

# Normy a standardizace testu

---

PSYB2590: ZÁKLADY PSYCHOMETRIKY | PŘEDNÁŠKA 5

11. 4. 2022 | HYNEK CÍGLER

„Norma: od slova normální, běžný“.

Co to znamená běžný?

V kontextu psychometrie.

**10**  
Joštova

**218**  
Brno-město

2019

**VYMÝVÁNÍ MOZKŮ**

**pondělí - pátek: 8 - 16 hod.**

**(Individuální zákroky po dohodě s rektorem)**

FAKULTA SOCIÁLNÍCH  
STUDIÍ

2020

MASARYKOVA UNIVERZITA

NAŠE ŠKOLA SE PYŠNÍ  
PRŮMĚRNÝM STUDENTSKÝM

**IQ: 63,15 %**

FAKULTA SOCIÁLNÍCH  
STUDIÍ

2021



2022

?  
?  
?  
?  
?  
?  
?  
?  
?  
?  
? ?



# Standardizace testu?

---

Soubor veškerých postupů, které slouží jako podklad a důkazy pro *standardní* rozhodování o jednotlivcích na základě testových metod.

- Proces tvorby podkladů.
- Proces dokazování, že tyto podklady jsou validní.

Pojem standardizace je proto poměrně široký a zahrnuje:

- Důkazy validity, reliability.
- Vytvoření norem, standardních skóru.
- Kodifikace postupů pro *standardní administraci, skórování a interpretaci skóru*.
- ...

# Standardizace testu

---

Tomáš Urbánek ([2010](#)) a další: 3 pojetí standardizace.

- **I. povrchní:** „... **metoda je přesně popsána** [...] jak má vypadat např. testový materiál, pomůcky nebo testový sešit, na jakém papíře, v jakých barvách a jakým písmem apod. Kromě toho je jasně definováno, **jak má být metoda používána**, tzn. komu, kým a za jakých podmínek smí být administrována, **jak má být vyhodnocována** a co znamenají získané výsledky.“
- **II. klamavé:** „... se spokojuje s existencí **jakýchkoli norem** ve smyslu popisu, jakých výsledků dosahují respondenti z nějakých jasně definovaných skupin. Ani tento požadavek není obtížné splnit, stačí jen použití metody spojit se sběrem dat a elementární statistickou prezentací výsledků...“
- **III. komplexní:** „... Součástí tohoto pojetí jsou i obě pojetí předchozí [...] je přinejmenším nutné prokázat, zda metoda měří daný atribut (**validita** a validizace) a s jakou přesností (**reliabilita**). Současně je nutno vyřešit **všechny speciální otázky**, které mohou nastat v souvislosti s testováním specifických charakteristik. To je pojetí uváděné např. ve Standardech pro pedagogické a psychologické testování (AERA, APA, NCME, 2001), ale je doporučováno i [[..., EFPA](#)].“



# Manuál diagnostického testu

---

**Teoretická východiska:** Co je měřeno, jaké jsou známé souvislosti, proč se to měří.

- Účel metody: Komu, kdy, proč, kým, kde...

**Postup administrace a skórování:** jak přesně se metoda zadává a skóruje.

- Tvorba hrubých skóru a převod na standardní/vážené skóry.

**Postup interpretace:** co výsledky znamenají.

- Součástí zpravidla i kazuistiky.

**Psychometrický manuál:** Dokládá výše uvedené na vzorku z cílové populace.

- Standardizační soubor, postupy konstrukce norem.
- Důkazy validity, reliability vzhledem k účelu metody.

Co z výše uvedeného je možné pouze přeložit ze zahraniční verze testu?

# Tvorba testu

---

Značné rozdíly mezi metodou určenou pro výzkum a pro individuální diagnostiku.

Tvorba nové testové metody pro praktické účely.

- + Kulturně adekvátní metoda.
- – Tvorba může selhat, vysoké nároky na přípravu...
- – Vysoké finanční náklady na průběžné pilotáže, analýzy...

Adaptace zahraniční metody pro praktické účely.

- Překlad vs. adaptace.
- + Zpravidla ověřená metoda, nižší nároky na velikost vzorku, pilotáže, méně práce.
- + Lze využít zahraniční důkazy validity, většinou rozsáhlejší teorie.
- – Cena licence (i několik milionů Kč), často časově omezená, poplatky...
- – Standardizační studie stejně musí být realizována.

# Design standardizační studie 1

---

Volba výběrové populace (pro koho je test určený)?

- Mezinárodní, národní, lokální, místní...

Kognitivní pilotáž/e (kvalitativní metodologie).

Kvantitativní pilotáž/e (podobné otázky, jako vlastní standardizace níže).

Způsob výběru vzorku a administrátorů.

- Náhodný, stratifikovaný, clusterový, příležitostný... Plánovaně chybějící data.
- Tvorba adekvátních clusterovacích proměnných (ČSÚ).
- Inkluzivní a exkluzivní kritéria.
- Zaškolení a výběr administrátorů.

Sběr dat.

- Párování respondentů s administrátory. Jak zajistit jejich ortogonalitu?
- Kódování dat.

# Design standardizační studie 2

---

Přepis dat, kontrola správnosti.

- Vyčištění dat, spárování datasetů atd.

Vážení respondentů?

- Clusterový/stratifikovaný výběr. Bude váženo vše/nic?

Položkové analýzy, analýzy reliability, validity.

- Uvnitř či napříč kohortami?
- Vyřazení položek, úpravy obsahu testu.

Tvorba norem.

- Vytvoření vyhodnocovacího softwaru, normalizačních tabulek...

Zkompletování manuálu.

Prodej. Spotřební materiál volně dostupný?

# Příklad nákladů: BACH

---

Školní dovednosti, cca 30 subtestů (čeština a matematika), cílová populace 5–22 + 55–80 let.

Konormace s WJ-IV, resp. TOMAL-SE; 2,5 roku vývoje, 3+1 rok standardizace v rámci projektu.

Pilotáže a vývoj (velmi hrubě): **1.100.000 Kč**

- Náklady na vývoj testu: odhad 500.000 Kč.
  - Tahle částka reálně nebyla vyplacena.
  - Zčásti je „sanována“ náklady na jiné personální náklady v rámci projektu.
- 500 individuálních administrací: 400.000 Kč.
- přepis dat: 50.000 Kč.
- tisk, poštovné: 150.000 Kč

