

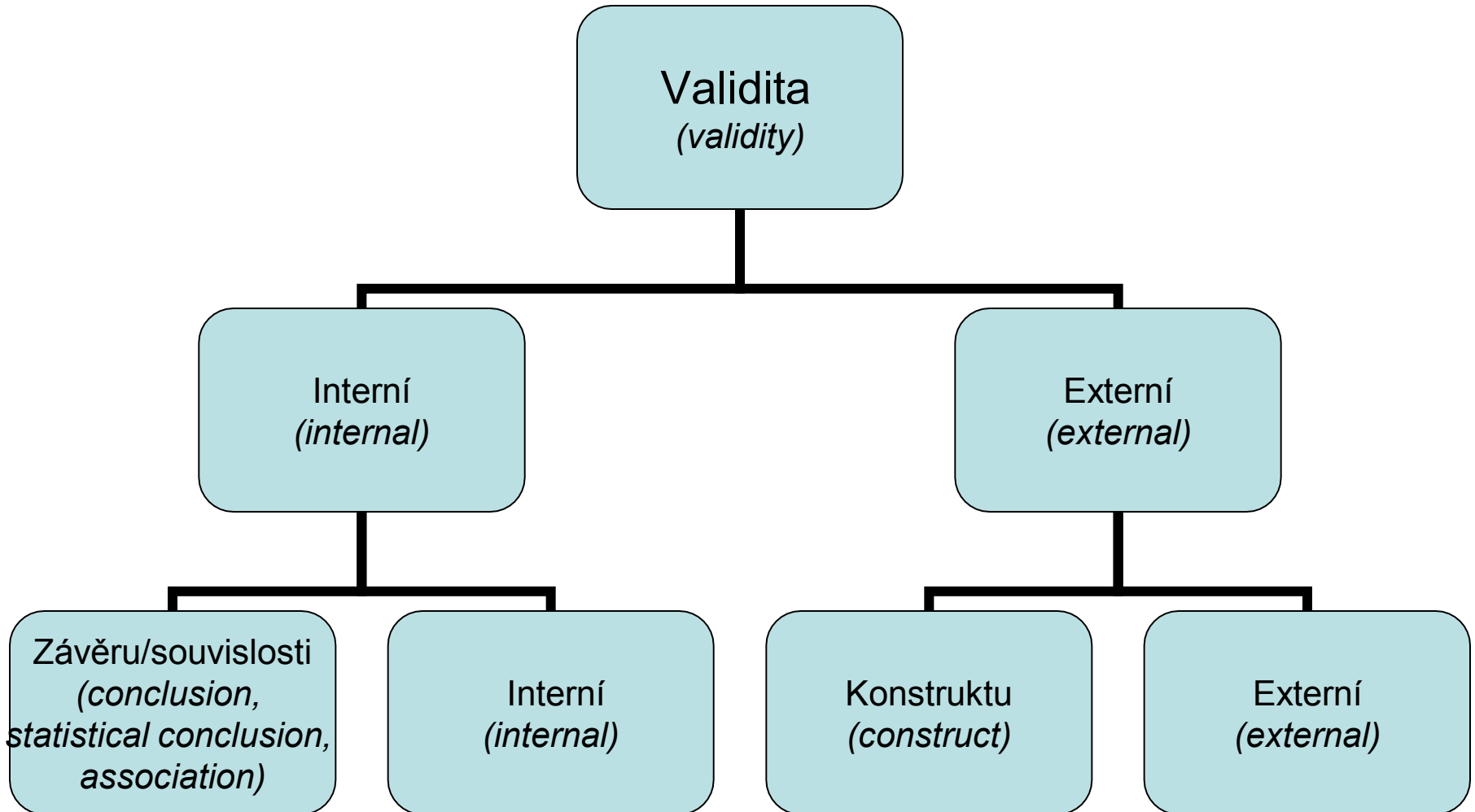
# „VALIDITA a RELIABILITA“

Nejen v sociálně-vědním výzkumu

M. Suchanec - díky Hagedaars 2006 and 2007, Campbell & Stanley 1966, Campbell & Cook 1979, Trochim 2006.

# Validita=platnost

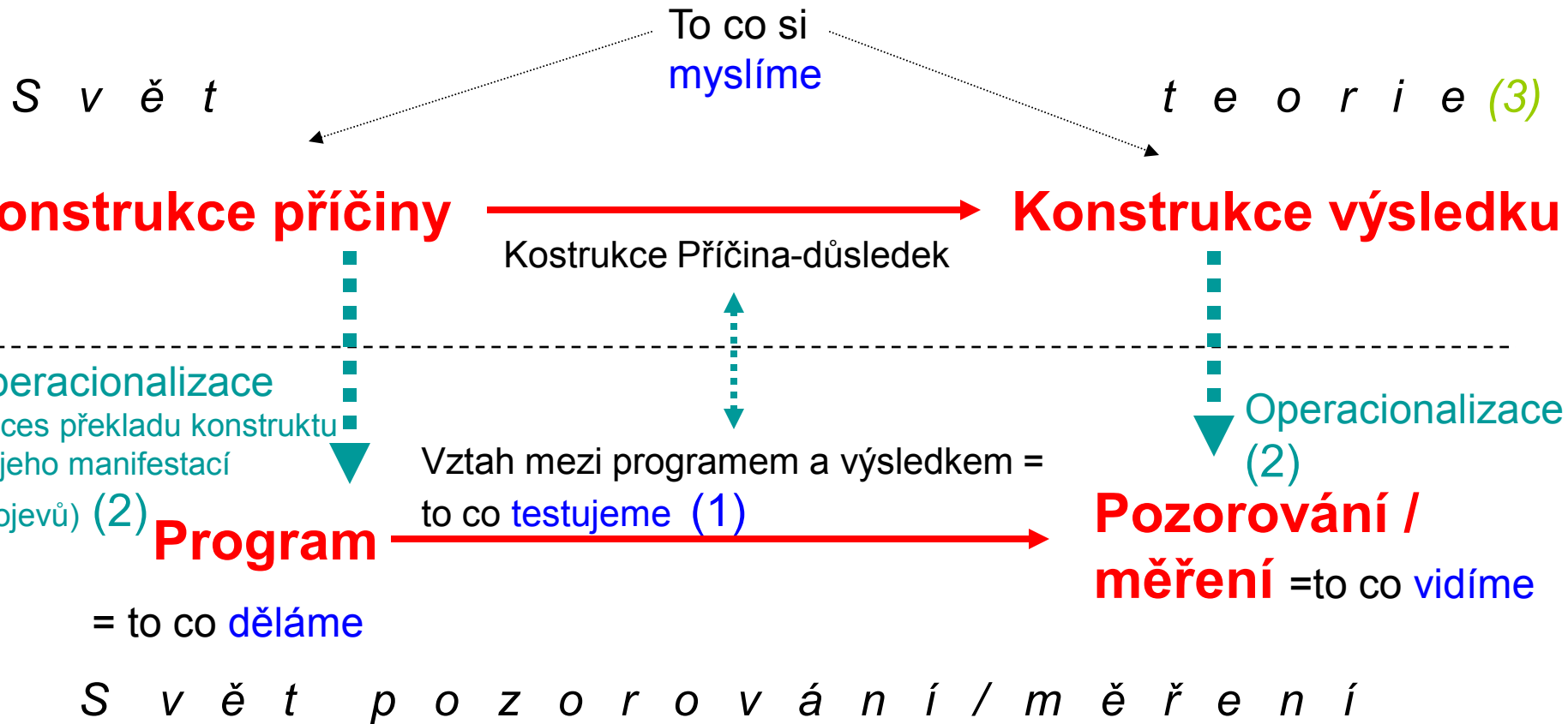
jaká, čeho?



# ....jaká, čeho?

- Závěrová = **Existuje vztah/souvislost** mezi dvěma proměnnými (např. programem a výsledkem)?
- Interní = Za předpokladu, že ano, **je vztah kauzální?**
- Konstruktová = Za předpokladu, že ano, **odpovídá program dobře naši konstrukci programu, a reflektuje naše měření dobře naši konstrukci měření** (Operacionalizovali jsme správně?)
- Externí = Za předpokladu, že ano, **můžeme zobecnit tento vztah na jiná nastavení?** (osoby, místa, čas)

# Validita ve vztahu k dvěma světům – teorie a měření



(1) Závěrová + interní = svět pozorování / měření

(2) Konstruktová = spojení mezi teorií a pozorovaným světem

(3) Externí = rozsah naší teorie

# Validita je kumulativní

**Kumulativnost** Každá otázka předpokládá kladnou odpověď na předchozí otázky

**Ilustrace=schody** každý je postavený na předchozích

„Olymp“ = dosažení komplexní validity

**Externí** = můžeme zobecňovat na jiné podmínky? (osoby, místa, čas)

**Konstruktivní** = můžeme zobecňovat zjištění na konstrukty?

**Interní** = je vztah kauzální?

**Závěrová** = existuje vztah mezi programem a výsledkem?



# Závěrová validita

Závěrová **vs.** interní – pouze vztah **vs.** je vztah kauzální?

4 možnosti a 2 chyby našeho rozhodnutí o vztahu:

vztah existuje a my ho najdeme - správně

