

PSYn4790 Psychometrika: měření v psychologii



Shoda posuzovatelů | 29. 10. 2019

Posuzování / hodnocení v psychologii?

- ▶ **Posuzovací škály**
 - ▶ Intenzita prožitků, příznaků nemoci, ...
- ▶ **Pozorování a observační studie**
 - ▶ Bylo / nebylo pozorováno nějaké chování? Do jaké kategorie zařadit to, co jsem pozoroval(a)?
- ▶ **Psychologická diagnostika**
 - ▶ Diagnostický nález, skóry z checklistu, ...
- ▶ **Hodnocení výkonu**
 - ▶ V rámci školní třídy, v testu, pořadí uchazečů při náboru zaměstnanců, ...
- ▶ **Kódování v kvalitativním výzkumu**
- ▶ ...napadne vás ještě nějaký příklad?



Vřelost

▶ 1 – Výrazný nedostatek lásky

- ▶ Takto jsou hodnoceni rodiče respondenta, kteří nejenže nebyli oporou jeden druhému, ale odmítali vzájemně spolupracovat nebo spolu soupeřili, nechovali se k sobě nikterak láskyplně či ohleduplně. Takto se posuzují vztahy charakteristické přítomností hněvu a nepřátelských projevů nebo vztahy, v nichž se rodiče k sobě chovali chladně a nezúčastněně. Toto hodnocení se využívá také v případech, kdy jeden z rodičů druhého psychicky či fyzicky týral či zneužíval. Manželství, která byla ukončena rozvodem, se hodnotí ▶

▶ 3 – Nedostatek vřelosti

- ▶ Vztah se vyznačuje mírnou, nicméně neadekvátní nebo nekonzistentní oporou. Potřeby jednoho nebo obou rodičů bývají občas uspokojeny, většinou jsou však přehlíženy. Tyto páry se vyznačují vzájemnou lhostejností, každý z partnerů žil víceméně vlastním životem, které se prolínaly pouze sporadicky. Toto hodnocení se užívá i pro páry, které spolu sice žily aktivně, ale jejich vzájemná interakce byla charakterizována spíše negativně, jednali spolu například s neúctou a s nedostatečným poskytováním opory.

▶ 5 – Ani neláskyplný, ani aktivně láskyplný

- ▶ Respondent hodnotí vztah svých rodičů jako „dobrý“ či „láskyplný“, ale neuvádí detaily, které by tento pohled potvrdily či vyvrátily. Pokud je k dispozici více detailů, lze říci, že rodiče poskytovali adekvátní emocionální oporu jeden druhému. Přestože nijak výrazně nerozuměli potřebám toho druhého, snažili se být si ve většině oblastí soužití nápomocni.

- ▶ Někteří respondenti se mohou při popisu soustředit na dovednosti rodičů v oblasti výchovy, a výzkumník/tazatel tak získává dojem, že manželství rodičů hrálo sekundární roli oproti výchově dětí, která byla pro pár prvořadá. Toto hodnocení také slouží jako průměrné hodnocení, pokud se manželé v minulosti nechovali k sobě láskyplně, ale tyto negativní epizody byly ve vztahu vystřídaný či vynahrazeny věrohodnými láskyplnými či obětavými činy.

▶ 7 – Láskyplný

- ▶ Přestože se ve vztahu mohly objevovat problémy, rodiče se vůči sobě projevovali láskyplným a chápajícím způsobem. Lze vytušit, že vztah byl plný důvěry a opory. Hodnocení 7 je odpovídající, pokud respondent souvisle a srdečně hovoří o vztahu rodičů a udává, že se k sobě pár choval s láskou, ale současně to dokládá menším množstvím specifických detailů.

▶ 9 – Velmi láskyplný

- ▶ Tito rodiče se k sobě aktivně chovali láskyplně a s vzájemnou náklonností a očividně se cítili dobře a užívali si vzájemnou společnost. Respondent uvádí konkrétní příklady, jak si byli jeho rodiče oporou sobě navzájem, partnersky, tak svým dětem jako rodiče. Poskytovali si navzájem přátelství a útěchu. Není nutné, aby byl vztah popisován jako absolutně perfektní, pro toto hodnocení se rozhodujeme tehdy, existují-li silné důkazy, že se rodiče navzájem milovali, respektovali a podporovali jeden druhého.

