

Chyba!

Zuzana Derflerová Brázdová

LF MU

Sémantika výrazu „chyba“

Chyba vs. Omyl vs. Pochybení vs. Zmýlená vs.
Zkreslení vs. Matoucí faktor

(Error – mistake – lapsus – faux pas – oversight –
bias – confounding factor)

- **Chyba** může představovat odchylku od normy nebo požadovaného stavu, popř. porušení daných pravidel.
- Ve statistice chyba není „omyl“, ale rozdíl mezi vypočítanou, očekávanou nebo změřenou hodnotou a hodnotou skutečnou nebo teoreticky správnou.

Výskyt chyb

- V rámci šetření je každá jednotlivá operace automaticky možným zdrojem chyb.

Příklady:

Nevhodně provedený výběr souboru, špatný design dotazníku, vadně kalibrovaný přístroj, nepravdivé odpovědi dotazovaných, nepozorně vkládaná data při zpracování, špatně zvolená statistická metoda, zavádějící interpretace

Má tedy vůbec smysl provádět nějaká šetření?

Ano, když:

1. budeme znát možné zdroje chyb
2. budeme se snažit jim předejít
3. budeme již vzniklé chyby poctivě vyhledávat
4. při interpretaci k nim nepokrytě přihlédneme

Chyby při měření I.

- Jakékoli opakované měření téhož jevu vykazuje mezi každými dvěma pokusy variace, tj. nejistotu ohledně správnosti.
- Tuto nejistotu lze nazývat chybou.
- Chyba není v tomto případě totéž co omyl!

Chyby při měření II.

- Vzhledem k axiomu, že žádné měření není přesné, hledáme hodnotu „nejblíže k něčemu“.
- Největší samozřejmá možná chyba bude tedy odpovídat $\frac{1}{2}$ měřicí jednotky.

Chyby při měření III.

Chybu lze vyjádřit jako

- absolutní

nebo

- relativní

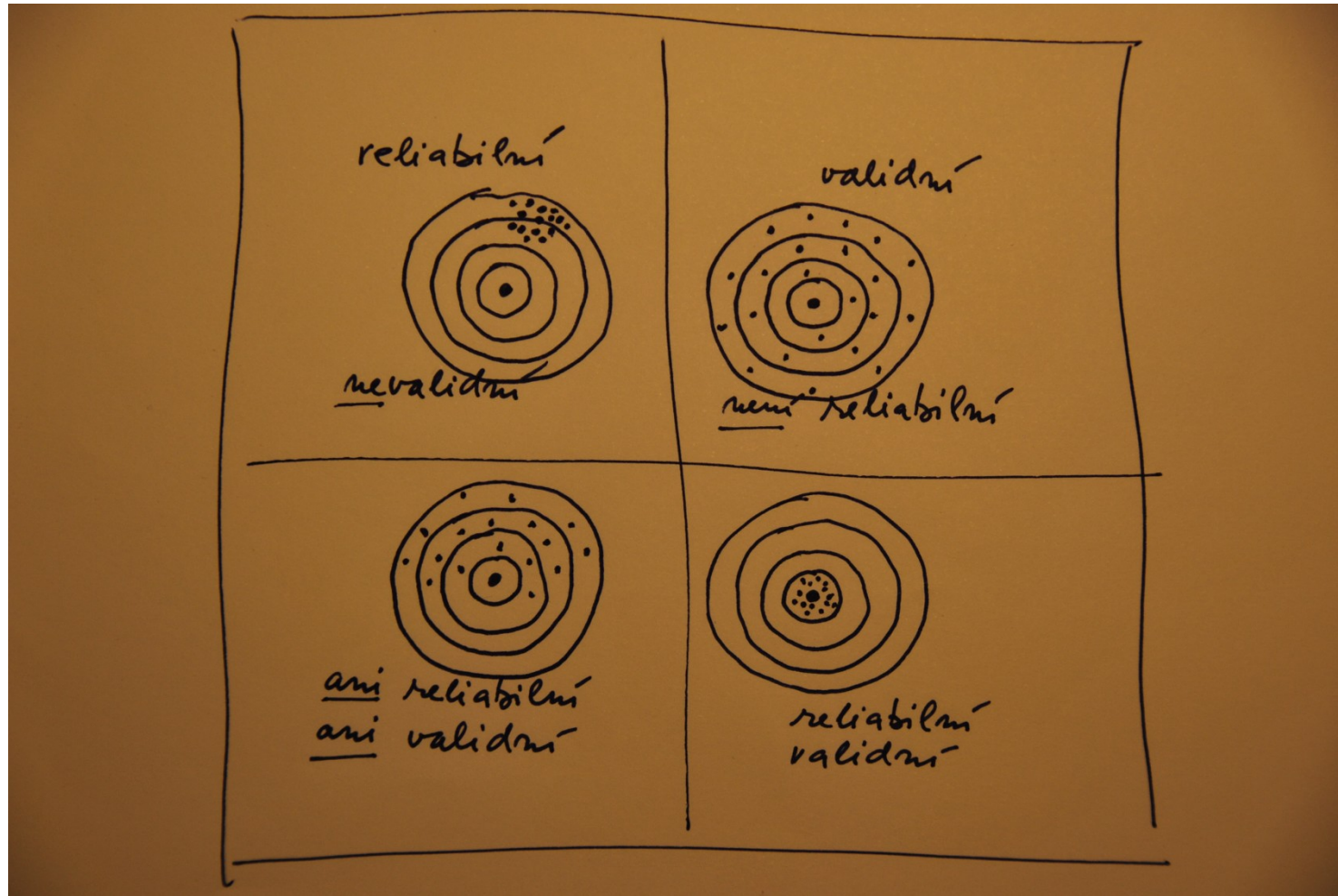
Relativní vyjádření chyby více indikuje význam chyby.