vztah existuje a my ho nenajdeme – chyba!

vztah neexistuje a my ho nenajdeme -správně

vztah neexistuje a my ho najdeme – chyba!

„Hledání jehly v kupce sena“ – dva problémy: maličká jehla a příliš mnoho sena - otázka poměru „Signál : Šum“

**Zdroje signálu**

1. Opravdová síla vztahu

**Zdroje šumu (=hrozby závěrové validity):**

1. Nízká spolehlivost nástroje měření/testu (špatná konstrukce dotazníku)
2. Nízká spolehlivost implementace programu (program je implementován v rozporu s teorií)
3. Náhodné irelevance v prostředí (ruch z ulice)
4. Náhodná heterogenita respondentů (každý pes jiná ves)

**Nízká statistická síla testu**

související komponenty: 1. velikost vzorku 2. velikost efektu 3. alfa úroveň 4. síla testu

Tzv. „**rybaření**“ = snaha manipulovat výsledky tím, že opakovaně analyzujeme data za lehce změněných podmínek

# Interní validita

- Existuje **alternativní příčina** výsledku?

Školní program  
na zvýšení  
matematické  
gramotnosti žáků



Matematická  
gramotnost  
žáků

Žáci sledovali v  
TV pořad  
„Počítání s  
Mirkem“

# Hrozby interní validity 1

## 1. Historie – událost mezi pre a post testem

př. účastník APZ souběžně absolvoval i jiný program, který by zvýšil jeho šanci na nalezení zaměstnání, popřípadě by například získal práci přes svého známého.

## 2. Zrání – přirozené procesy (stárnutí, růst)

Př. Účastník APZ se právě nachází v období, kdy je již obvyklé, že si uchazeč najde práci, a tudíž by si ji našel i bez programu. Alternativně, uchazeč se postupem času může stávat zkušenější v hledání práce.