Proč se zabývat shodou?

- ▶ Kdo může zaručit „objektivitu“ posuzování / hodnocení?
 - ▶ I pokud jsou hodnotící kritéria jasně definována, jsou stejně chápána a používána?
- ▶ Zajištění reliability výzkumné metody
 - ▶ **Hodnocení** na posuzovacích škálách, pozorování chování, hodnocení výkonu
 - ▶ **Administrace** diagnostických metod – vliv administrátora
- ▶ Zajištění interní validity výzkumných designů
 - ▶ Shoda posuzovatelů, pozorovacích schémat atp.



Co dělat s (ne)shodou?

- ▶ Shodu můžeme „vynutit“ (např. použít průměrné hodnocení)
 - ▶ Tím se ale připravujeme o informace
- ▶ ...nebo ji můžeme nějak kvantifikovat a vyjádřit její míru
 - ▶ Míra (ne)shody je důležitý a interpretovatelný údaj.
- ▶ Po kvantifikaci můžeme (ne)shodu efektivněji studovat
 - ▶ Jak velké jsou mezi hodnotiteli rozdíly?
 - ▶ Jsou tyto rozdíly náhodné?
 - ▶ Jsou tyto rozdíly systematické (např. rozdílně „přísní“ hodnotitelé)?



Dvě hlavní použití míry (ne)shody

- ▶ Lze několik různých hodnocení „redukovat“ na jediný údaj?
 - ▶ Kolik spolu mají hodnocení „společného“, jde stále o tu stejnou proměnnou?
- ▶ Jaká je reliabilita takovéto redukce v případě...
 - ▶ ... průměrného/výsledného hodnocení několika hodnotiteli?



Dva hlavní typy neshody

- ▶ 1. Nesystematický rozdíl mezi hodnotiteli
 - ▶ 2. Systematický rozdíl mezi hodnotiteli
-
- ▶ ...zpravidla ale pozorujeme kombinaci obou typů



Nominální proměnné

- ▶ 1. Nesystematický rozdíl – náhodný rozdíl
- ▶ 2. Systematický rozdíl mezi hodnotiteli – rozdíl v poměru

- ▶ *Příklad:* Při náboru do armády posuzují dva psychologové, jestli se rekruti hodí spíš na pilota nebo na tankistu.
 - ▶ Jeden z psychologů může dávat více závěrů „pilot“ než druhý (systematický rozdíl v poměru – ale nemusí to být tak „čisté“)
 - ▶ Oba psychologové mají tento poměr stejný, ale neshodnou se v x % případů (nesystematický rozdíl)



(alespoň) Ordinální proměnné

- ▶ 1. Nesystematický rozdíl – (ne)shoda v pořadí
- ▶ 2. Systematický rozdíl mezi hodnotiteli – (ne)shoda v průměru

- ▶ *Příklad:* Během náboru zaměstnanců mají dva psychologové za úkol obodovat každého uchazeče na stupnici 0-10
 - ▶ Jeden z psychologů je „přísnější“ a hodnotí každého méně body
(systematický rozdíl v průměru)
 - ▶ Oba psychologové se neshodnou na tom, kdo je nejlepší, kdo druhý nejlepší, třetí nejlepší, atd. (nesystematický rozdíl)

Jaké otázky si klást?

- ▶ Kdo se má shodovat s kým?
 - ▶ **Shoda administrátorů:** Vede individuální vyšetření různými administrátory ke stejným výsledkům? (WISC...)
 - ▶ **Shoda hodnotitelů:** Ohodnotí již získaný protokol různí lidé stejně? (ROR; kvalitativní výzkum).
 - ▶ **Intra-rater reliability:** Obdobné otázky, ale pro jednoho administrátora/hodnotitele v různých časech.