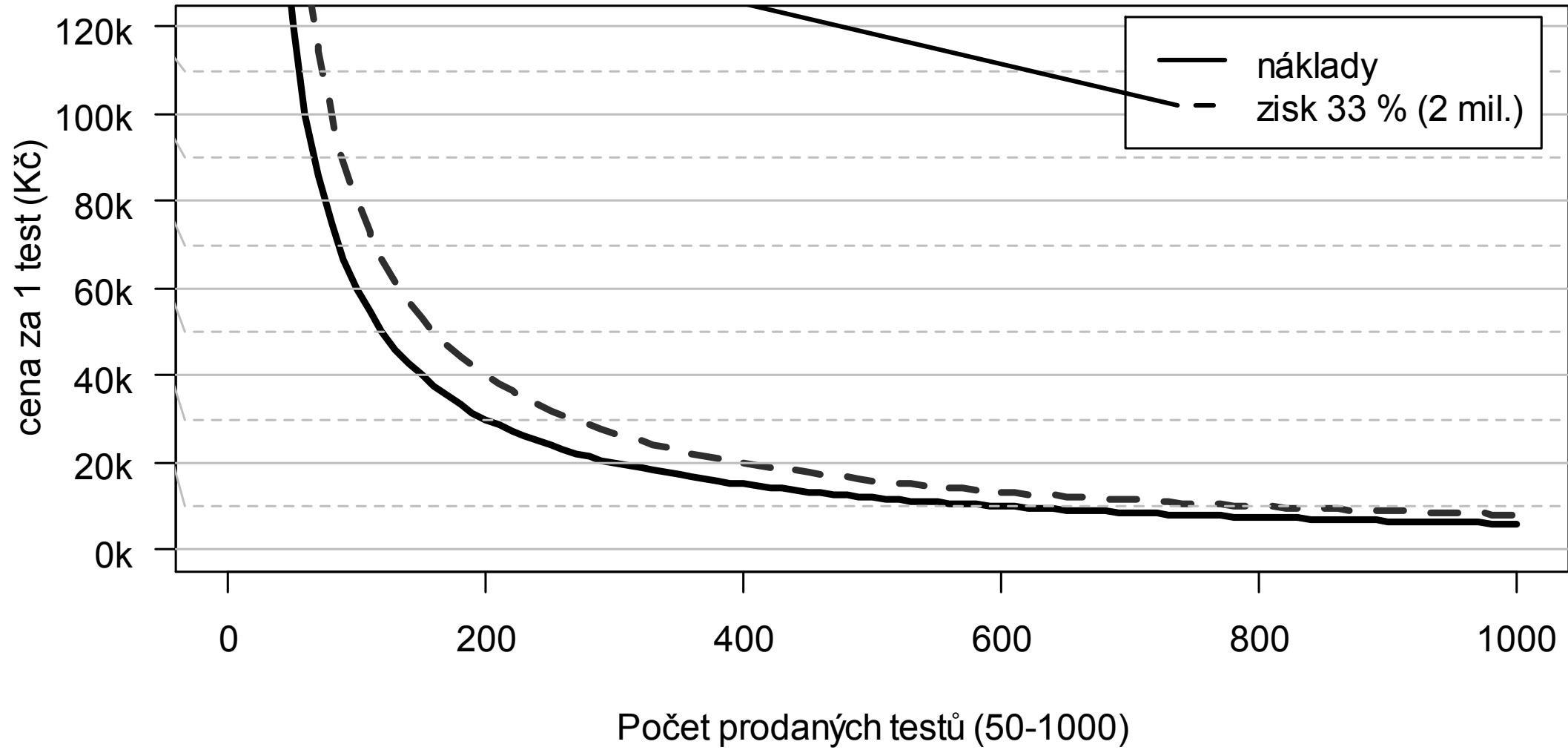
Standardizace (TAČR): **5.346.133 Kč**

- Sběr dat: přes 2.000.000 Kč
  - (navýšeno v průběhu projektu, částka není jistá).
- Odměna respondentům: 140.000 Kč
- Přepis dat: 170.000 Kč
- Školení (lektoři): 50.000 Kč
- Tisk, IT, nahrávací studio, grafika apod.: 440.000 Kč
- Poštovné: 80.000 Kč
- Personální náklady jiné: 1.000.000 Kč.
- Režie, nájmy...: 800.000 Kč

Celkové náklady: minimálně **6 milionů Kč.**

- (V personálních nákladech se zčásti překrývá pilotáž a standardizace).

# Cena 1 testu podle množství prodeje



# Příklad nákladů: BACH

Jsou ty náklady skutečně tak vysoké?

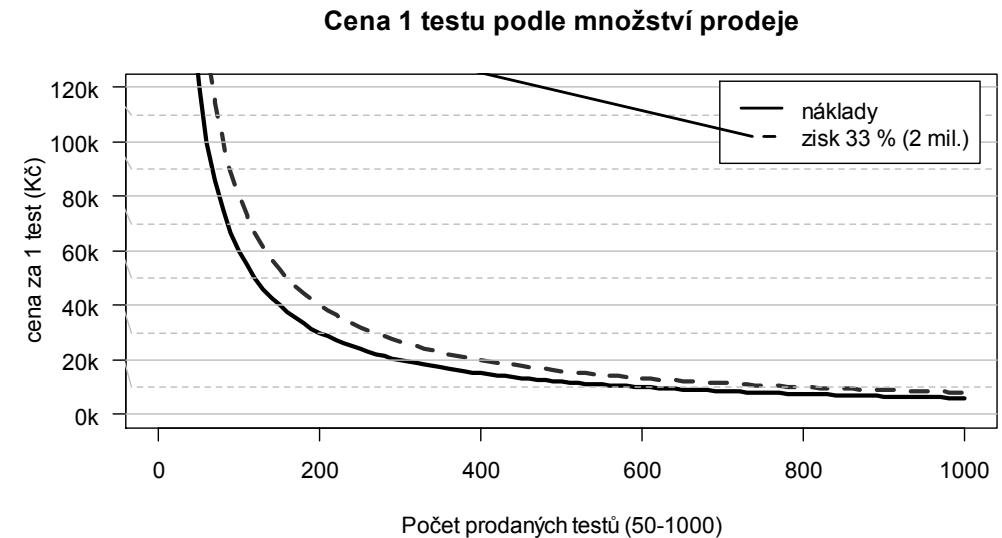
- Příklad: prodá se 200 ks testu → nákupní cena 40.000 Kč.
- Test bude aktuální 10 let, 80 vyšetření/rok.

Náklady na jednoho klienta: **50 Kč.**

- Plus cca **50 Kč** spotřební materiál.

Srovnejte s personálními náklady:

- Průměrná měsíční mzda v PPP (2021): 44.200 Kč → 280 Kč/hodina
- Náklady zaměstnavatele: cca **350 Kč/h.**
- Při délce vyšetření 4 hodiny (včetně psaní zpráv): **1400 Kč/vyšetření.**



## Standard 9.22

Test users have the responsibility to respect test copyrights, including copyrights of tests that are administered via electronic devices.

„1. přikázání psychologické diagnostiky:“

**Nezkopíruješ!** 😊

AERA, APA, & NCME. (2014).

*Standards for Educational and Psychological Testing.*

Washington: American Educational Research Association.

## Standard 9.21

Test users have the responsibility to protect the security of tests, including that of previous editions.



# Standardy a zdroje

Doporučení a standardy  
pro vývoj testů

Guidelines

Recenzní model testu



INTERNATIONAL TEST COMMISSION

ITC Guidelines for Translating and Adapting Tests  
(Second Edition)

**Version 2.4**

# Krok 1: Vyvinout dobrý test

---

Downing, S.M., & Haladyna, T.M. [eds.] (2006). *Handbook of Test Development*. L. Erlbaum.

- Sborník/učebnice pro vývoj testů. Nejen psychometrika, ale i tipy pro vydavatele.

## **Další doporučení a guidelines pro vývoj testů:**

International Test Commission: <https://www.intestcom.org/page/28>

- Doporučení Mezinárodní komise pro testování – opravdu doporučuji. Není ale závazné pro ČR.

Případně ještě BPS [Guidelines on testing and test use](#).

- Další doporučení od britské psychologické společnosti. Rovněž není závazné pro ČR.

APA/AERA/NICM: **Standards for Psychological Testing** (CZ/EN verze)

ISO 10667-1, 10667-2: Norma pro testování v B2B prostředí

- [ISO 10667-1 Assessment service delivery : Procedures and methods to assess people in work and organizational setting. Part 1, Requirements for the client](#)
- [ISO 10667-2 Assessment service delivery : Procedures and methods to assess people in work and organizational setting. Part 2, Requirements for service providers](#)

# Krok 1: International Test Commission

---

Seznam doporučení:

- [The ITC Guidelines for Translating and Adapting Tests](#)
- [The ITC Guidelines on Test Use](#)
- [The ITC Guidelines on Computer-Based and Internet-delivered Testing](#)
- [The ITC Guidelines on Quality Control in Scoring, Test Analysis and Reporting of Test Scores](#)
- [The ITC Guidelines on the Security of Tests, Examinations, and Other Assessments](#)
- [The ITC Guidelines on Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal](#)
- [The ITC Guidelines for the Large-Scale Assessment of Linguistically and Culturally Diverse Populations](#)

# Krok 2: Zhodnocení kvality testu

---

Testy bývá zvykem recenzovat.