## 3. Testování – pre-test ovlivňuje post-test

př. V případě, že je test stejný, může se účastník test naučit, popřípadě první test ho může motivovat k tomu, aby v druhém testu dosáhl lepšího výsledku.



## Hrozby interní validity 2

4. Instrumentace – změna v nástroji měření nebo výzkumníka (struktura dotazníku, změna pozorovatele)  
př. změna lektorů, kdy jejich hodnocení bude ovlivněno jejich subjektivními a vzájemně se lišícími metodami.
  
5. Úmrtnost – nenáhodné „odpadávání“ ze studie  
př. Nemotivovaní opustí program a míra motivace zároveň ovlivňuje úspěšnost na trhu práce

## Hrozby interní validity 3

6. Regrese k průměru – nenáhodný vzorek který podrobíme programu (např. žáci s nejhoršími výsledky) bude v post-testu blíže k průměru než v pre-testu (podmínkou je normální distribuce populace)

7. Selektce – kontrolní vs. programová skupina nejsou stejné a tyto rozdíly zahrnují faktory jež ovlivňují výsledek

př. zdroje selektce: a) absence randomizačního procesu, b) samovýběr, c) existence určitých výběrových kritérií pro přijetí do programu, která souvisí s úspěšností programu, (např. výběr nejvíce znevýhodněných nezaměstnaných).

# Ilustrace hrozeb interní validity



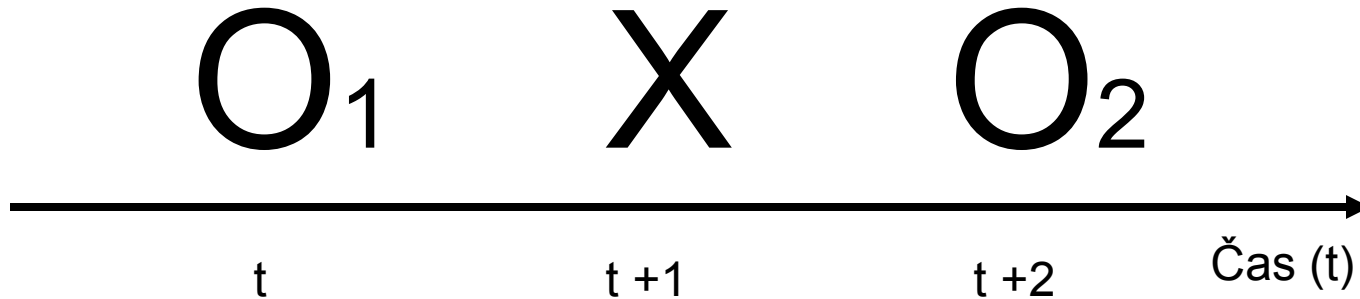
X = vystavení / expozice programu / události / léčbě / opatření (*exposure*)

O = měření / pozorování efektu expozice (*observation*)

Př. Učitel učí své žáky matematiku a po pololetí je vyzkouší a zjistí že matematiku umí ....usoudí, že je dobrý učitel.....

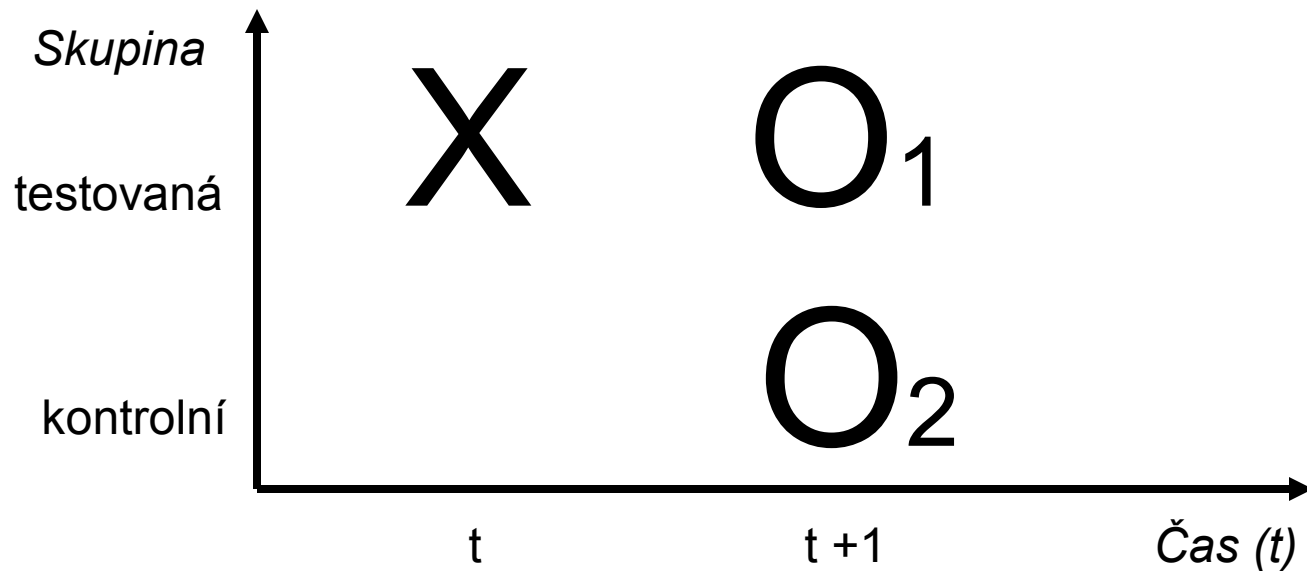
Co když ale žáci uměli matematiku už předtím než je začal učit?

