

PSY117/454

Statistická analýza dat v psychologii

Přednáška 4

KVALITY STATISTIK (+NORMÁLNÍ ROZLOŽENÍ)



Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital.

Aaron Levenstein

Srovnání ukazatelů střední hodnoty

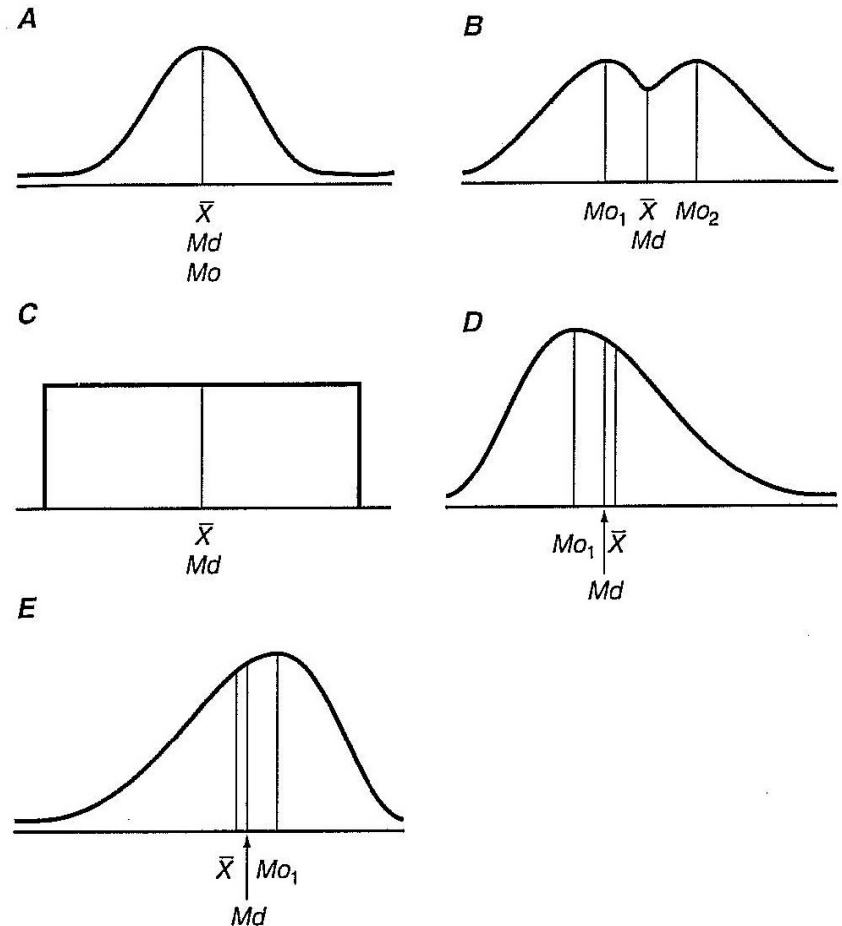
M_o – nejčastější hodnota

M_d – střed setříděného souboru, min součet abs. odchylek

M – těžiště, min součet umocněných odchylek (nejmenší čtverce)

Výpovědní hodnota - závislost na tvaru rozložení - robustnost
výskytu extrémních hodnot - rezistence
... úrovní měření

ROBUSTNOST a REZISTENCE jsou žádoucími vlastnostmi statistik



Srovnání ukazatelů rozptýlenosti

VR – empirický rozsah (podle úrovně měření v měřítku či v počtu bodů na teoretické škále)

IQR – pásmo 50% prostředních

s, s^2 – průměrná mocnina odchylky, s v jednotce měření

- Na rozdíl od středních hodnot nejsou ukazatele rozptýlenosti navzájem numericky srovnatelné. např. výška: $VR=38, IQR=12, s=8, s^2=64$
 - Sdílí R a R s odpovídajícími ukazateli střední hodnoty.
 - Kromě zešikmení a extrémních hodnot je u variability dobré ještě dávat pozor na empirické meze měření (strop/podlaha).
-

Popisné statistiky - mezishrnutí

- Pokud nejsou data na intervalové úrovni měření a nejsou pěkně symetricky normálně rozložená, jsou lepšími popisnými statistikami POŘADOVÉ statistiky.
 - Hmm ... a proč jsou tedy všude, kam se podívám průměry a směrodatné odchylky?
-

Estimační kvality statistik I

Kvality statistiky jako prostředku odhadu „skutečné“ hodnoty v populaci

TABLE 5.1

The Expected Values of the Range, s^2 , and s as a Function of Sample Size of n Observations from a Random Sample from a Normal Distribution in which $\sigma = 10$

<i>If $\sigma = 10$ n</i>	<i>Expected Value of the Range</i>	<i>Expected Value of s^2</i>	<i>Expected Value of s</i>	<i>Expected Value of Range/s</i>
2	11	100	8.0	1.4
5	23	100	9.4	2.4
10	31	100	9.73	3.2
20	37	100	9.87	3.7
50	45	100	9.95	4.5
100	50	100	9.97	5.0
200	55	100	9.987	5.5
500	61	100	9.993	6.1
1,000	65	100	9.997	6.5