- USA: Burosův institut. <https://buros.org/mental-measurements-yearbook>
- EU: EFPA Review Model. <http://assessment.efpa.eu/>

V České republice časopis Testfórum: [www.testforum.cz](http://www.testforum.cz)

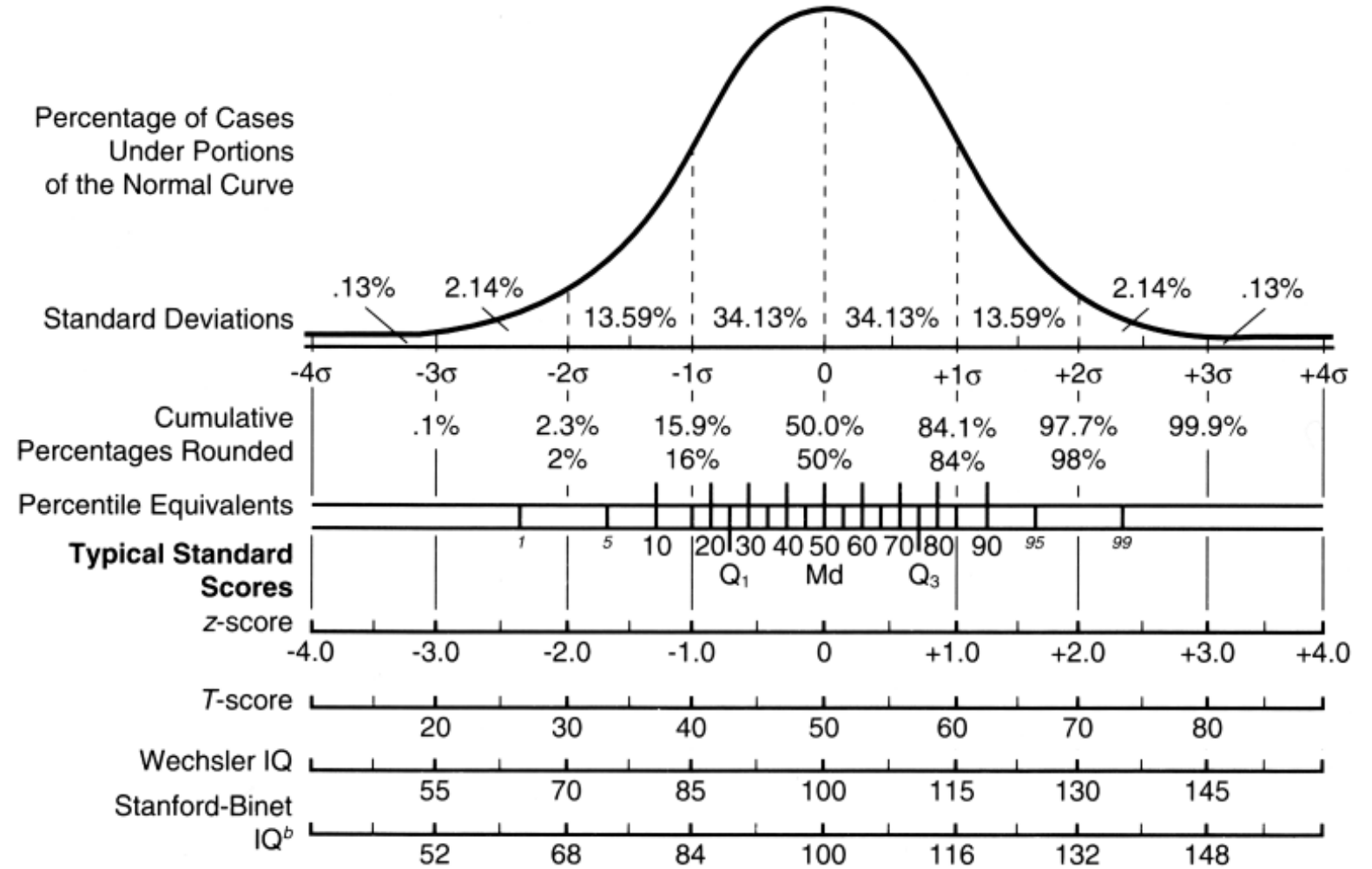
- Seznam dosud recenzovaných testů: <https://testforum.cz/recenze>

## TESTFÓRUM

Časopis pro psychologickou diagnostiku

# Normy

Percentage of Cases Under Portions of the Normal Curve



# K čemu jsou normy

---

Snaha vyhnout se **chybám intepretace**. Normy dávají smysl výsledkům testování:

- Porovnáním výsledku s výsledky populace;
- porovnáním výsledku s kritériem;
- porovnáním výsledků navzájem napříč testy.

Zamezení osobních chyb, svévolné intepretaci „čísel“.

Částečné „překonání“ problému měření v sociálních vědách.

- Normy jsou tím, co udává „škálu“ měření. Tvorba „jednotek“ (IQ apod.).
- Důsledkem je často neoprávněná reifikace výsledku měření
  - Např. ztotožnění IQ = inteligence.

# Proč vlastně normy?

---

Proč jsou normy v psychologii nezbytné?

I pokud by měření v psychologii bylo intervalové, není poměrové.

Neexistuje tedy jasně definovaný referenční bod.

- Referenční bod je nutné stanovit arbitrárně.

Je nutné zvolit i jednotku; typicky je závislá na vzorku (populaci).

- Na čem bývá založena jednotka např. ve fyzice? Historicky etalon.
- Proč je jednotka závislá na vzorku problém?
- Šlo by to řešit jinak?

# Typy norem

---

## „Klasické“ normy

- Mezinárodní, národní normy, místní normy.
- Nahodilé normy (více různých specifických populací v případě, že není dostupný reprezentativní vzorek).
- Uživatelské normy.
- Lokální normy, normy pro specifické populace.

**Referenční** (normy) vs. **kriteriální** (arbitrární kritérium) testování.

**Expektační tabulky** – odhady pravděpodobnostiho výskytu jevu, klinické odpovědi apod.

- Nepředpokládá náhodný vzorek, spíše vypárovaný oproti pacientům.
- Často v podobě grafu pravdivě vs. falešně pozitivní odpovědi. Podobné ROC analýze.

„**Typologie**“ – specifický příklad ipsativních skóru. Pozor na ně!

- Specifické nároky na data, na měřený atribut.
- Kontinuální rys by měl mít bimodální rozdělení.



# Druhy skóru<sup>1</sup>

---

## HRUBÉ SKÓRY

**Sumační indexy** – prostý součet položek.

- Nebo průměr, který má výhody i nevýhody.

**Lineární kombinace** – každá pol. má jinou váhu, např. na základě faktorové analýzy.

- Někdy též vážené nebo kompozitní skóry.
- Odhad faktorového skóru / faktorové skóry.

**Jiné; zejm. IRT odhady** (Analogie hrubých skóru v CTT – theta, EAP/MAP, W-skóry)

## ODVOZENÉ SKÓRY (VŠE OSTATNÍ)

**Percentilové skóry:** Percentily, decily, percentilové pořadí a další (kvantily, percentilové rozpětí...), steny, staniny...

**Standardní skóry:** IQ(100;15), T(50;10), T(500; 100), z-skóry, Wechslerovy vážené skóry W(10;3)...

**Vývojové skóry:** Mentální věk (age-equivalent score, grade-equivalent score), index relativní výkonnosti (RPI), zóna vývoje

**Ipsativní skóry**

<sup>1</sup> <http://prirucka.ujc.cas.cz/?slovo=skóre> ☺

# Standardní skóre

---

Lineární transformace hrubých skóre na odvozené. **Z-skór:**

$$z = \frac{X - \bar{X}}{\sigma_X}$$

**Standardní skór:**

$$S = \sigma_S \cdot z + \bar{S} = \frac{\sigma_S}{\sigma_X} (X - \bar{X}) + \bar{S}$$

- $S$  – standardní skór,  $\sigma_S$  – cílová SD,  $z$  – z-skór,  $\bar{S}$  – cílový průměr,  $\sigma_X$  – SD HS,  $\bar{X}$  – průměr HS,  $X$  – hrubý skór.

**Předpoklady:**

- Průměrně/přiměřeně obtížné položky a tedy i normální rozdělení hrubého skóru.
- Pokud předpoklad neplatí: nelineární transformace podle tabulky. Kde se vezme ta tabulka? 😊

# Normalizace rozložení (mírné zešikmení)

---

McCallova plošná standardizace.

