

Validita: Realistické pojetí

PSYb2590: Základy psychometriky | **Přednáška 3**

5. 3. 2024 | Petr Palíšek & Petra Hubatková

Merkur – Země – Venuše – Slunce

Co nepatří mezi ostatní?

A) Slunce, protože je to jediná hvězda

B) Země, protože je to jediné z nabízených těles, kde existuje život

C) Země, protože je to jediné z nabízených těles, které není viditelné na obloze

D) ...







**I'M GOING ON
AN ADVENTURE!**

Validita: Cesta tam a zase zpátky

Validita

Validita výzkumu: platnost výsledků studie vzhledem ke skutečnosti.

Validita diagnostické metody: několik různých pojetí závislých na čase / filosofické pozici autorů

Validita

Validita výzkumu: shoda mezi interpretací výsledků studie a skutečností

Validita diagnostické metody: několik různých pojetí závislých na čase / filosofické pozici autorů

Validita na počátku věků

TEST JE VALIDNÍ, POKUD MĚŘÍ TO, CO MÁ MĚŘIT.

Jednoduchá myšlenka, stará skoro jako konstrukce psychologických testů sama:

T. L. Kelley, žák Thorndika, ne ten sociální psycholog (1927):

- Validita = the **extent** to which a test “**really measures what it purports to measure**” (p. 14) and is appropriate for a “*specifically noted purpose*” (p. 30)

Ve druhé přednášce o validitě budeme mluvit i o jiných pojetích založených na druhé části této věty.

Kolik validit znáš, tolikrát jsi psychometrikem

Operacionalisté/antirealisté: Výsledky validního testu korelují s kritériem (např. výsledky inteligenčního testu predikují pracovní výkon)

Logičtí pozitivisté: Výsledky validního testu mají odpovídat teoretickým očekáváním („zákonům“)

Konstruktivisté: Validita výsledků testu je tvořena v interakci testu a interpretujícího – klíčová je tak validní interpretace

Realisté: Test má měřit skutečný atribut. Validní je tedy takový test, který měří existující konstrukt.

Kolik validit znáš, tolikrát jsi psychometrikem

Operacionalisté/antirealisté: Výsledky validního testu korelují s kritériem (např. výsledky inteligenčního testu predikují pracovní výkon)

Logičtí pozitivisté: Výsledky validního testu mají odpovídat teoretickým očekáváním („zákonům“)

Konstruktivisté: Validita výsledků testu je tvořena v interakci testu a interpretujícího – klíčová je tak validní interpretace

Realisté: Test má měřit skutečný atribut. Validní je tedy takový test, který měří existující konstrukt.

Validita (dle Dennyho Borsbooma)

Test je validní, pokud měří to, co má měřit.

Z toho plyne, že pro hodnocení validity testu je podstatné:

1. jestli to, co měří (tj. **atribut**), existuje, nebo nikoliv (*jinak to nemá smysl měřit*)
2. jestli je s tímto atributem test propojen (*jinak to možná má smysl měřit, ale test to nedokáže*)

Takový přístup k validitě se dá popsat jako **realistický** (tj. vycházející z realistické ontologie)

Validita (dle Dennyho Borsbooma)

V realismu lze: **Test je validní, pokud měří to, co má měřit.**

Přesněji specifikovat na:

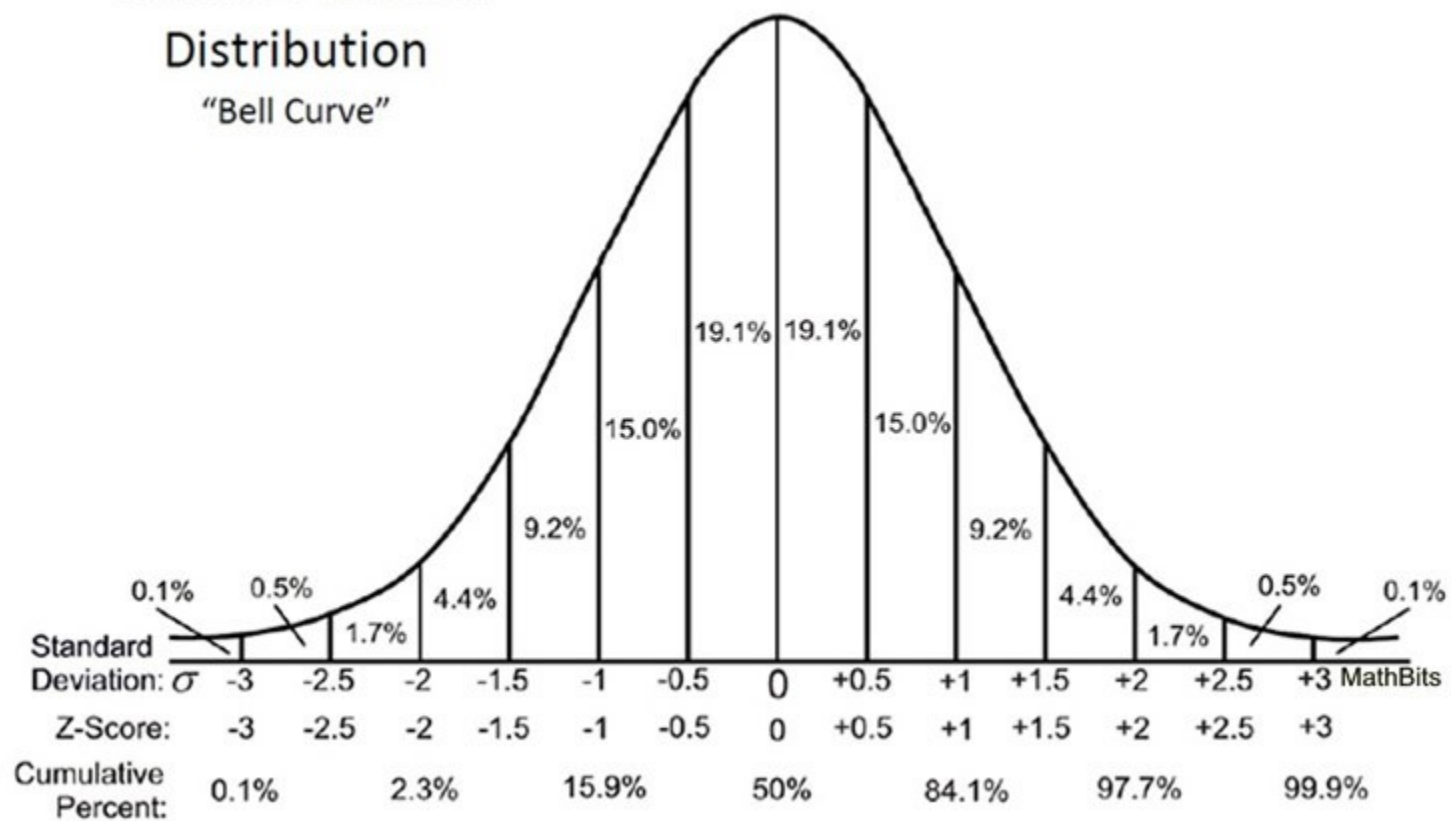
Test je validní, pokud

(1) **měřený atribut existuje** a zároveň

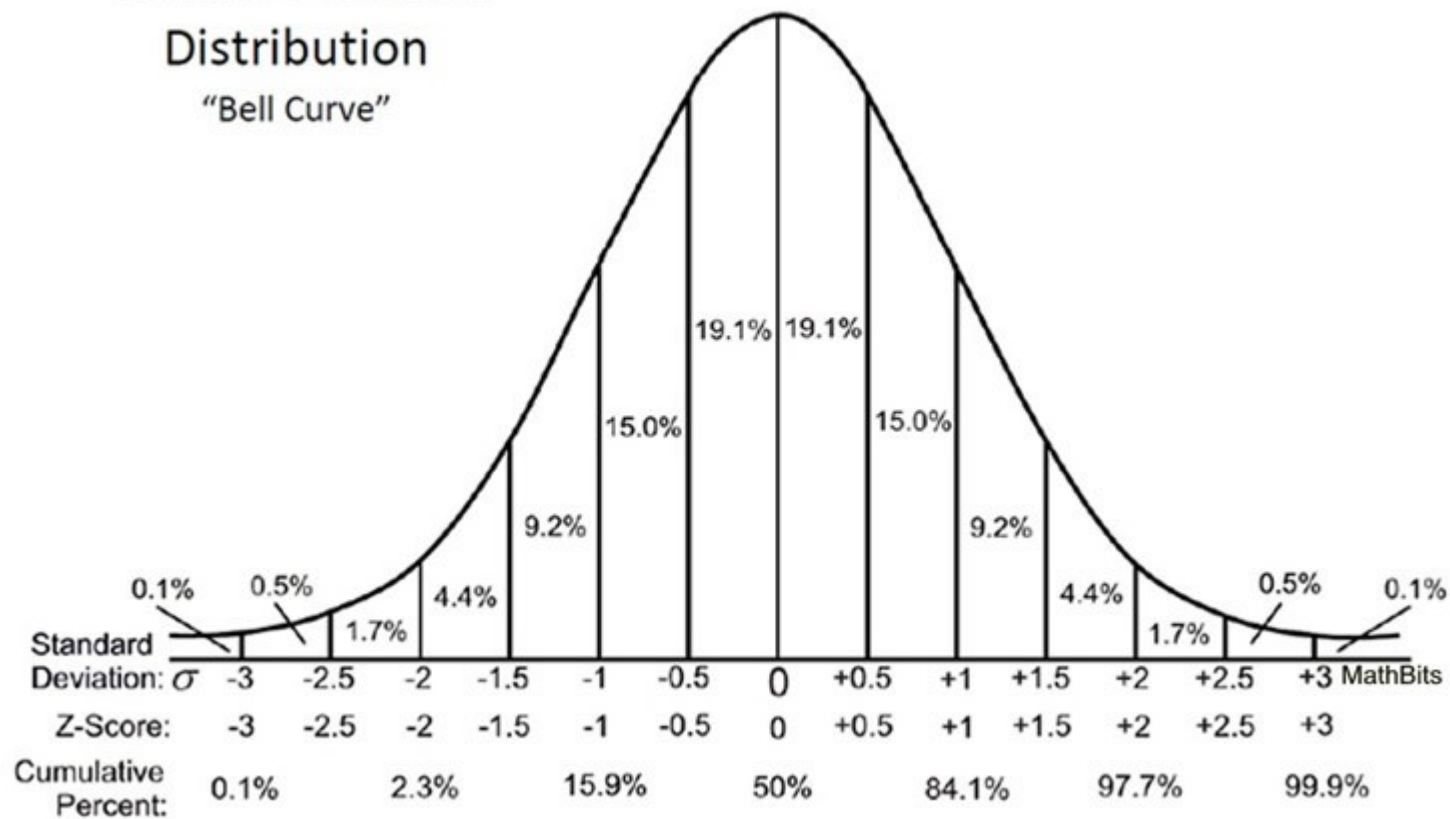
(2) **variabilita v atributu kauzálně způsobuje variabilitu v testových skórech.**



Standard Normal Distribution "Bell Curve"



Standard Normal Distribution "Bell Curve"



REWARDS STORE

Available Points:
ø5,025

✕

All
▾

| | Reward | Point Cost | ▲ |
|--|------------------|------------|---|
| | Instant Fun | ø100 | |
| | Instant Hygiene | ø100 | |
| | Confident Potion | ø200 | |
| | Energized Potion | ø200 | |
| | Flirty Potion | ø200 | |
| | Focused Potion | ø200 | |
| | Happy Potion | ø200 | |



Příklady z reálného
života?