- ▶ Kolik bylo hodnotitelů?
 - ▶ Jeden, ve alespoň dvou různých časech
 - ▶ Dva
 - ▶ Více



Typy proměnných a související hypotézy

- ▶ **Nominální nebo ordinální**
 - ▶ Jaká je absolutní/relativní míra shody 2 nebo více osob?
- ▶ **Ordinální**
 - ▶ Jaká je míra shody v pořadí hodnocených osob?
 - ▶ Jaká je míra shody ve střední hodnotě?
 - ▶ Absolutní shoda (pořadí a střední hodnoty dohromady).
- ▶ **Intervalová/poměrová**
 - ▶ Jaká je míra shody v pořadí hodnocených osob?
 - ▶ Jaká je míra shody ve střední hodnotě?
 - ▶ Absolutní shoda (pořadí a střední hodnoty dohromady).
- ▶ V psychologické diagnostice je typickým postupem ověření shody v případě položek nominálními/ordinálními statistikami (analogie korigovaných korelací se škálou) a pro celkové skóry intervalovými statistikami.



Nominální proměnné (n = 2)

▶ Cohenovo kappa

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

- ▶ Pro dva hodnotitele.
- ▶ Kolikrát je jejich shoda vyšší než náhodná shoda?
- ▶ P_o = pozorovaná shoda hodnocení
- ▶ P_e = shoda hodnocení očekávaná na základě prosté náhody



Nominální proměnné (n = 2)

► Cohenovo kappa

	0	1	SUM
0	35	13	48
1	3	49	52
SUM	38	62	100

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o = pozorovaná shoda
hodnocení

$$= \frac{35 + 49}{100} = 0.84$$

P_e = očekávaná shoda hodnocení
na základě prosté náhody

P_e = P-nost, že oba řeknou „0“ a nebo že oba řeknou „1“
(sčítáme pravděpodobnosti)



Nominální proměnné (n = 2)

► Cohenovo kappa

	0	1	SUM
0	35	13	48
1	3	49	52
SUM	38	62	100

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o = pozorovaná shoda
hodnocení

$$= \frac{35 + 49}{100} = 0.84$$

P_e = očekávaná shoda hodnocení
na základě prosté náhody

$$P_e = \left(\frac{38}{100} * \frac{48}{100} \right) + \left(\frac{62}{100} * \frac{52}{100} \right) = 0.505$$

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.84 - 0.505}{1 - 0.505} = 0.677$$



Nominální proměnné (n = 2)

► Cohenovo kappa

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.84 - 0.505}{1 - 0.505} = 0.677$$

- Interpretace (volná): „Shoda hodnotitelů je **68 %** mezi prostou náhodou a 100% shodou.“
- Interpretace (exaktnější): Podíl nárůstu shody oproti náhodné shodě je 0,68 maximálního možného nárůstu.
- Cohenovo kappa nabývá hodnot mezi -1 a 1
- SPSS: Analyze/Descriptives/Crosstabs: Statistics/Kappa



Nominální proměnné ($n > 2$)

- ▶ Cohenovo kappa je určeno jen pro dva hodnotitele.
- ▶ Pro n hodnotitelů je zobecněním **Fleissovo kappa**.
- ▶ Stejná logika a interpretace, pouze složitější výpočet.
- ▶ SPSS: plug-in SPSS Fleiss Kappa
 - ▶ Pro dva hodnotitele je výsledek identický s Cohenovým kappa a oproti běžnému SPSS dialogu poskytuje interval spolehlivosti.



Ordinální proměnné

- ▶ Lze do jisté míry použít běžné statistiky, které už znáte:
- ▶ Shoda středních hodnot (přísnost hodnotitelů):
 - ▶ 2 hodnotitelé: Mann-Whitney („neparametrický t-test“)
 - ▶ N hodnotitelů: Kruskal-Wallis („neparametrická ANOVA“)
- ▶ Shoda pořadí:
 - ▶ 2 hodnotitelé: Běžná pořadová korelace (Spearman, Kendall) pro shodu pořadí
 - ▶ N hodnotitelů: Kendallův koeficient konkordance (W) – viz dále
- ▶ ...ale máme k dispozici lepší nástroje 😊