- Každému  $X$  je přiřazeno percentilové pořadí.
- Percentilům je přiřazen T-skór za předpokladu normálního rozdělení.
- + teoreticky „dobré“ vyhlazení.
- - percentily jsou zatížené vysokou výběrovou chybou.

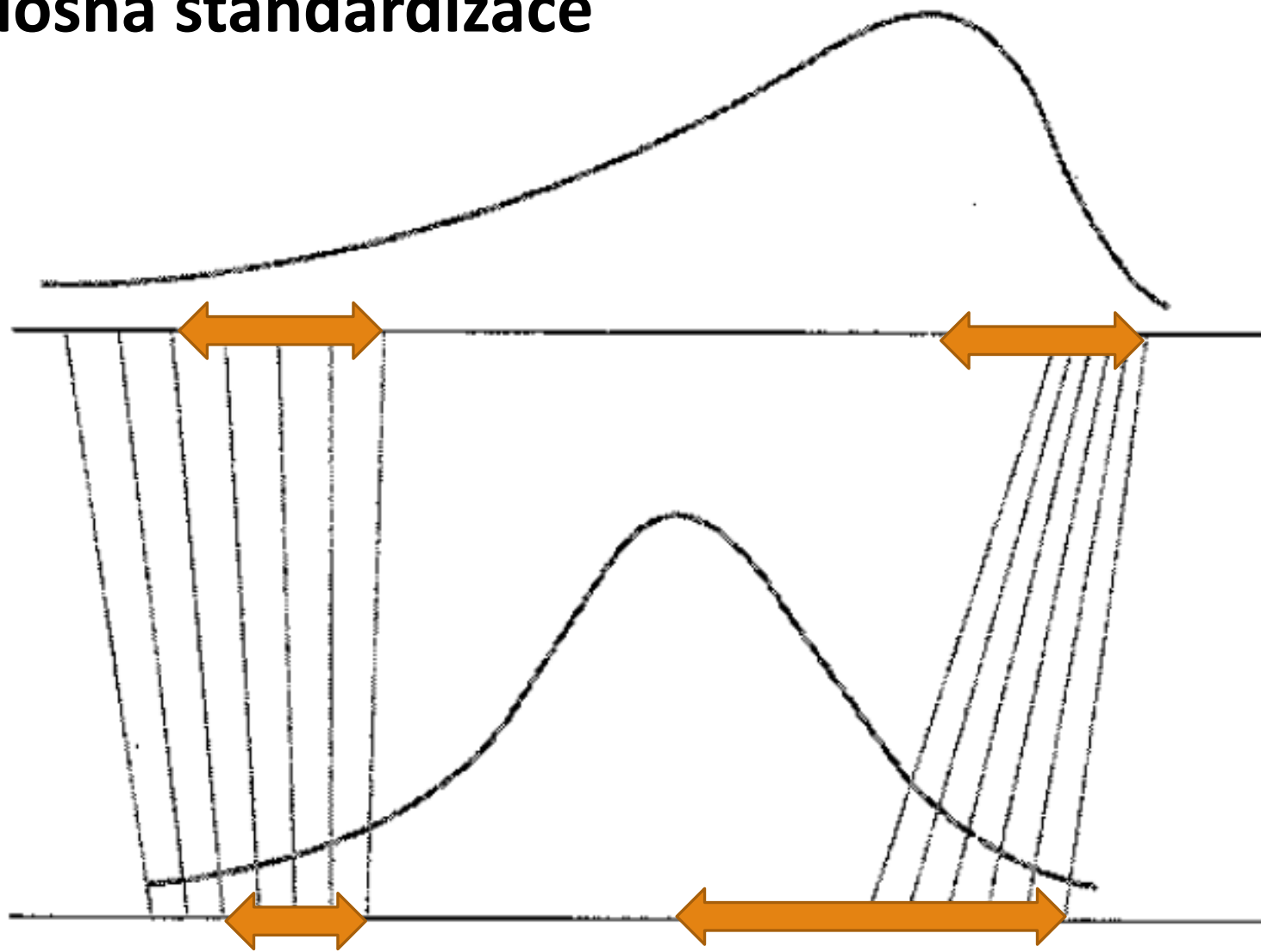
Normalizace podle mediánu.

- Samostatná SD pro lepší a horší respondenty.
- + odhad jen 3 parametrů ( $M$ ,  $SD_{\text{lower}}$ ,  $SD_{\text{upper}}$ ), menší výběrová chyba.
- - slabší vyhlazení, obtíže s konstrukcí CI.

Jiné nelineární transformace včetně kontinuálního normování.

**Transformuje se nejen skór, ale i jeho SE/CI!**

# McCallova plošná standardizace



# Normalizace podle mediánu

---

Nejjednodušší způsob normalizace skóre.

Předpoklady:

- Normální rozdělení má průměr (přibližně) shodný s mediánem.
- Předpokládáme, že každá polovina rozložení sama o sobě odpovídá přibližně normálnímu nezešikmenému rozložení, jen s jinými parametry.

Postup:

- 1. Rozdělíme respondenty na dvě poloviny podle mediánu.
- 2. Ručně spočítáme SD horní a dolní poloviny.
  - Nejde o SD uvnitř poloviny, ale odhad SD napříč polovinami, když by druhá polovina měla stejné, avšak zrcadlově otočené rozložení.
- 3. SD použijeme zvláště pro výpočet SE a SS v obou polovinách.
  - Co s přechody přes medián a SE, CI?

Např. Woodcock-Johnson IV US (výhodné při vyhlazování skóre).

# Extrémní zešikmení = problém

---

Výrazný efekt stropu nebo podlahy.

- Velká komplikace – ideálně by žádný respondent neměl mít max. nebo min. skóre.
- V případě těžkého testu i získání jediného bodu hrubého skóre posune respondenta velmi výrazně na všech škálách (percentil, stand. skóre).

Extrémně snadný/obtížný test.

- Např. při měření patologie.

Neexistuje kontinuální latentní proměnná, ale kvalitativní latentní „třída“.

V těchto případech je standardní skóre nevhodné.

- Percentil nebo spíše kriteriální skórování.

# Doporučení ke standardním skórum

---

Veškeré skóry jsou zaokrouhleny na celá čísla (kromě z-skóru, ty na 2 desetiny).

APA doporučuje T-skóry; IQ skóry výhradně pro měření výkonu v kognitivních testech.

Se skórem je vždy reportována chyba, např. formou CI (doporučuje se 90%).

- Vyjma stenů a staninů.

Steny a Staniny jsou považovány za „rozpětí“, konstruovány jsou na základě plošné transformace.

- Steny  $N(5,5; 2)$ , staniny  $N(5; 2)$ .
- Spíše marginální použití.

# Percentily

---

Procento osob, které mají **horší** hrubé skóre než hrubé skóre daného člověka.

- U škál s malým množstvím možných skóru prakticky nejde dosáhnout percentilu 100.
- Percentilové pořadí (percentil rank) – **stejně nebo horší** hrubé skóre **než daný percentil**.
- U dlouhých škál je rozdíl zanedbatelný, u krátkých je potřeba vědět, s čím pracujeme.
  - V případě nespojitě proměnné (v psychologii prakticky vždy) se liší percentily a percentilové pořadí mírně liší.

Odhad většinou na základě pozorovaného rozložení a ne normální distribuční funkce.

- Naopak standardní skóry často založené na percentilu (viz McCallovu plošnou transformaci).
- Ale co chyba měření a výběrová chyba? Může vést k rozdílu percentilu a standardního skóre.
- Je zvykem „vyhladit“.