# Ilustrace hrozeb interní validity



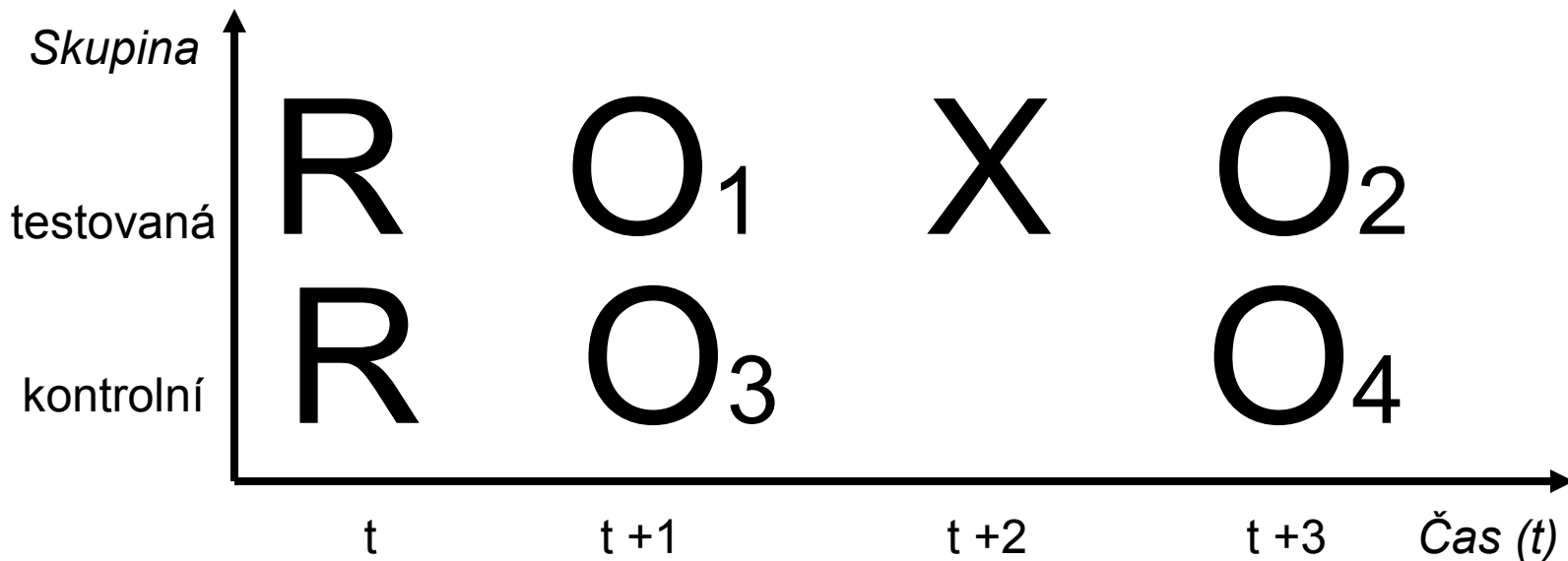
- ...je zapotřebí zjistit (O1) na jaké úrovni byli jeho dítka předtím než je začal učit nebo (viz další slajd)...
- ....ale co když mezi testy (O1 a O2) děti sledovaly pořad „počítáme s Mirkem“? = hrozba historie!
- ....a co když se děti prostřednictvím testu O1 naučili lépe zvládnout stejný test O2 = hrozba testování!
- ....a co když test O1 byl méně náročný než test O2 = hrozba instrumentace!

# Ilustrace hrozeb interní validity



- ....porovnat výsledky svých žáků s výsledky žáků které matematiku neučil.
- ...Co když ale jeho žáci mají pod čepicí a jsou z výběrové třídy, zatímco žáci které neučil jsou z poloviny diskalkulici? = hrozba **selekce!**

# Ilustrace hrozeb interní validity



...nahodile (např. kostka) přiřadíme žáky do obou skupin tak aby jejich charakteristiky byly náhodně rozptýlené 😊

X = vystavení / expozice programu / události / léčbě / opatření (*exposure*)

O = měření / pozorování efektu expozice (*observation*)

R = nahodile (např. kostka) přiřadíme žáky do obou skupin (*randomization*)

# Konstruktová validita

- Operacionalizovali jsme správně?
  - a) přeložili jsme správně naše myšlenky nebo teorie do našeho programu?
  - b) operacionalizovali jsme správně náš konstrukt do jeho manifestací (projevů)?

# Konstruktová validita

- Validita „překladu“
  - Zjevná (face) validita
  - Obsahová validita
- Validita ve vztahu k vnějššímu kritériu (criterion-related)
  - Prediktivní validita
  - Konkurenční validita
  - Souběžná (convergent) validita
  - Rozlišující, též „založená na členství ve známé skupině (discriminant)



- Validita „překladu“
  - Jak „přesně“ jsme přeložili konstrukt
  - Správnost operacionalizace na základě teoretické „reflexe“ – teoretická definice konstruktů vs. jeho operacionalizace

## – Zjevná (face) validita

- Nejslabší způsob posouzení konstruktové validity
- Porovnám operační definice (test, obsah aktivit programu) s jeho nálepkou („matematická schopnost“, „program rekvalifikace“)

## – Obsahová validita

- Stanovíme kritéria naplnění programu př. „program prevence těhotenství mladistvých“ (cílová skupina, preventivní vs. léčebný, obsahová kritéria: základní informace o těhotenství a prevenci)
- Pouze programy naplňující kritéria jsou obsahově validní