BA DUM TSSS



Příklady z reálného života?

Důsledky realistického pojetí

Validita je vlastností testu

Ať test používáte k čemukoliv, (1) existence atributu a (2) vztah testu a atributu zůstávají pořád stejné

Test validní buď je, nebo není

Atribut může mít s testem různě silný vztah, ale na tom nezáleží, pro validitu stačí splnit dvě podmínky viz výše

Test musí být ukotven v psychologické teorii

- měření neexistuje na zelené louce – nelze měřit atribut, u něhož nevíme, jak funguje, nebo co to je
- musí existovat nějaký kauzální mechanismus

Co to znamená, že je nástroj ukotvený v teorii?

Jaký je data generující proces?

Obsahová validita

Co to znamená, že je nástroj ukotvený v teorii?

Jaký je data generující proces?

*Somewhere in the **chain of events that occur between item administration and item response**, the measured attribute must play a causal role in determining what value the measurements out-comes will take; otherwise, the test cannot be valid for measuring the attribute. Importantly, this implies that the problem of validity cannot be solved by psychometric techniques or models alone. On the contrary, it must be addressed by substantive theory. Validity is the one problem in testing that psychology cannot contract out to methodology.*

Atribut = latentní proměnná?

Atribut je realističtější pojem, více spojený se “schopností” (ability)

Konstrukt zahrnuje širší spektrum fenoménů, které nemusí odpovídat představám o “schopnosti” ve stylu g-faktoru

Příklad: Balance scale task (Inhelder and Piaget, 1958)

Jak děti rozumí vztahu mezi vzdáleností a váhou?

Teoretický předpoklad: Existují různá stádia kognitivního vývoje. V různých vývojových obdobích děti (preoperační stádium, stádium konkrétních operací, stádium formálních operací) budou k úkolu přistupovat jinak (s tím, jak se jim vyvíjí kognitivní funkce)

Na základě sesbíraných dat potom provedli analýzu latentních tříd, jejíž výsledky odpovídaly předpokládané čtyři vývojová stádia



Co to znamená, že je nástroj ukotvený v teorii?

Obsahová validita

Whenever I proposed to substantive psychologists that their theories should provide information with respect to the question of how a latent attribute determined the responses to questionnaire items, they looked at me as if they witnessed water burning. However, most of these same psychologists will have little problems in identifying why certain items should be included in a test. Usually, their answers either implicitly or explicitly explain how the items tap attributes that hang together systematically.

Obsahová validita

racionální posuzování validity

Odpovídá **obsah testu** měřenému konstruktů a účelu měření?

- Můžeme odpovědi na položky (v zamýšlené populaci) považovat za projevy, znaky, symptomy různých úrovní měřeného konstruktů?
- Zachycují položky všechny důležité druhy projevů (symptomy, znaky..)?
- Je vliv příbuzných, podobných či souvisejících konstruktů minimální?

Někdy též **výběrová validita**: jsou položky dobře *vybrané* z univerza možných položek?

Důkazy obsahové validity:

- Posouzení metody/položek experty.
- Racionální argumentace: Design tvorby položek. Soulad s teorií v plné šíři.

Obsahová validita vs. „face validity“:

- **Zjevná validita**: shodu konstruktů s metodou posoudí i laik správně.
- **Zdánlivá validita**: metoda ve skutečnosti měří něco jiného, než si laik myslí.