Estimační kvality statistik II

- Nezkreslenost
 - tj. že systematicky nenad(pod)hodnocuje
 - např. s podhodnocuje
- Konzistence
 - s velikostí vzorku roste přesnost odhadu
- Relativní účinnost
 - jak rychle roste přesnost se vzorkem
 - zde vítězí M nad Md a strhává s sebou i další momentové statistiky
 - *jejich výhodou je i snadné počítání s nimi*

Popisné statistiky - shrnutí

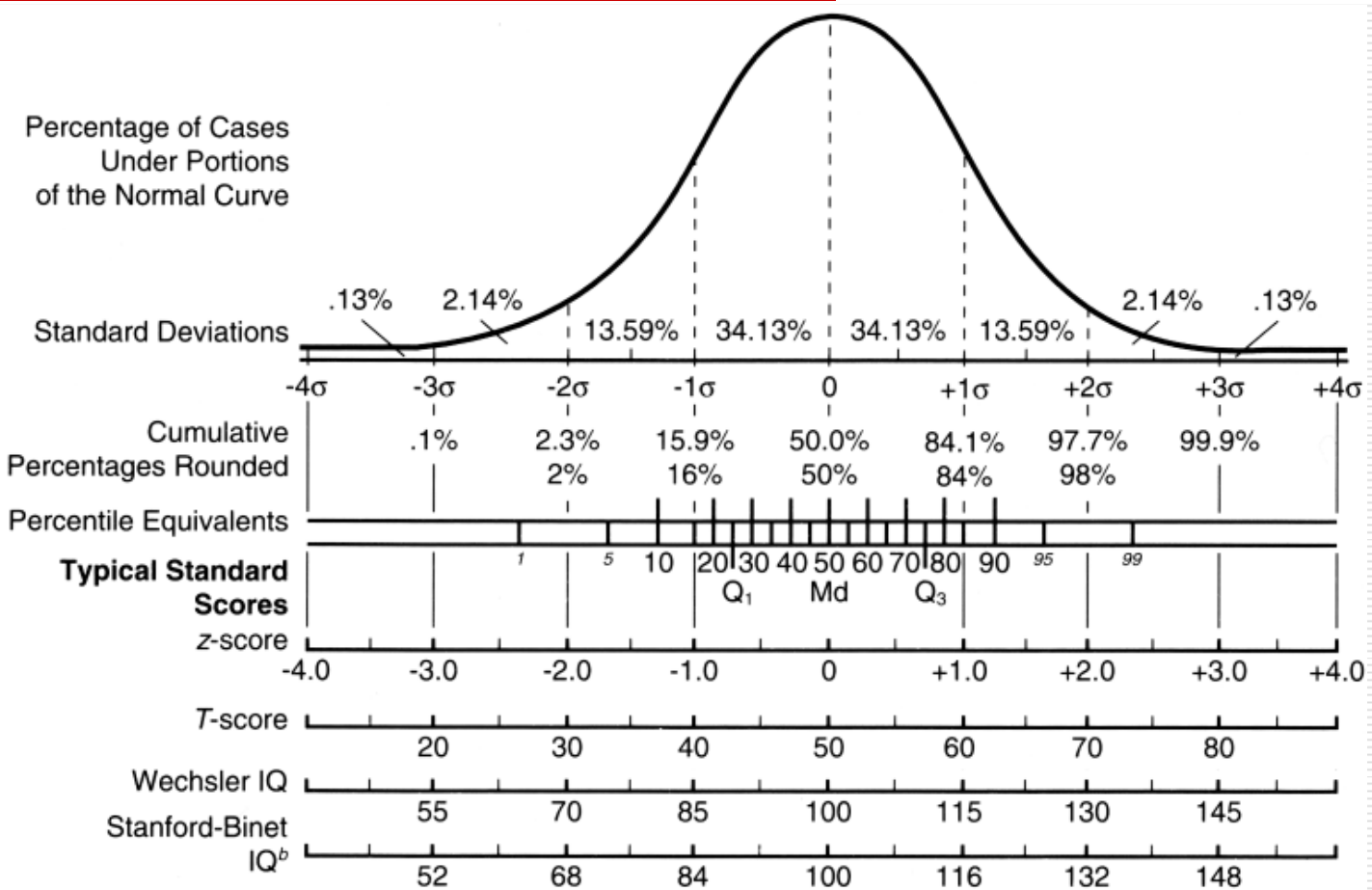
Zvažujeme

- úroveň měření
- tvar rozložení – symetrie, normalita
- cíl studie – pouze popis X usuzování, porovnávání

Tedy...

- Je-li cílem především deskripce dat, pak použijeme **POŘADOVÉ** ukazatele. Připojíme-li i odchylkové, nic nezkazíme.
 - $N, min, Q_1, Md, Q_3, max$
 - boxplot
 - pro individuální skóry **percentily**
 - Je-li cílem další usuzování, porovnávání apod., používáme **ODCHYLKOVÉ** ukazatele
 - N, m, s (N, M, SD)
 - popis rozložení
 - pro individuální skóry **z-skóry**
-

Normální rozložení podruhé I.





Ze z-skórů na percentily a zase zpět aneb *area under the curve*.

□ Normální rozložení.xls v ISu