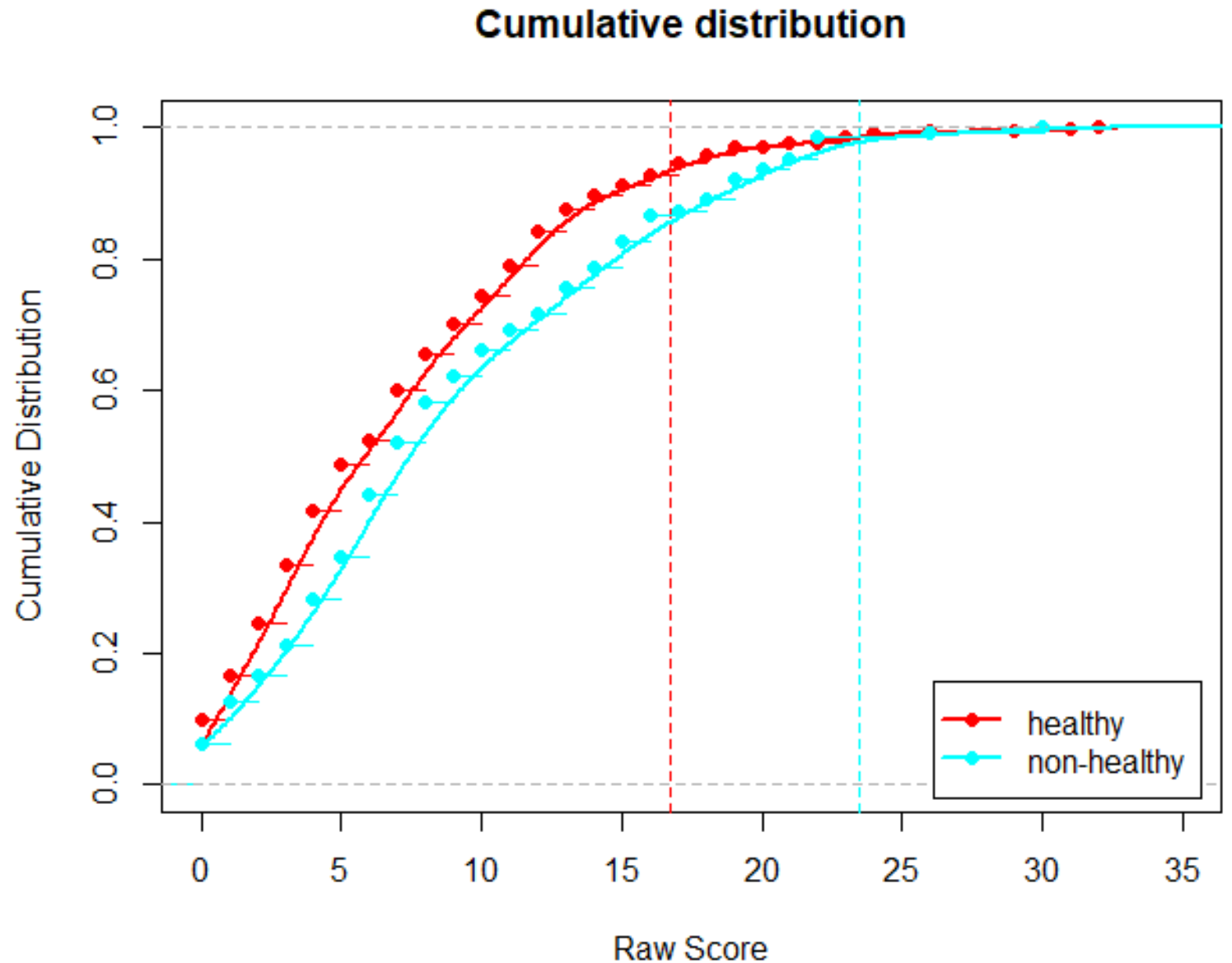
Příklad na vyhlazení  
percentilových norem

Beckův inventář depresivity  
(BDI)

- Svislé části označují dva kritické skóry.
- $N_h = 450$
- $N_n = 127$

balíček ks v R

- Kernel cumulative distribution



# Kontinuální normování

---

Dříve uvedené nelineární transformace a vyhlazení percentilu se označují jako **vertikální vyhlazení** testových skóru.

Existuje ještě **horizontální vyhlazení**, označované jako „**kontinuální normy**“.

Oboje slouží pro zpřesnění norem, zmenšení výběrové chyby či snížení požadavků na velikost vzorku se používá vyhlazení.

- **Horizontální vyhlazení** – *napříč* referenčními kategoriemi.
- **Vertikální vyhlazení** – *uvnitř* referenčních kategorií (viz percentily a SS dříve).

Princip: V případě 20 kategorií bychom potřebovali  $20 \times 2 (M+SD) = 40$  parametrů.

- Vyhlazení odhadne všech 20 průměrů s pomocí 3 parametrů. Ušetří se informace.
- Dílčí odchylky od reprezentativity souboru dostávají nižší význam.

# Vyhlazení norem (kontinuální normování)

---

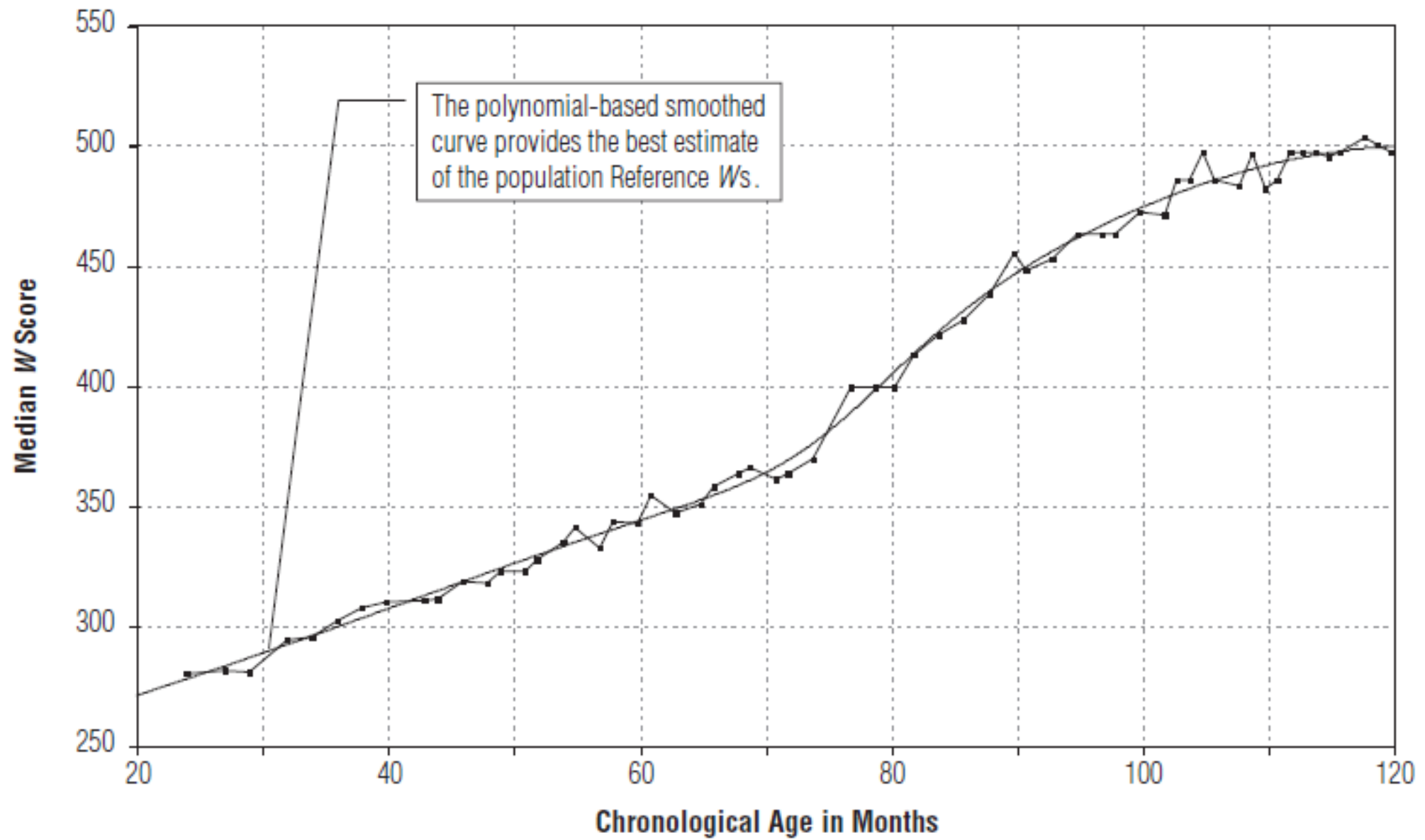
Cíle vyhlazování:

- Redukce výběrové chyby (trade-off mezi  $N$ , počtem parametrů a výběrovou chybou).
- Odstranění nekonzistencí, které by dělaly problém při interpretaci výsledku.

