификацију. Корелациони коефицијент између предикторских варијабли износи 0.60 и статистички је значајан са грешком мањом од 5%.

Умјесто закључка

Поред поменутих метода утврђивања ваљаности, како предикторских тако и критеријских варијабли, постоје и друге методе.

Ваљаност мјерних инструмената није једини критериј да би неки тест, анкета, упитник и сл. били прихваћени и употријебљени. Постоје и друге метријске карактеристике мјерних инструмената као што су поузданост (релијабилност) и осјетљивост (дискриминативност) тестова. Наиме, битно је напоменути да ваљаност и поузданост тестова није једна те иста ствар. У неким случајевима битно се разликује ваљаност од поузданости тестова. Пожељно је да мјерни инструменти: тестови, анкете и упитници, поред високе ваљаности, посједују високу поузданост и осјетљивост.