- Validita ve vztahu k vnějššímu kritériu (criterion-related)
  - „Chová se“ operacionalizovaný konstrukt v souladu s teorií?
  - rozdíl obsahová vs. kriteriální
    - Obsahová : srovnává konstrukt vs. konstrukt
    - Kriteriální: srovnává konstrukt vs. výkon

## – Prediktivní validita

- Porovnává předpověď s výsledky
- Schopnost predikovat něco co by měl být schopen predikovat
- Např. náš test matematické schopnosti je validní, pokud žáci, kteří na něm dopadli dobře mají i dobré výsledky ve škole

## – Konkurenční, též „validita založená na členství ve známé skupině“

- Schopnost rozlišit lidi, které by test měl být schopen rozlišit
- Např. náš test posouzení maniodeprese by měl od sebe oddělit lidi s maniodepresí a schizofreniky
- Test „rasisimu“ v rasistických uskupeních by měl vykazovat jiné výsledky než stejný test v obecné populaci

- Souběžná (convergent) validita
  - Mezi výsledky našeho testu aritmetických schopností a jiného testu stejných schopností by měla být vysoká korelace
- Rozlišující (discriminant)
  - Mezi výsledky našeho testu aritmetických schopností a cizího testu verbálních schopností by měla být nízká korelace

# Konstruktová validita

## Hrozby

1. Špatná definice významu konstruktů
2. Použití jediné verze programu v jednom čase místě zabraňuje zachycení plné šíře konceptu programu
3. Použitím jediného způsobu měření „sebedůvěry“ nelze dokázat že opravdu měříme „sebedůvěru“
4. Skupina podstupující program se zároveň účastní jiných programů
5. Efekt programu zahrnuje efekt testování
6. Zmatečné označení konstruktů (lék namísto určitá dávka léku)

# Externí validita

Můžeme zobecňovat na jiná nastavení?  
(lidé, čas, místo)