Chyby při měření IV.

Možnosti předcházení technickým chybám při měření:

1. Precizování měření užitím co nejmenších měřících jednotek
2. Použití korektních technik, např. vyvarování se „paralaxy“
3. Opakování měření, abychom získali věrohodný průměr
4. Práce v kontrolovaných podmínkách, vždy stejných

Reliabilita a validita



Komponenty chyby

$$x = T + e,$$

$$\text{kde } e = e_r + e_s$$

e_r náhodná chyba

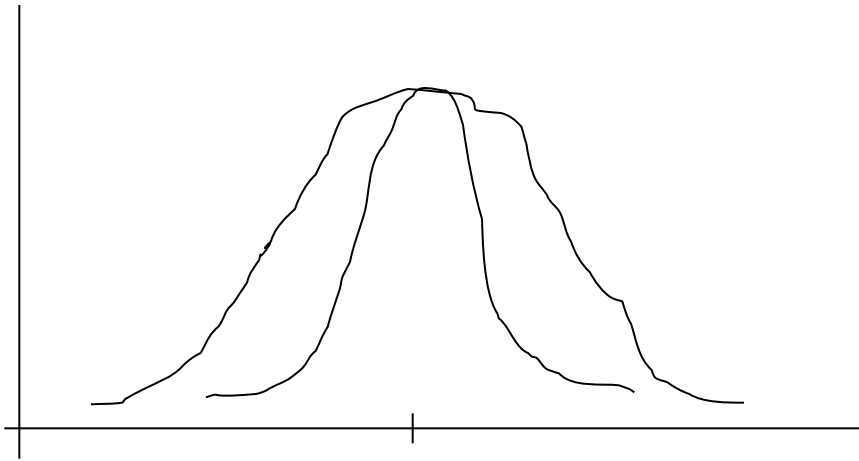
e_ssystematická chyba,

takže

$$x = T + e_r + e_s$$

Náhodná chyba

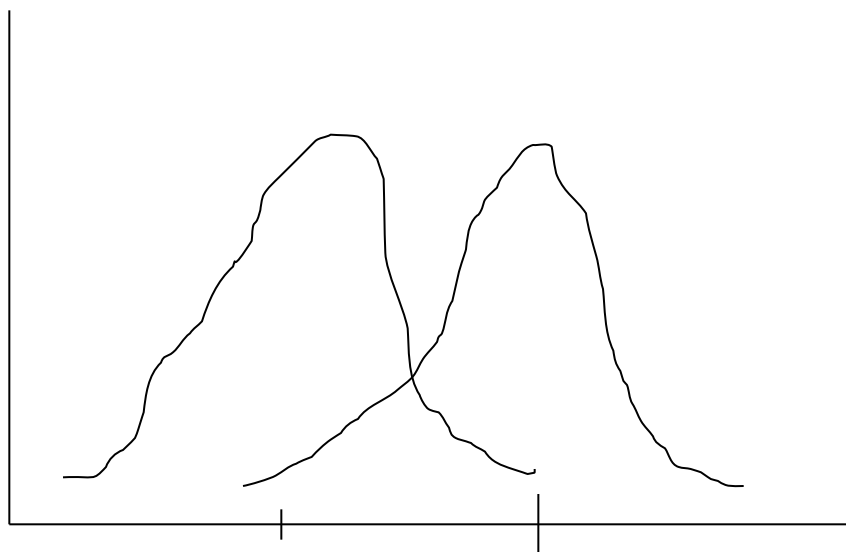
Random error neboli šum



- *Neovlivňuje průměr, ale ovlivňuje variabilitu kolem průměru*

Systematická chyba

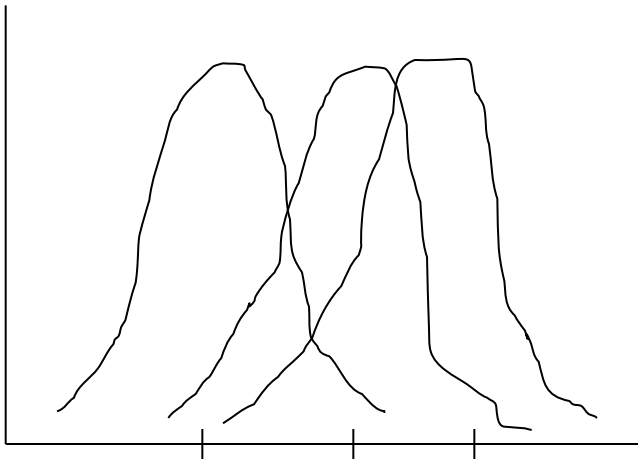
Systematic error *neboli* bias (omyl)



Ovlivňuje průměr – proto termín „bias“
(zkreslení)

Co s tím?

- Jestliže různá měření téhož vykazují sice stejnou variabilitu, ale jiné průměry, je možné použít triangulaci k přiblížení se pravdivé – skutečné hodnotě



Úroveň měření

Neznalost úrovně měření je příčinou chyby.

