

# Standardní chyba měření

PSYb2590: Základy psychometrie (Seminář 2)

22. 3. 2021

# Standardní chyba měření (SEM)

$$SEM = \sigma_e = \sigma_X \sqrt{1 - r_{xx'}}$$

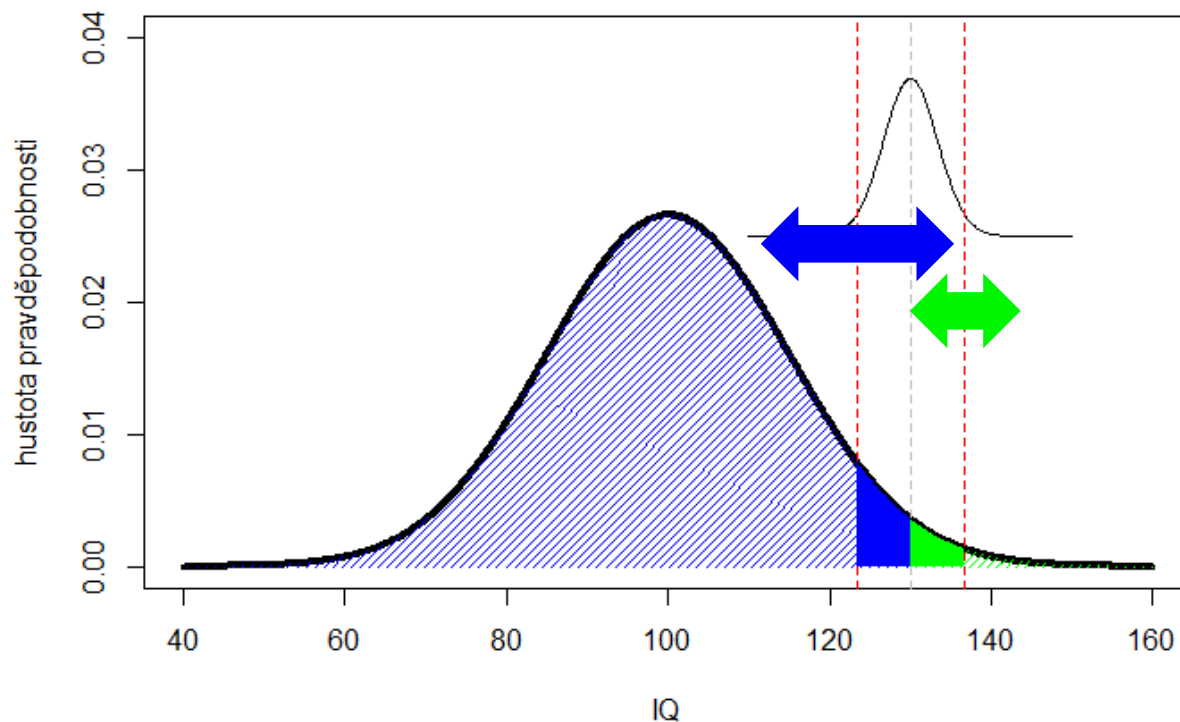
- Směrodatná odchylka jednotlivých paralelních testů (pozorovaných skóre) kolem jejich průměru (pravého skóre).
- Lze využít pro konstrukci intervalu spolehlivosti pro pravý skór
  - $CI = E(\tau) \pm z_p SEM$  ;
  - $z_p$  je kvantil normálního rozložení:  $z_{95\%} = 1,96$ ;  
 $z_{90\%} = 1,64$ ;  $z_{80\%} = 1,28$ ;  $z_{68\%} = 1$ .
- Ale...

# Regresní model CTT

- Naměřil jsem klientovi IQ 130 v inteligenčním testu.
  - Náhodně vybraný z populace, nemáme žádné další informace.
  - $r_{xx'} = 0,8$ .
  - Jaká je nejpravděpodobnější hodnota jeho „pravého“ IQ?

- $\sigma_X^2 = \sigma_T^2 + \sigma_e^2$

- Správná odpověď je 124.