## hrozby

1. nenáhodný vzorek
2. lokální, časová specifika

# Reliabilita

=

spolehlivost

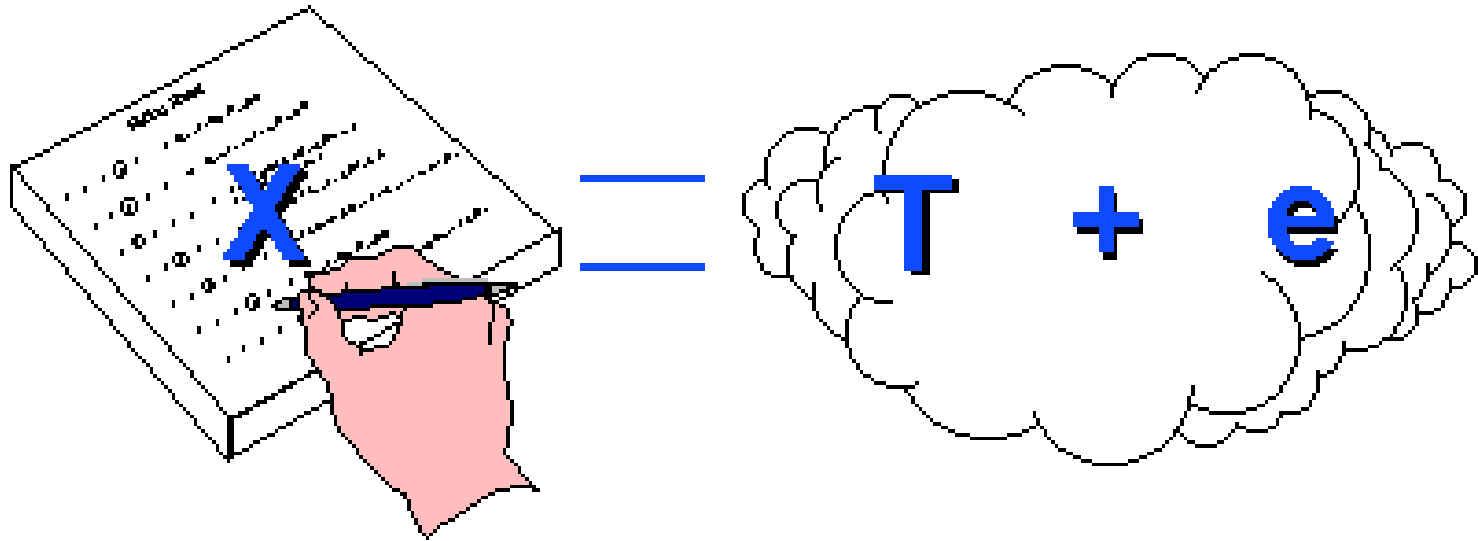
opakovatelnost

konzistence



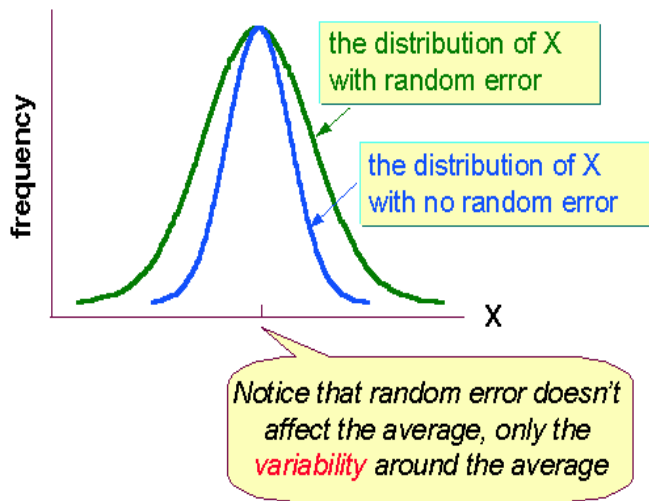
# Teorie klasického testu

$$\text{Observed Score} = \text{True Ability} + \text{Random Error}$$



$$\text{var}(X) = \text{var}(T) + \text{var}(e)$$

$$\text{Reliabilita } (0,1) = \text{var}(T) / \text{var}(X)$$

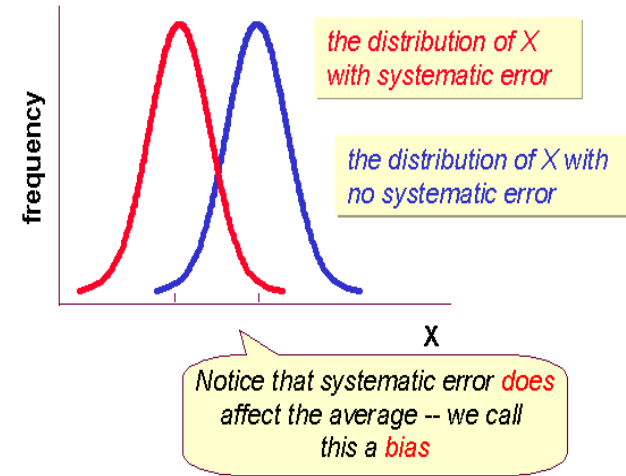


$$X = T + e$$

Two Components:

- $e_r$  • Random Error
- $e_s$  • Systematic Error

$$X = T + e_r + e_s$$



## Náhodná chyba:

Při statistickém testu se někteří studenti cítí špatně (pili alkohol, nespali), někteří dobře (strávili den toulkou v přírodě se svou láskou) – **průměrný výsledek vzorku se namění**

### Jak chyby eliminovat?

- 1) Pilotovat výzkumný nástroj
- 2) Pokud sběr dat probíhá pomocí rozhovorů nebo pozorování, trénovat výzkumníky
- 3) Pozorně přepisovat data do počítače
- 4) Použít statistické procedury
- 5) Triangulovat (použít více měřících a analytických technik)

## Systematická chyba:

Při testu je ruch z ulice, pořád někdo chodí do místnosti – ovlivněno jedním směrem (negativně) je více studentů – **průměr vzorku se mění**

# Odhad reliability

protože neznáme var (T), lze reliability jen odhadnout - pomocí korelace:

$$\text{Kovariance (X1, X2) / sd(X1) * sd(X2)}$$

Nalevo: sdílený rozptyl mezi X1 a X2 indikuje var(T) neboť T je to jediné co X1 a X2 sdílejí (chyby jsou různé)

Napravo: měříme-li totéž jen s mírným zpožděním, očekáváme že  $\text{sd}(X1)=\text{sd}(X2)$ , potom  $\text{sd}(X1)*\text{sd}(X2) = \text{sd}(X1)^2 = \text{var}(X)$

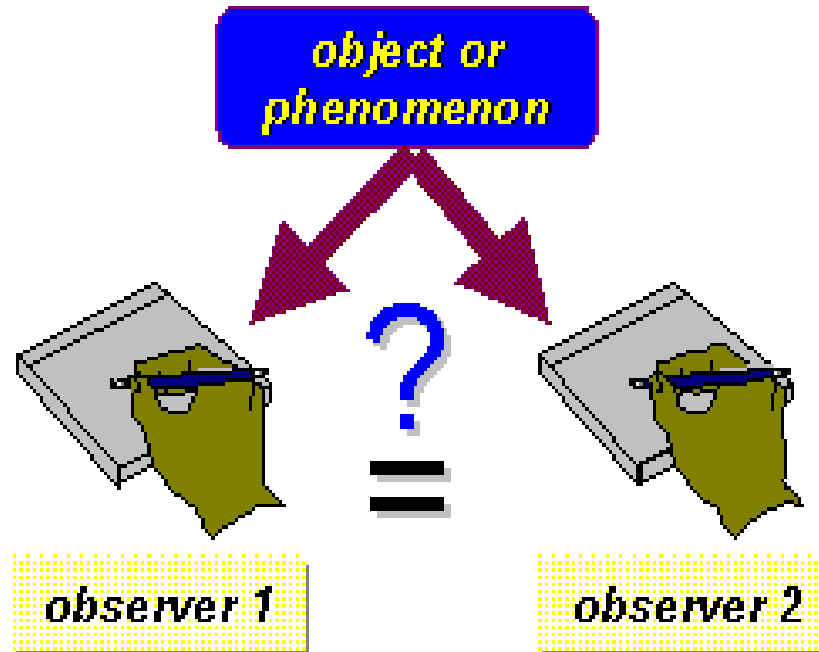
$$\begin{aligned} \text{Tedy (X1, X2) / sd(X1) * sd(X2)} \\ = \text{var (T) / var (X)} \end{aligned}$$