Celá řada postupů pro různé druhy vyhlazení.

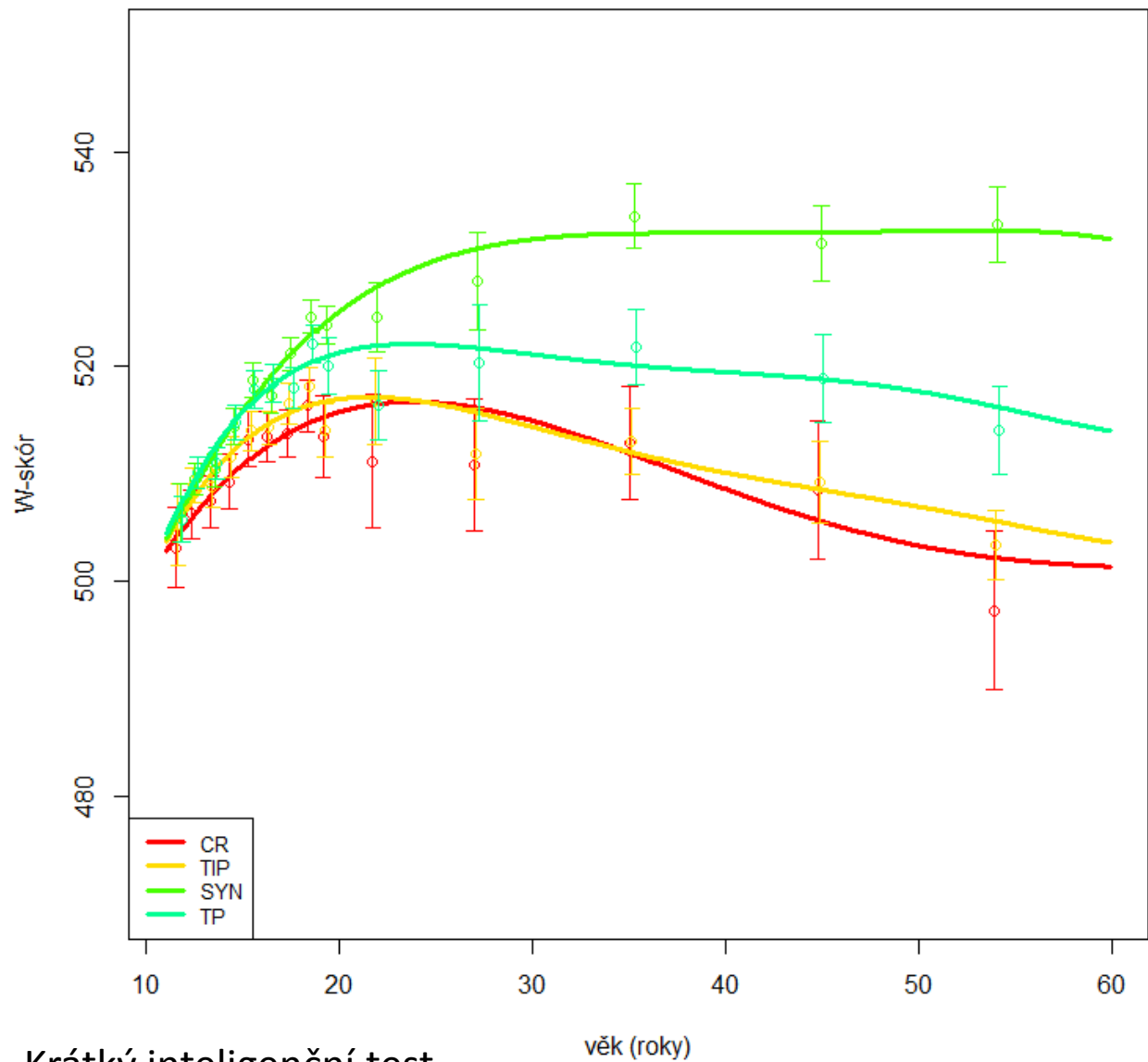
- „Ruční“ korekce/vyhlazení 😊
- Kernel density smoothing – pro vyhlazení percentilů uvnitř kategorie.
- Polynomy, frakční polynomy, spline smoothing – vyhlazení  $M$ ,  $SD$  napříč kategoriemi.
- Vyhlazení prostřednictvím Taylorových polynomů (R balíček [cNORM](#)).

**Další výhoda:** Normy s přesností na měsíc či den (není nutné mít „široké“ kohorty).



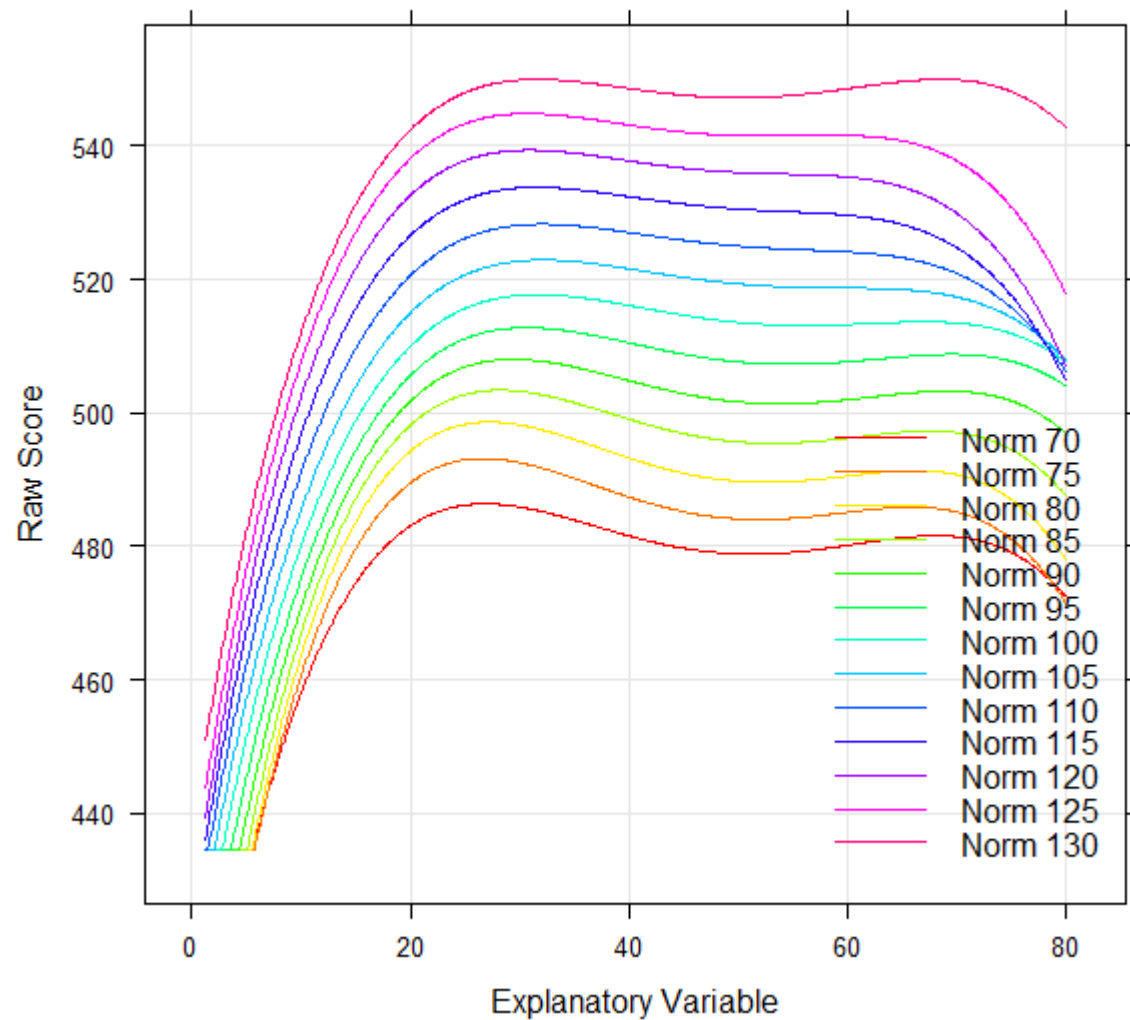
*Note.* A similar process is completed with the standard deviations (*SDs*) for Letter-Word Identification.