Validita položek

Každá položka obsahuje dvě hlavní rozptylové složky

- =příčiny rozdílů mezi lidmi v pozorování/odpovědích

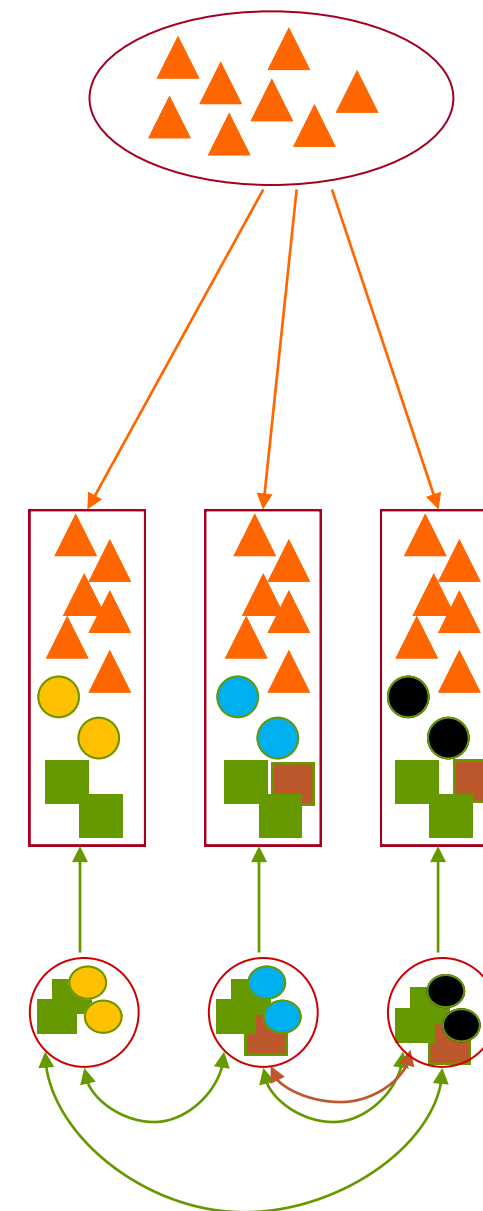
První je **společná** - týká se měřeného atributu.

- Pozorování/Odpověď je **způsobena** měřenou latentní proměnnou. ▲

Druhá složka je **specifická** pro danou položku.

- Náhodné vlivy na pozorování/odpovědi v daném čase. ●
- Další systematické vlivy/důvody na pozorování/odpovědi
- Specifické faktory

Položky by se měly „překrývat“ ve společném rozptylu ▲ a dobře pokrývat celé významové spektrum měřeného atributu a zároveň co nejméně „překrývat“ ve specifických složkách. ■■



Validita položek

Reliabilita testu: podíl společného a celkového r.: $\frac{\text{▲}}{\text{▲} + \text{●} + \text{●} + \text{●} + \text{■} + \text{■}}$

Specifický, ale nesdílený rozptyl ● ● ● snižuje pouze reliabilitu.

Specifický, ale sdílený rozptyl ■ ■ může reliabilitu nadhodnocovat.

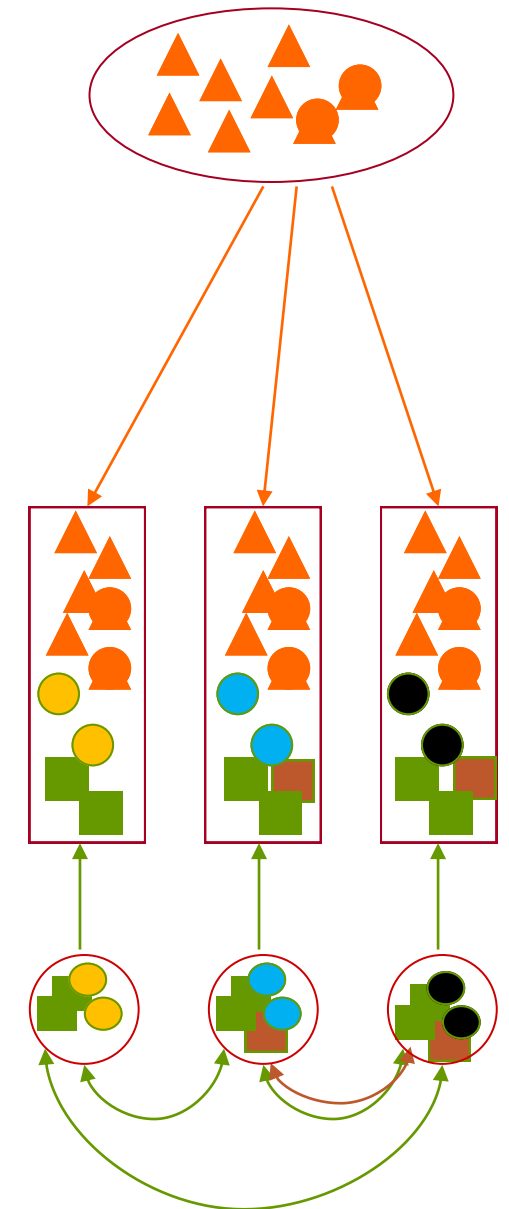
- Např. Cronbachovo alfa, split-half atp. Nelze dost dobře odlišit od ▲ .