korelace mezi dvěma pozorováními stejného měření je odhadem reliability

# Typy reliability

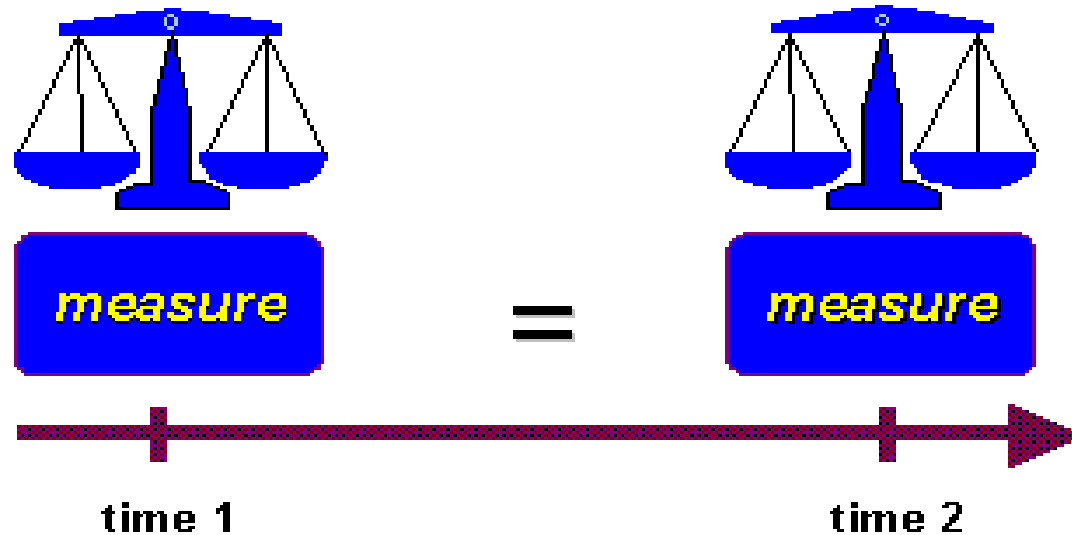
- **1) Konzistence mezi výzkumníky**
  - Míra shody mezi různými výzkumníky pozorujícími stejnou věc
- **2) test-retest reliability**
  - Konzistence mezi dvěma měřeními stejné věci (stejnými testy) v různých časech
- **3) Parallel-Forms Reliability**
  - Konzistence mezi výsledky dvou paralelních testů konstruovaných stejným způsobem
- **4) Interní konzistence testu**
  - konzistence výsledků mezi dvěma polovinami stejného testu
    - 4a) split-half reliability – reliability dvou půlek
    - 4b) Reliability Cronbachovy alfy

# Konzistence mezi výzkumníky



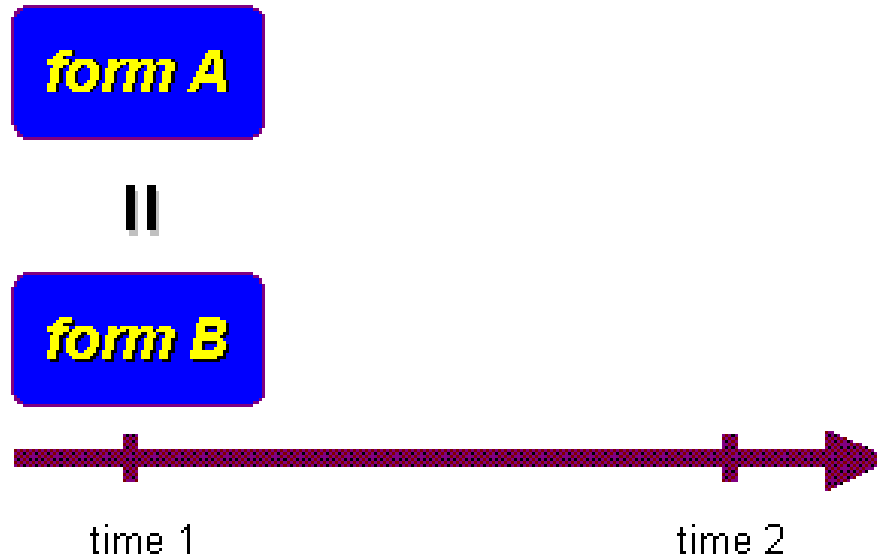
- Lidé = nekonzistentní, rozptýlení, unavení, sní, dezinterpretují
- 2 způsoby
  - Nominální proměnné – procentuální shoda
  - Spojité proměnné - korelace
- Zvýšení konzistence – trénink
- Předpoklad: záznam (test/pozorování) vyplňují výzkumníci

# test-retest reliabilita



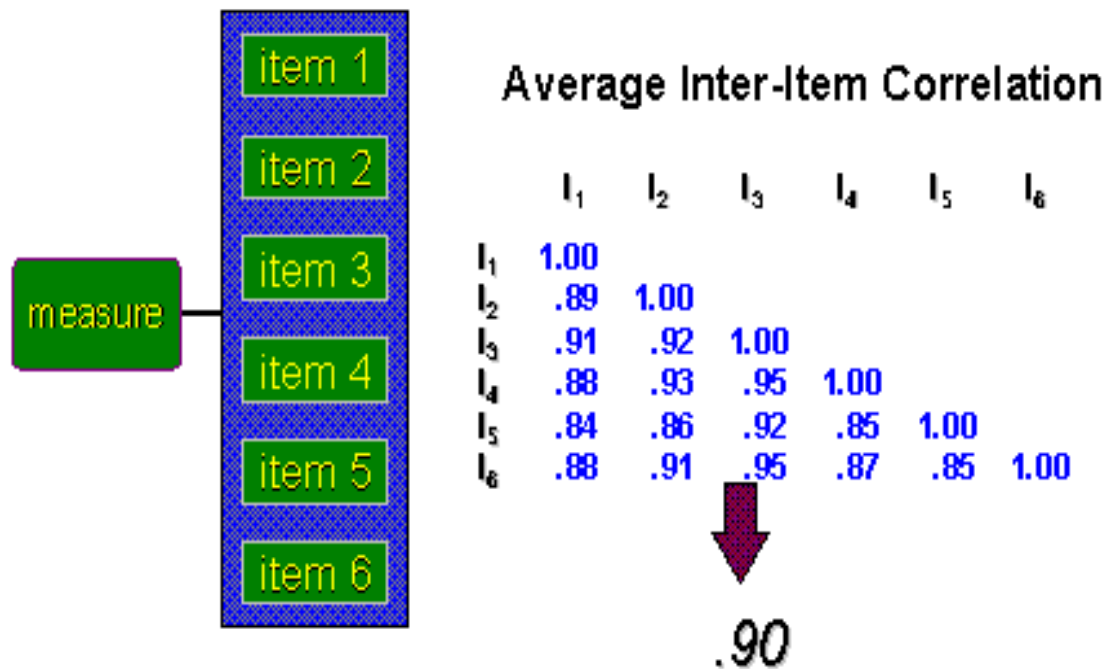
- Stejný test stejnému vzorku ve 2 různých časech
- Čím větší prodleva, tím nižší korelace (čím menší prodleva více podobné faktory způsobují chybu v měření)
- Předpoklad: krátká časová prodleva