Technický manuál testu WJ-IV



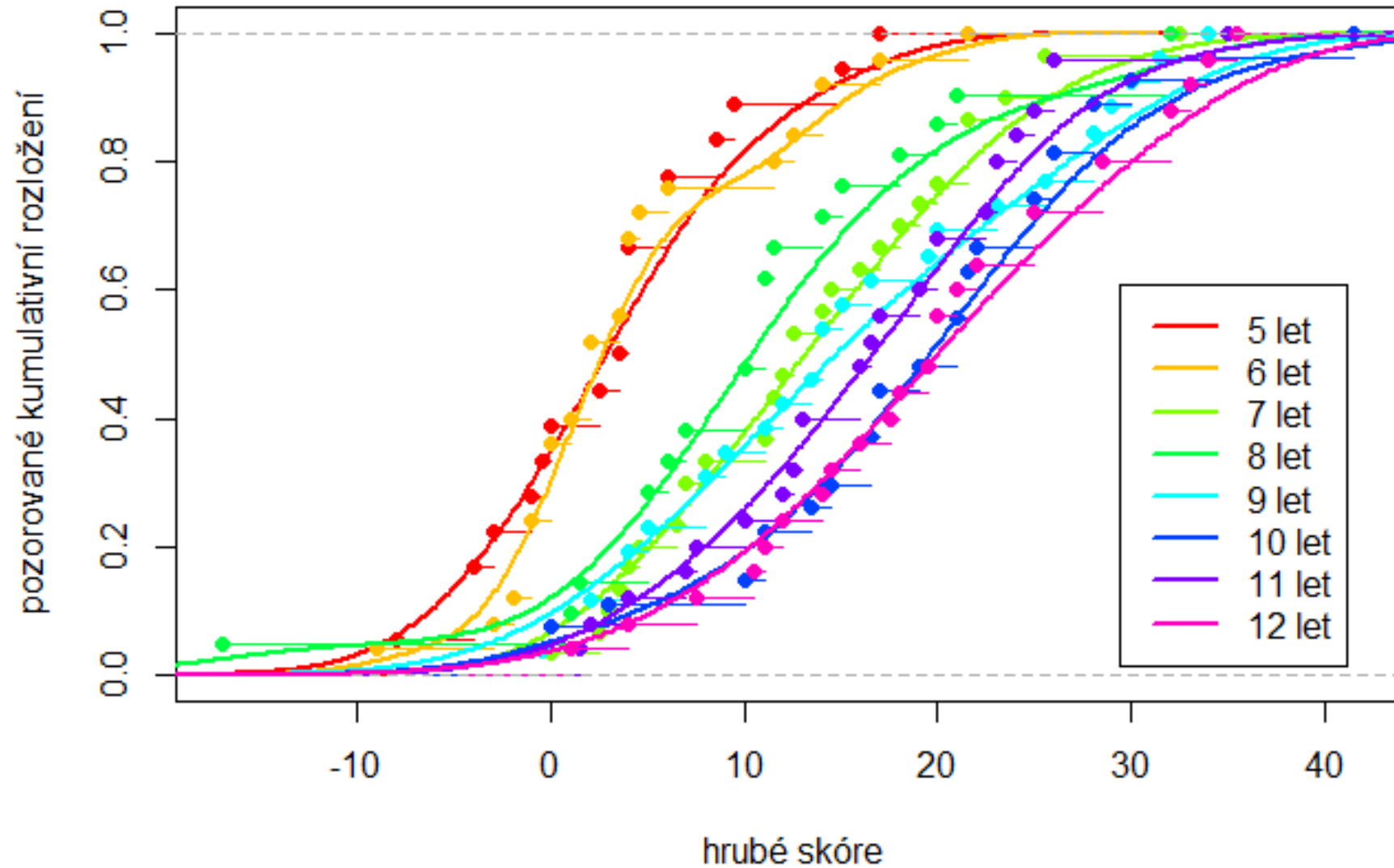
Krátký inteligenční test

### Norm Curves



WJ-IV COG CZ (pracovní analýzy)  
cNORM package

# dívky



Test vytváření příběhů

# Vývojové skóry

---

**Věkové ekvivalenty (age equivalent)** – jakému věku odpovídá dané skóre?

- Věk, v němž respondenti průměrně dosahují daného skóre.
- Analogie „mentálního věku“ (Binet) – dnes se tento termín nepoužívá.

**Ročníkový ekvivalent** – totéž, ale pro ročník/třídu.

**Zóna vývoje** – věkové skóre v podobě rozsahu.

- Rozsah na základě chyby měření, nebo častěji na základě stadiální křivky vývoje.

# Raschovské skóry (théta)

---

Kategorie skórů založená na Teorii odpovědi na položku (IRT), konkrétně 1parametrovém (Raschově) IRT modelu.

- Viz poslední přednáška.

Analogie hrubého skóre v CTT. Výhodnější např. pro sledování vývoje.

## **W-skóre.**

- Referenční bod: Právě 10leté děti mají průměrně  $W=500$ .
- Univerzální *jednotka*: Pokud někdo s  $W=A$  má 50% pravděpodobnost na správnou odpověď na určitou položku, pak někdo jiný s  $W=A+10$  má 75% pravděpodobnost, resp.  $W=A-10$  25%.

## **Index relativní výkonnosti**

- RPI – Relative Proficiency Index
- Ve formátu  $XX/90$ , např.  $47/90$  nebo  $94/90$ .
- „S jakou pravděpodobností respondent zvládne úkol, který jeho vrstevníci zvládnou s 90% pravděpodobností?”



# Ipsativní skórování

---

Nejsou zpravidla skóry v pravém slova smyslu:

- Standardní skóry srovnávají interindividuální variabilitu.
- Ipsativní skórování srovnává intraindividuální variabilitu.