Ordinální proměnné (n=2)

- ▶ Můžeme k nim přistupovat jako k nominálním proměnným, ale výsledkem je obvykle podhodnocení shody
- ▶ Řešením je „vážená kappa“.
- ▶ Neshody jsou váženy různým způsobem – čím dále od diagonály, tím jde o větší neshodu
 - ▶ Jak vážit?
 - ▶ lineární váhy (vzdálenost od diagonály)
 - ▶ kvadratické váhy (vzdálenost od diagonály²)

shoda		hodnotitel A		
		1	2	3
hodnotitel R	1	15	12	1
	2	9	23	5
	3	0	8	17

lineární váhy		hodnotitel A		
		1	2	3
hodnotitel R	1	0	1	2
	2	1	0	1
	3	2	1	0



Ordinální proměnné (n=2)

- ▶
- ▶ **Příklad vpravo:**
 - ▶ Běžná kappa, $\kappa = 0,401$
 - ▶ Lineární váhy, $\kappa_{wlin} = 0,502$
 - ▶ Kvadratické váhy, $\kappa_{wquad} = 0,620$
- ▶ **SPSS: plug-in SPSS Weighted Kappa**

shoda		hodnotitel A		
		1	2	3
hodnotitel R	1	15	12	1
	2	9	23	5
	3	0	8	17

lineární váhy		hodnotitel A		
		1	2	3
hodnotitel R	1	0	1	2
	2	1	0	1
	3	2	1	0

Ordinální proměnné ($n > 2$)

▶ Vážená Fleissova kappa

- ▶ Kombinace Fleissovy kappy a vážené Cohenovy kappy
- ▶ Bere v potaz shodu pořadí i středních hodnot

▶ Kendallův koeficient konkordance (W)

- ▶ Určeno pro shodu pořadí
- ▶ Původní verze nepočítá s „remízami“, ale existuje zobecněná verze
- ▶ Odpovídá na otázku, nakolik hodnotitelé udávají stejné pořadí toho, co hodnotí



Intervalové proměnné

- ▶ Opět lze do jisté míry použít běžné statistiky
- ▶ Shoda průměrů (přísnost hodnotitelů):
 - ▶ 2 hodnotitelé: t-test
 - ▶ N hodnotitelů: one-way ANOVA
- ▶ Shoda pořadí:
 - ▶ 2 hodnotitelé: Pearsonova korelace
 - ▶ N hodnotitelů: ICC2 (viz dále)
- ▶ ...ale máme k dispozici lepší nástroje 😊 (ano, už zase...)



Intervalové proměnné

- ▶ Teorie zobecnitelnosti 😊
- ▶ Pro zjednodušení jsou definovány 2×3 základní typy intra-class korelací, které jsou konkrétními speciálními případy teorie zobecnitelnosti.
 - ▶ Intra-class korelace: Jak moc se podobají hodnoty v rámci stejných tříd (hodnocených osob)?
 - ▶ Vnitrotřídní korelace.
 - ▶ Inter-class korelace: Jak moc se podobají hodnoty napříč třídami (hodnotitelem A a hodnotitelem B).
 - ▶ Příkladem je Pearsonova korelace.