# Regresní model CTT

- Naměřené hodnoty se pohybují kolem pravé hodnoty, nikoliv naopak. Jinými slovy: chyba měření je **chybou pravého skóru**, nikoliv pozorovaného.
  - Výsledkem je tzv. regrese k průměru.
  - Intervaly spolehlivosti jsou „asymetrické“ kolem naměřené hodnoty.
- Viz doporučení z povinné literatury (Dudek, [1979](#))

# Regresní model CTT

- Máme 2 proměnné:  $X$  a  $T$ 
  - $X$ :  $M(X)$ ,  $VAR(X) = SD(X)^2$
  - $T$ :  $M(T) = M(X)$ ,  $VAR(T) = r_{XT}VAR(X)$
  - $r_{XT} = \sqrt{r_{XX'}}$  ....  $r_{XT}^2 = r_{XX'}$
- Můžeme predikovat  $X$  z  $T$  nebo  **$T$  z  $X$**
- $T = bX + a$ 
  - $b = r_{XT} \sqrt{(r_{XT}^2 VAR(X) / VAR(X))} = r_{XT}^2 = r_{XX'}$
  - $a = M(T) - (r_{XX'}M(X)) = M(X) - (r_{XX'}M(X))$
- $T = r_{XX'}X + M(X) - (r_{XX'}M(X)) = r_{XX'}(X - M(X)) + M(X)$