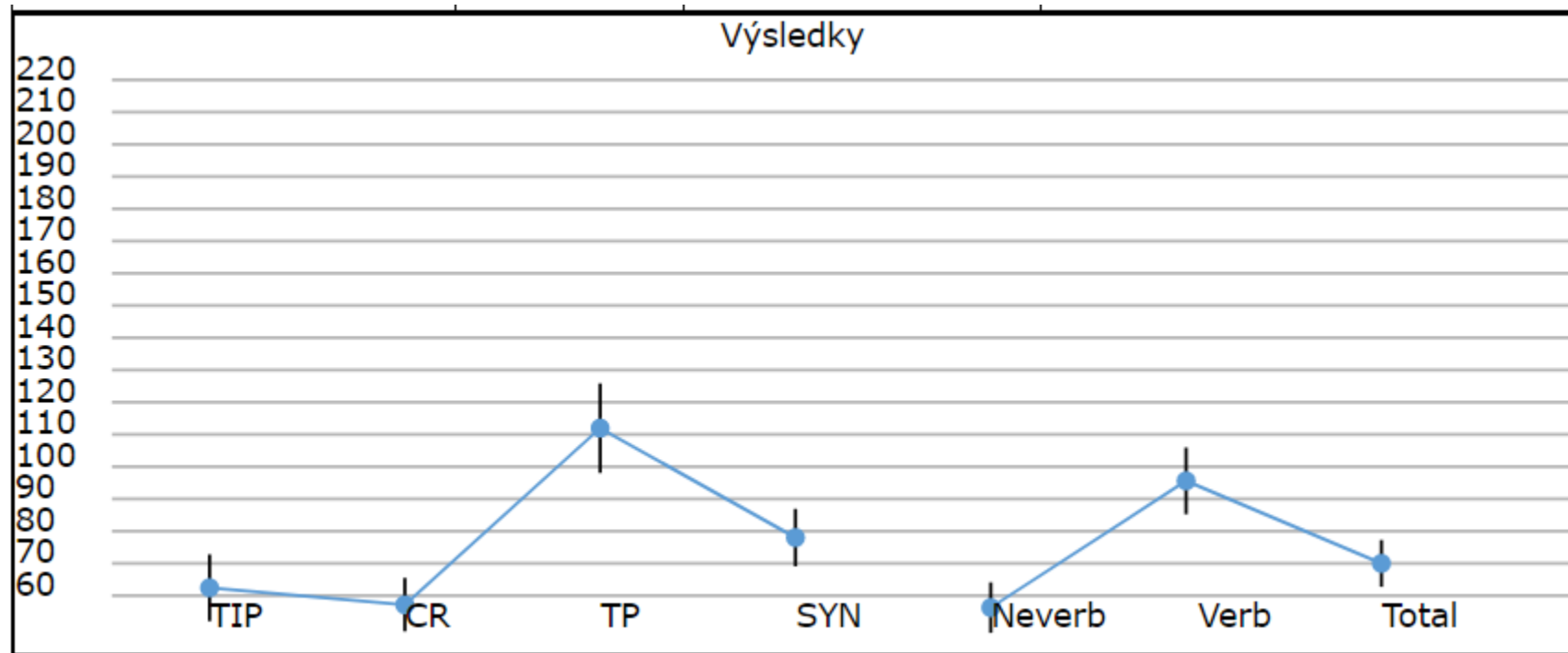
Založené na diskrepanci

- Předpokládáme, diskrepance mezi subtesty/faktory v inteligenčním testu může ukazovat na SPU.
- Kariérní poradenství – dotazník volby povolání („co člověka baví víc?“).

Používá se standardní chyba rozdílu, případně je rozdíl subtestů přímo standardizován.

Analýza profilu.

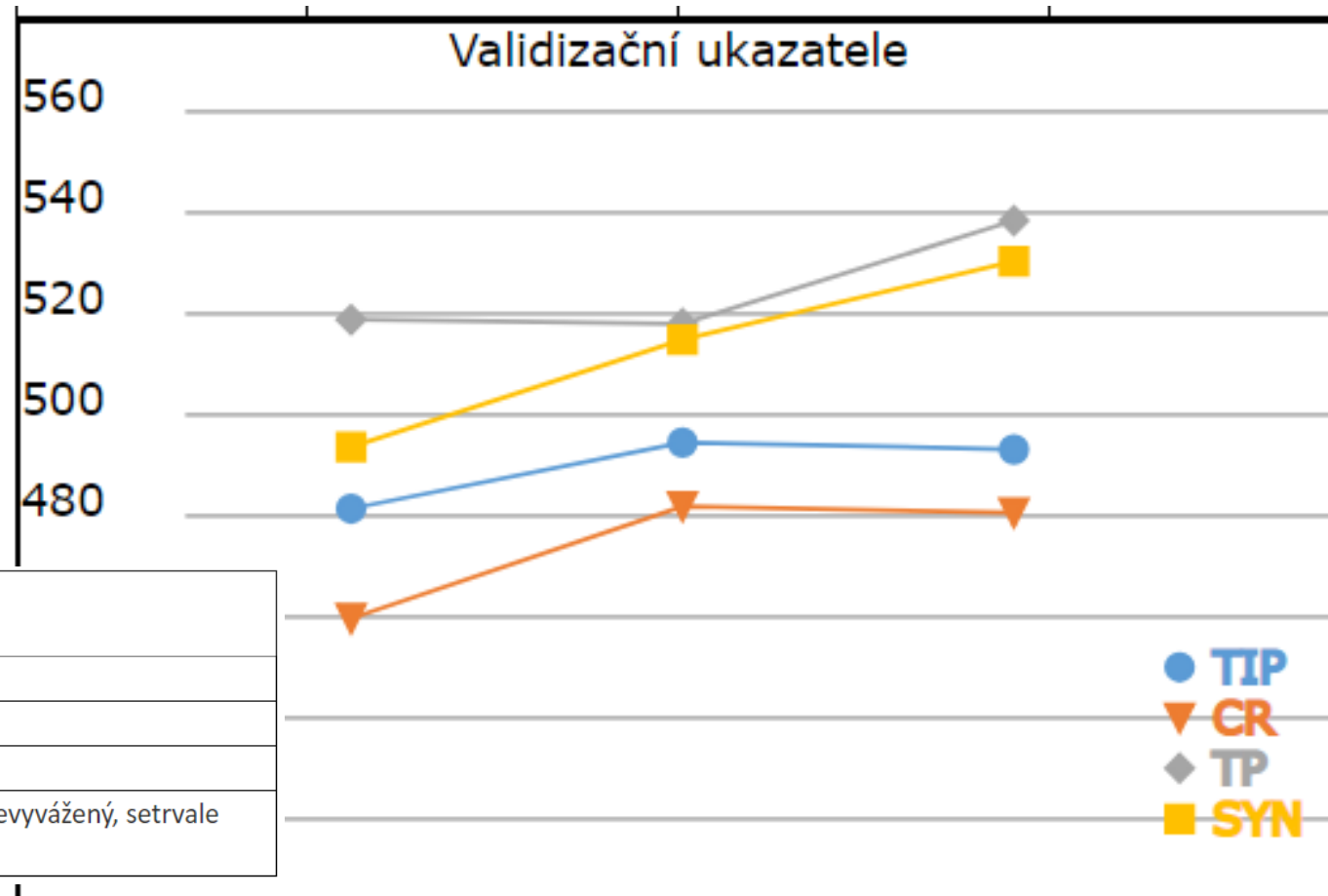
# Ipsativní skórování (více testů)



Krátký inteligenční test

# Ipsativní skórování (v rámci testu)

Krátký inteligenční test



| Validizační ukazatele | Část 1 | Část 2 | Část 3 | $\chi^2$ (df = 2) | p     | pozn   |
|-----------------------|--------|--------|--------|-------------------|-------|--|
| TIP                   | 481,37 | 494,44 | 493,09 | 1,6               | 0,449 |  |
| CR                    | 459,63 | 481,83 | 480,58 | 3,816             | 0,148 |  |
| TP                    | 518,76 | 517,97 | 538,37 | 2,11              | 0,348 |  |
| SYN                   | 493,55 | 514,81 | 530,31 | 12,677            | 0,002 | Pozor, v subtestu je patrný nevyvážený, setrvale rostoucí výkon. |

# Součtové/vážené skóry

---

Příklad: Máme inteligenční test. Chceme spočítat celkový skór (g-faktor). Můžeme:

- 1. sečíst všechny položky napříč subtesty.
- 2. standardizovat každý test a pak sečíst subtesty.
- 3. standardizovat každý test a vzít jejich vážený součet.

Výhody? Nevýhody?

Hlavní komplikace:

- Nelze sčítat nevážené subtesty (a tedy ani položky), mají jinou SD.
- Nelze předpokládat, že všechny vážené subtesty mají stejný vztah s g-faktorem. Na rozdíl od položek nepředpokládáme „náhodný výběr“ z domény.
- Efekt stropu, podlahy. U dětí různé „váhy“ pro různé referenční skupiny.
- Vliv chyby měření (testy s nižší reliabilitou mají nižší váhu). Různá chyba pro různé referenční skupiny.

Např. Wechsler: součet standardizovaných subtestů.

Např. Woodcock-Johnson: vážený průměr nestandardizovaných subtestů.

# Součtové/vážené skóry

**Formativní vs. reflektivní měření** na druhé úrovni vzhledem k chybě měření.

A. Reflektivní celkový skór.

- Celkový skór je odhad g-faktoru.
- Specifické rozptyly považovány za chybu.
- Vyšší míra chyby měření.

B. Formativní celkový skór.

- Celkový skór je jednoduše průměrem subtestů.
- Specifické rozptyly nehrají roli.
- Nižší míra chyby měření.

Zpravidla testy používají variantu B.

Různé odhady reliability/chyby měření.