Je důležité vědět, zda čísla, která hodnotím a interpretuji, vyjadřují hodnotu:

1. Nominální
2. Pořadovou (ordinální)
3. Intervalovou
4. Poměrnou

Nominální hodnoty

Číslo je jen popisné, analogické jménu.

Nelze jej matematicky zpracovávat!

Příklad: Hráč s číslem 12 na dresu se střetl s hráčem s číslem 36.

Pořadové – ordinální hodnoty

Pořadové označení může vyjadřovat měření v čase nebo dokonce kvantitu, ale bez záruky matematických hodnot.

Příklad: konzumace zeleniny vůbec = 0, 1x měsíčně = 1, 1x týdně = 2, denně = 3, 5x denně = 4

Rozdíly mezi kategorií 0 a 2 nejsou stejné jako mezi 1 a 3.

Intervalové hodnoty

- Čísla mají skutečně tu matematickou hodnotu, kterou vykazují v měření. Lze tedy počítat průměr, SD.

Příklad: měření v cm, kg, g

Pozor na měrné jednotky s logaritmickou povahou!

Měření poměrné

- Pozor! Číslo 0 se chová skutečně matematicky, tj. jako 0.

Chyby při designu souboru

- Randomizace!
- Velikost populace – očekávaná četnost výskytu sledovaného jevu v populaci – velikost vybraného souboru