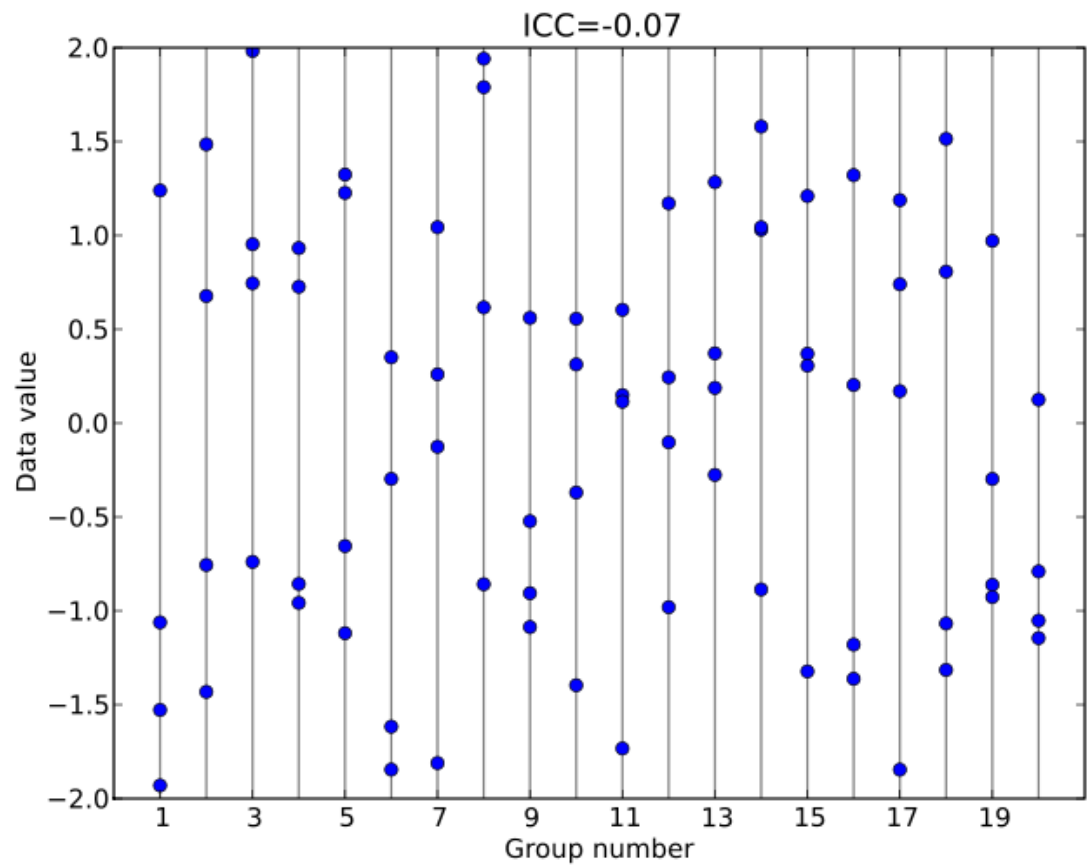
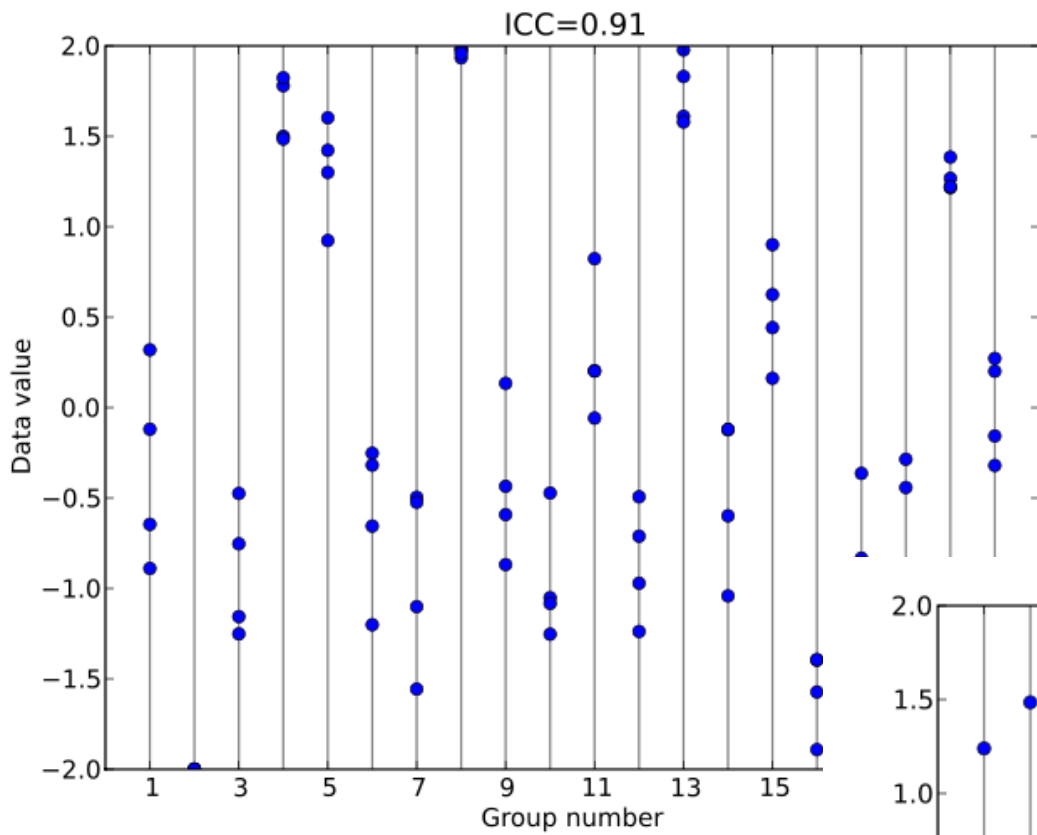
Intra-class / vnitrotřídní korelace

