

PSY117/454

Statistická analýza dat v psychologii

Přednáška 3

MÍRY CENTRÁLNÍ TENDENCE a VARIABILITY

Centrální tendence (=střední hodnoty)

- CT je údaj, jímž se snažíme popsat rozložení jedné proměnné
- Jeho kouzlo i zrádnost je právě v tom, že je to právě jeden údaj.
- CT udává průměrnou, typickou, reprezentativní, *očekávanou* hodnotu
 - Co se těmito slovy míníme, záleží na tom, jakou míru CT se rozhodneme použít

Modus, medián a průměr

□ Modus - kategoriální typická hodnota

- nejčastější hodnota, h. s nejvyšší četností
- jediná možnost u nominálních dat, u vyšších úrovní často užitečnou volbou

\hat{X}, Mo

□ Medián – pořadová střední hodnota

- hodnota prvku uprostřed uspořádaného souboru
- při sudém počtu prvků je mediánem kterékoli číslo z intervalu mezi nejbližší vyšší a nejbližší nižší hodnotou (konsenzuálně střed intervalu)
- pořadová data a výše

\tilde{X}, Md

□ Aritmetický průměr – deviační střední hodnota

- jak ho znáte ze školy
- pouze intervalová a poměrová data
- velmi citlivý na extrémní hodnoty

\bar{X}, M, m

Střední hodnoty - poznámky

- je třeba je umět spočítat ručně (a zopakovat si práci se sumačním symbolem Σ)
- i vážený průměr
- jak je ovlivní datové transformace přičtení konstanty a násobení konstantou
- vhodnost použití 3M (Hendl s.95)

AJ: weighted mean, add, multiply

Míry variability (rozptýlenosti)

- nominální – entropie – nepoužívá se
- pořadové
 - (variační) rozpětí = $X_{max} - X_{min}$ (extrémně kolísá s velikostí vzorku)
 - kvartilové rozpětí = $Q_3 - Q_1$
- odchytkové (deviační, momentové) ukazatele
 - průměrná absolutní odchylka ($\Sigma|x|/n$) – nepoužívá se
 - průměrná odchylka na druhou – rozptyl
 - populační ($\Sigma x^2/n$) vs. výběrový ($\Sigma x^2/(n-1)$)
 - součet odchylek na druhou = suma čtverců
 - směrodatná odchylka (standardní odchylka)
 - odmocnina rozptylu - návrat k původní jednotce

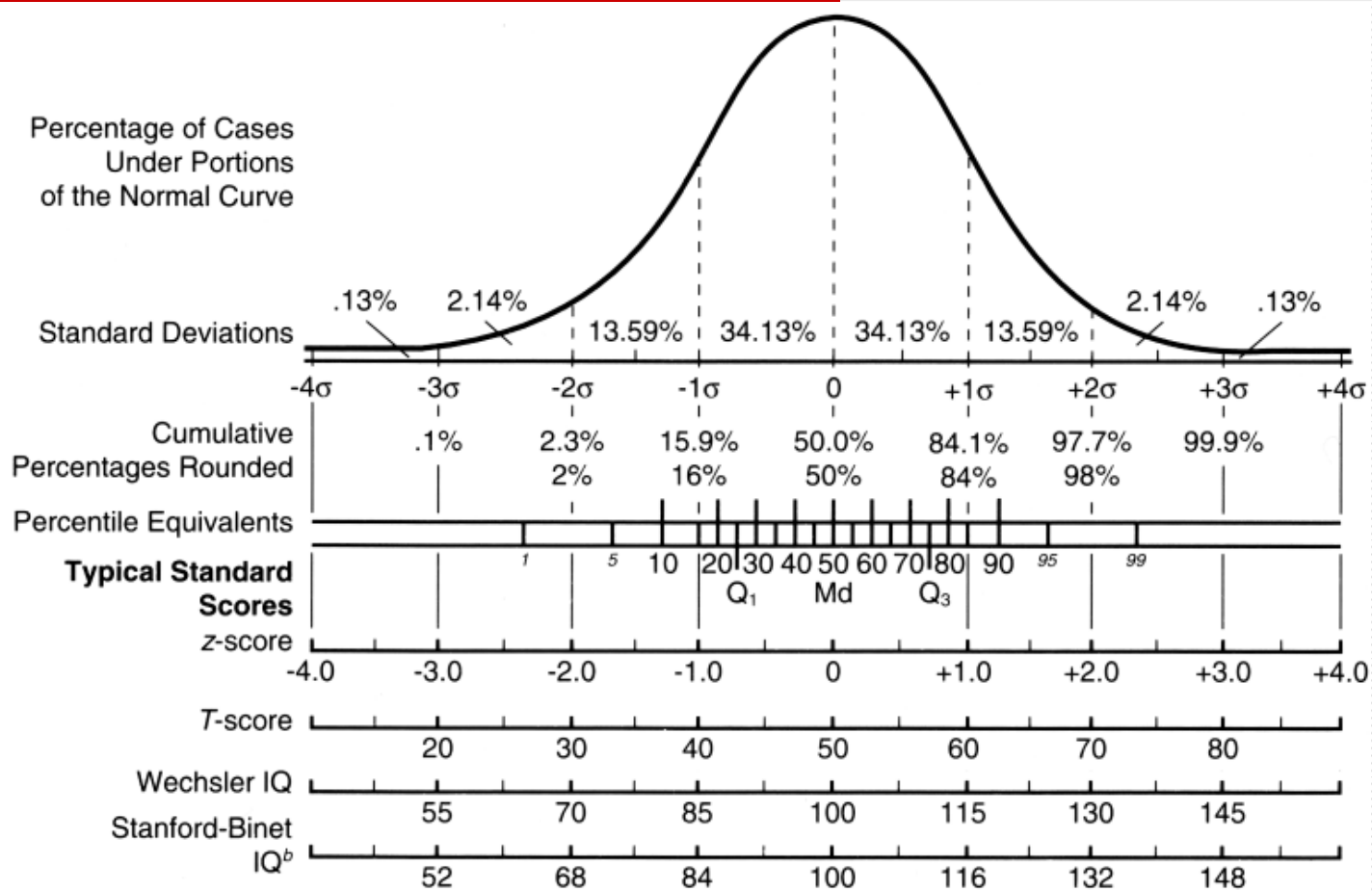
Souhrn

- Kategoriální deskriptivy
 - modus, (entropie)
 - Pořadové deskriptivy
 - medián, kvartily, percentily (a jiné *kvantily*)
 - kvartilové rozpětí
 - Odchylkové, momentové deskriptivy
 - aritmetický průměr
 - rozptyl, směrodatná odchylka (k=2)
 - zešikmení (k=3) = $\sum x^k / n$
 - špičatost (strmost) (k=4)
-

z-skóry, standardizované s.

- transformace sady skóřů tak, aby $m=0$, $s=1$
– jednotkou měření se stává s
- $z_i = (x_i - m) / s$
- u přibližně normálně rozložených dat o lidech je většina (přes 90%) lidí mezi -3 a 3
- ze z-skóřů pak např. T-skóřy ($m=50, s=10$), IQ-skóřy (100, 15) apod.

Oblasti pod křivkou normálního rozložení



Statistické zkratky a značky

- různé systémy, je třeba dobře popisovat
- N, n = velikost vzorku (podvzorku)
- X_i = skór i-té osoby u proměnné X
- x_i = deviační skór, odchylka od průměru
- M, m, \bar{x} = průměr
- SD, s = směrodatná odchylka
- s^2 = rozptyl

AJ: statistical notation, sample size, subsample, score, deviation score