# Reliabilita paralelních forem testu



- Vytvořím velkou sadu otázek posuzující stejný konstrukt (např. spokojenost)
- náhodně rozdělím na dvě sady
- Každou sadu dám stejnému vzorku
- Korelace mezi dvěma sadami je odhad reliability

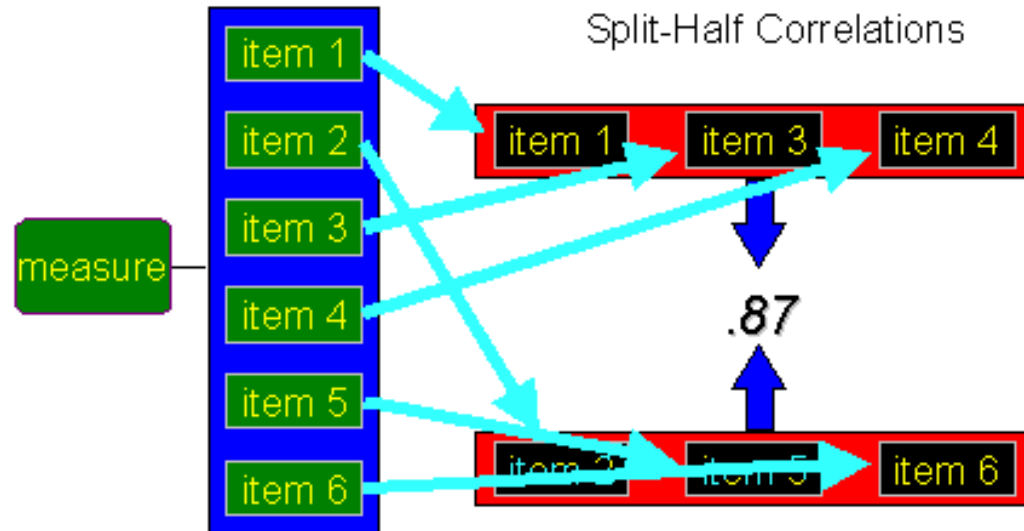
# Reliabilita mezi jednotlivými otázkami testu (inter-item rel.)



- Spočítáme korelaci mezi každým párem otázek
- 6 otázek = 15 párů = 15 korelací
- Průměr 15 korelací je odhadem reliability

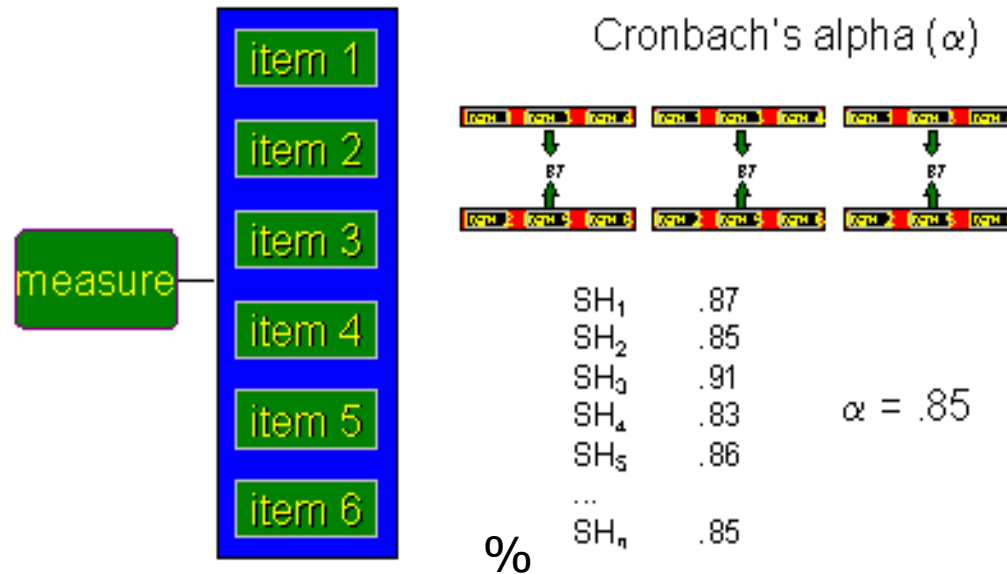


# Split-half reliability – reliability dvou půlek



- Náhodně rozdělíme všechny otázky které mají měřit stejný konstrukt do dvou sad
- Výsledná korelace mezi dvěma sadami je odhadem reliability

# Cronbachova alfa



- Průměr všech korelací mezi všemi možnými kombinacemi dvou půlek testu