- ▶ Tři typy / modely (proč modely?):
 - ▶ ICC1: každá „věc“ je hodnocena stejným počtem „náhodných“ hodnotitelů, kteří jsou ale pokaždé jiní (žádný hodnotitel nehodnotí víc jak jednu „věc“)
 - ▶ ICC2: každá „věc“ je hodnocena stejným počtem „náhodných“ hodnotitelů, ti jsou pokaždé stejní (každý hodnotitel hodnotí každou „věc“)
 - ▶ Typicky tohle je to, co chcete.
 - ▶ ICC3: každá „věc“ je hodnocená stejným počtem nenáhodných hodnotitelů, ti jsou pokaždé stejní (každý hodnotitel hodnotí každou „věc“)

Intra-class / vnitrotřídní korelace

- ▶ Tyto tři modely se dále dělí podle toho, jestli reálně dochází k:
 - ▶ Udělení jednoho hodnocení jedním hodnotitelem -> $ICC(x, 1)$
 - ▶ Udělení průměrného hodnocení od všech hodnotitelů -> $ICC(x, k)$

- ▶ V kombiaci s předchozím slidem:
 - ▶ $ICC(1, 1)$, $ICC(1, k)$
 - ▶ $ICC(2, 1)$, $ICC(2, k)$
 - ▶ $ICC(3, 1)$, $ICC(3, k)$



Krippendorfova alfa

- ▶ **Zobecnění konceptu klasického koeficientu alfa (např. Cronbachovy).**

- ▶ Cronbachova alfa: $\alpha = 1 - \frac{\text{chybový rozptyl}}{\text{celkový rozptyl}}$

- ▶ (plus nějaké korekce na počet stupňů volnosti)

- ▶ Krippendorfova alfa: $\alpha = 1 - \frac{\text{pozorovaná neshoda}}{\text{očekávaná neshoda}} \sim$

- $\sim 1 - \frac{\text{rozdílnost v hodnocení subjektů}}{\text{rozdílnost subjektů} + \text{rozdílnost v hodnocení subjektů}} (= \text{celková rozdílnost})$

- ▶ **Použitelné pro nominální, ordinální i intervalové proměnné a libovolný počet hodnotitelů.**

- ▶ Jen se různým způsobem vyjádří pozorovaná a očekávaná neshoda.
 - ▶ **Díky tomu stejný význam napříč různými typy proměnných, koeficienty lze srovnávat!**
 - ▶ Použitelné i v případě chybějících dat.



Kudy na to?

- ▶ V programu R balíčky:
 - ▶ irr
 - ▶ raters
 - ▶ concord
 - ▶ něco málo i v psych package
- ▶ Různé pluginy do SPSS
- ▶ Ad-hoc programy, např. on-line kalkulačka
<https://nlp-ml.io/jg/software/ira/>



Intepretace výsledku

- ▶ Jak velká „shoda“ je zapotřebí?
- ▶ Co vlastně znamená neshoda?
 - ▶ V diagnostice
 - ▶ Ve výzkumu
 - ▶ U kódování testu?
 - ▶ U sledování záznamu?
 - ▶ U hodnocení výkonu/kompetence?