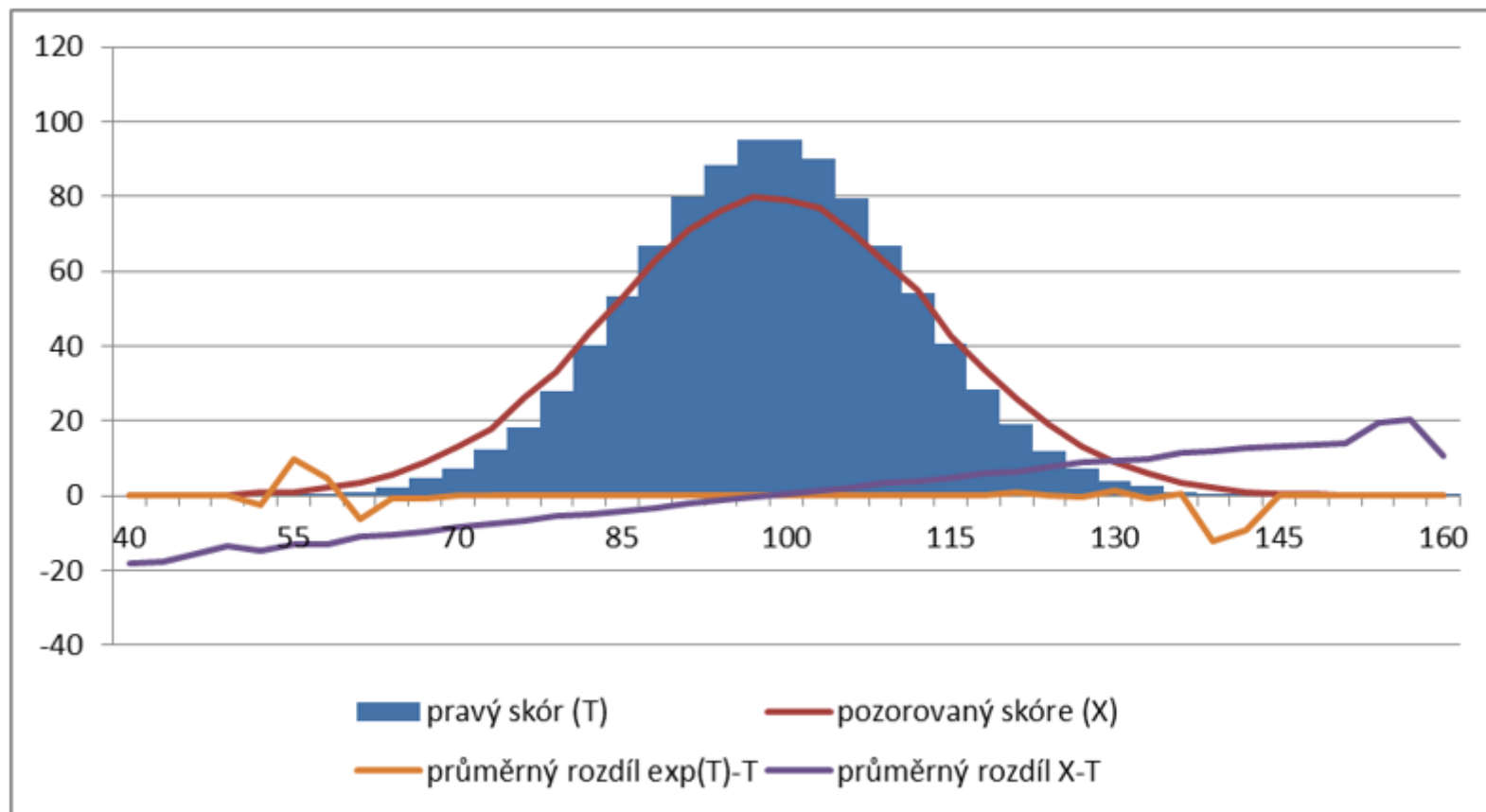
# Regresní model CTT

$$\begin{aligned} E(T|x) &= r_{xx'}x + (1 - r_{xx'})M_x \\ &= M + (X - M)r_{xx'} \end{aligned}$$

- $E(T|x)$  : očekávané pravé skóre na základě pozorovaného
- $r_{xx'}$  : reliabilita („směrnice“)
- $M_x$  : průměrné skóre;  $((1 - r_{xx'})M_x$  je „průsečík“)
- Čím větší je reliabilita, tím větší je vliv pozorovaného skóre a menší vliv populačního průměru (a naopak).

# Regresní model CTT

Obrázek 2: Simulace chyb měření a odhadu ( $N = 100\ 000$ )



# Pravé vs. pozorované skóre

- SD pravého vs. pozorovaného skóre:

$$r_{xx'} = \frac{\sigma_{\tau}^2}{\sigma_x^2} \rightarrow \sigma_{\tau} = \sqrt{r_{xx'}} \sigma_x$$

- SD pravého skóre má  $\sqrt{r_{xx'}}$ krát menší SD než pozorované skóre.
- Někdy se pro konstrukci standardních skóru používá právě odhad směrodatné odchylky pravého skóre  $\sigma_X \sqrt{r_{xx'}}$  (např. WISC-IV<sup>UK</sup>).

- Preferovaný postup.
- V takovém případě při standardizaci použijeme standardní chybu odhadu pravého skóre:

$$SE_{\tau} = \sigma_X \sqrt{r_{xx'}} \sqrt{1 - r_{xx'}}$$

- (Nemá vliv na další výpočty s již standardizovanými skóry.)



# Postup výpočtu intervalu spolehlivosti měření

## 1. Výběr vhodného koeficientu reliability

- Typicky vnitřní konzistence.

## 2. Odhad pravého skóre.

- $E(T|x) = r_{xx'}x + (1 - r_{xx'})M_x = M_x + (X - M_x)r_{xx'}$

## 3. Výpočet standardní chyby měření.

- $SEM = \sigma_e = \sigma_X \sqrt{1 - r_{xx'}}$

## 4. Volba šířky intervalu (hladiny spolehlivosti).

- $z_{95\%} = 1,96$ ;  $z_{90\%} = 1,64$ ;  $z_{80\%} = 1,28$ ;  $z_{68\%} = 1$

## 5. Konstrukce CI kolem odhadu pravého skóre.

- $CI = E(T|x) \pm z \cdot SE$

# Asymetrické intervaly spolehlivosti

r	0,001		0,2		0,7		0,9		0,999	
SE	14,99		13,42		8,22		4,74		0,47	
IQ	E(t)	95% CI	E(t)	95% CI	E(t)	95% CI	E(t)	95% CI	E(t)	95% CI
40	100	[70,6–129,3]	88	[61,7–114,3]	58	[41,9–74,1]	46	[36,7–55,3]	40	[39,1–41]
60	100	[70,6–129,3]	92	[65,7–118,3]	72	[55,9–88,1]	64	[54,7–73,3]	60	[59,1–61]
80	100	[70,6–129,4]	96	[69,7–122,3]	86	[69,9–102,1]	82	[72,7–91,3]	80	[79,1–80,9]
100	100	[70,6–129,4]	100	[73,7–126,3]	100	[83,9–116,1]	100	[90,7–109,3]	100	[99,1–100,9]
120	100	[70,6–129,4]	104	[77,7–130,3]	114	[97,9–130,1]	118	[108,7–127,3]	120	[119,1–120,9]
140	100	[70,7–129,4]	108	[81,7–134,3]	128	[111,9–144,1]	136	[126,7–145,3]	140	[139–140,9]
160	100	[70,7–129,4]	112	[85,7–138,3]	142	[125,9–158,1]	154	[144,7–163,3]	160	[159–160,9]