Sdílený specifický rozptyl ■ ■ navíc snižuje validitu.

- Systematické zkreslení, které nelze oddělit od společného rozptylu ▲ .

Společný rozptyl (konstrukt) může být tvořen více fasetami: ▲ a ● .

Validitu snižuje i to, pokud část společného rozptylu ● není v položkách obsažena.



Tvorba položek

Dobré položky

- Jsme si jistí společným faktorem – *Opravdu měřený konstrukt způsobuje tohle?*
- Známe i specifické faktory (může jich být hodně) a nejlépe je i kontrolujeme.
- Máme představu o tom, jakými náhodnými vlivy může být odpověď zasažena.

Položky dohromady v jedné metodě.

- Měří všechny **společný faktor** (nebo známe jejich „faktorovou strukturu“).
- Pokrývají celou paletu možných projevů (a různých úrovní) konstruktů.
- Specifické faktory nesdílí.
- Nechybí jim žádný z „aspektů“ konstruktů.

Validita – vztah k teorii

Bez teorie není možné posoudit ani obsahovou validitu, ani data generující proces

However, a century of experience with test construction and analysis clearly shows that it is very hard to find out where the scores are coming from, if tests are not constructed on the basis of a theory of item response processes in the first place.

Statistika / metodologie nás v tomto sama o sobě nemůže spasit, její síla vychází jen z propojení s teorií

As is often the case in psychology, we have beautiful models but too little theory to go with them

“statistical machinery” vs. vysvětlení

Problémy v realistickém ráji

Co když latentní proměnná není to nejlepší vysvětlení pro atribut?

Problematické ověření v kontextu psychologie

Příliš vágní teorie

Validizace

*The fact that the crucial ingredient of validity involves the causal effect of an attribute on the test scores implies that the **locus of evidence for validity lies in the processes that convey this effect.***

Proces získávání empirických důkazů o validitě

Kognitivní interviews (např. Narrative Response Models)

Kognitivní modelování (např. pomocí reakčních časů a time-series dat)

Shoda modelů měření s daty (např. faktorová analýza)

Denny Borsboom

- holandský psychometrik *1973
- University of Amsterdam
- student Gideona Mellenbergha
- autor *Measuring the Mind* (jeho disertace)
- vycházel např. z Michella, Bollen, Molenaara
- nejprve formuloval striktní koncepci validity (cca 2004), pak rozpoznal, že je obtížné dle ní provádět validizaci (cca 2013), načež přešel k síťovým modelům (cca 2020). Nyní buduje psychologické teorie



Co si odnášíte?

Take home messages (za nás)

1. Existuje více pojetí validity. Zde se věnujeme tomu **realistickému**, protože je moderní a ve srovnání s konkurencí dobře vyargumentované.
2. Validita v realismu: (1) **atribut existuje** a zároveň má (2) **kauzální vztah** s pozorováními vzniklými na základě testu.
3. Validizace je ale obtížná, protože musí být založena na **obsahové validitě, validitě odpověďových procesů a statistickém modelování**.
4. Obsahová validita je nutnou podmínkou pro smysluplnou konstrukci testu. V položkách nemá nic chybět, ani přebývat.
5. Celkově je **zásadní role psychologické teorie**.