# Rozdíl dvou pozorovaných měření

- Nejjednodušeji: srovnání, zda se CI nepřekrývají.
  - Příliš striktní, malá síla testu.
- Standardní chyba rozdílu:
  - $SE_{A-B} = \sqrt{SE_A^2 + SE_B^2}$
  - V případě jediného testu:  $SE_{A-B} = \sigma_X \sqrt{2} \sqrt{1 - r_{xx'}}$
- Očekávaným rozdílem je 0, interval se konstruuje kolem nuly.
- Předpokládá se nezávislost chyb měření.

# Predikce budoucího pozorovaného skóre

- Standardní chyba predikce:

- $SE_{pred} = \sigma_X \sqrt{1 - r_{xx'}^2}$

- Nepredikujeme z měření na pravý skór, ale z měření na měření; proto je nutné reliabilitu ještě jednou umocnit.
- Očekávaným skórem je odhad pravého skóre, konstruuje se kolem predikce.
- Typicky se využívá test-retest reliabilita.

# Více různých druhů chyb

- Více chyb pro více účelů. Přehled:
  - Dudek, F. J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin*, 86(2), 335–337. <https://doi.org/10.1037/0033-2909.86.2.335>
  - Harvill, L. M. (1991). An NCME Instructional Module on Standard Error of Measurement. *Educational Measurement: Issues and Practice*, 10(2), 33–41. <https://doi.org/10.1111/j.1745-3992.1991.tb00195.x>
  - Cígler, H., & Šmíra, M. (2015). Chyba měření a odhad pravého skóru: Připomenutí některých postupů Klasické testové teorie. *Testforum*, 4(6), 67–84. <https://doi.org/10.5817/TF2015-6-104>
- Pro účely PSb2590 stačí výpočty uvedené v prezentaci.
- <https://hynekcigler.shinyapps.io/kalkulacka/>
  - (Pozor, jen vývojová verze, není dokončeno.)

# Praktické cvičení 1

- Zbyněk byl vyšetřen testem hudebního nadání a v testu dosáhl 40 bodů.
- Víte, že průměrné skóre je  $M=60$  ( $SD=20$ ) a reliabilita  $r=0,7$ .
- Jaký je interval spolehlivosti tohoto měření?
- Zbyněk není spokojen s výsledkem a nechá se vyšetřit znovu. V jakém intervalu bude nejspíše ležet jeho druhý výsledek?

# Praktické cvičení 2

- Zbyšek byl vyšetřen testem matematických schopností, jeho T-skóre je  $T=70$ .
- Víte, že vnitřní konzistence je  $\alpha=0,9$  a test-retest po 3 měsících  $r=0,8$ .
- Jaký je interval spolehlivosti tohoto měření?
- Na základě testování Zbyšek podstoupí 3měsíční kurz rozvoje matematických schopností. Po jeho ukončení je znovu vyšetřen. Jakého skóre musí dosáhnout, aby byl kurz „úspěšný“?

# Praktické cvičení 3

- Zbyněk byl vyšetřen psycholožkou dvěma talentovými testy – testem hudebního nadání a testem matematického nadání.
- V testu **hudebního** nadání získal Zbyněk **70 bodů z 90** možných a v testu **matematického** nadání **75 bodů ze 100** možných.
- Víme, že test **hudebního** nadání má přibližně normální rozložení o průměru 50b ( $SD=20$ ) a test **matematického** nadání má také normální rozložení o průměru 45 ( $SD=15$ ).
- Reliabilita testu **hudebního** nadání je  $r_H=0,8$ , reliabilita testu **matematického** nadání je  $r_M=0,9$ .
- Ve které z testovaných oblastí má Zbyněk výraznější talent?