

Normální rozdělení

Úvod

- Potřebujeme znát pro další statistickou práci
- Může být použité pro srovnání populací
- Jak vypadá rozdělení hodnot proměnné?
- Vypadá podobně jako zvon nebo nějak jinak?

Na jaké znalosti budeme dnes stavět?

- Typy proměnných
- Histogram
- Rozptyl
- Průměr
- Směrodatná odchylka

Opakování typy proměnných

- Nominální
 - Ordinální
 - Intervalové
 - Poměrové
-
- Kategorické x kardinální

Opakování histogram

- Graf četnosti
- Kolikrát se daná hodnota objevuje v datovém souboru
- V spss – analyze – descriptives - frequencies

Opakování Průměr

- „geometrický“ střed
- suma vzdáleností na jednu stranu se rovná sumě vzdáleností na druhou stranu

Opakování rozptyl a sm.odch

- Jak moc se jednotlivé případy liší od průměrné hodnoty
- Suma umocněných odchylek
- Směrodatná odchylka je odmocninou rozptylu

Terminologická vsuvka

- „normální rozdělení“ je termín
- Rozdělení odlišné od „normálního“ může být pořád normální
 - Počet zlomenin

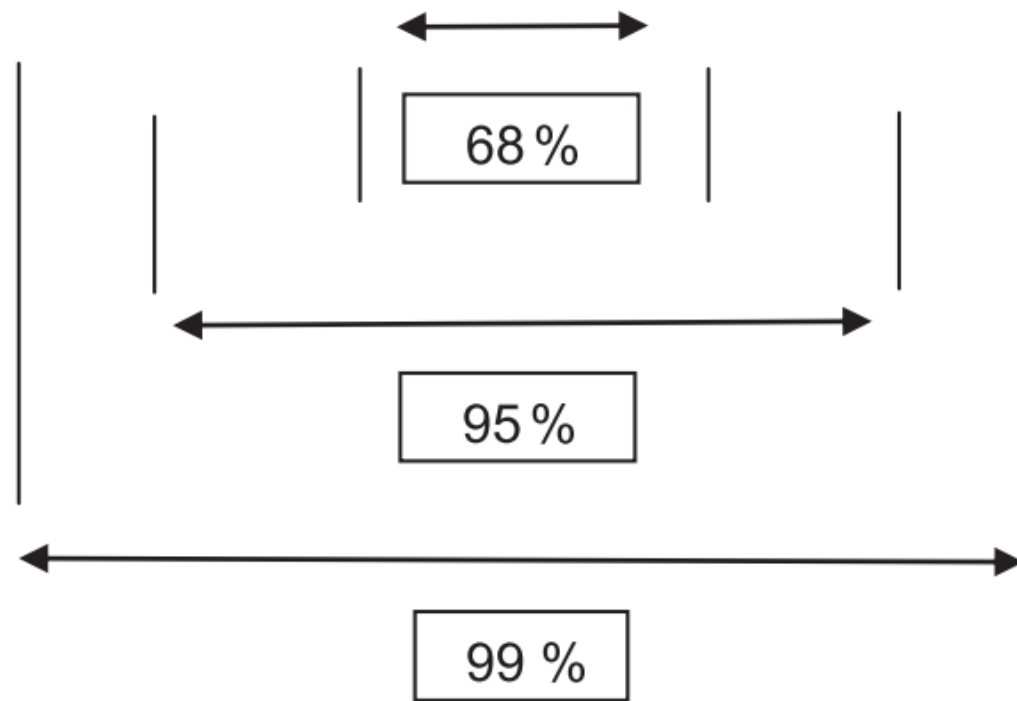
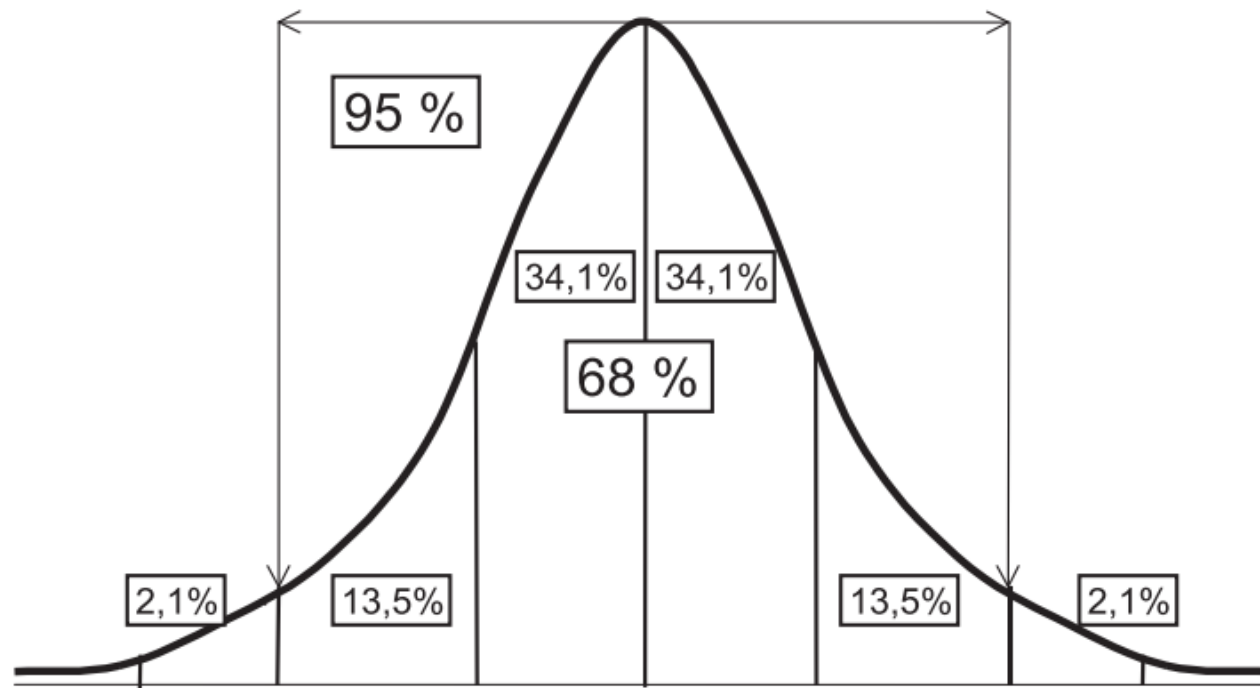
Co je to rozdělení

- Kolikrát se opakuje nějaká hodnota v datovém souboru
- Různé podoby rozdělení
 - Normální
 - Poisonovo
 - Uniformní

Jak normální rozdělení vypadá

- Symetrické rozložení zvonovitého tvaru
- Jeden vrchol
- Průměr = medián

- 67 % případů v intervalu průměr \pm 1 sm.odch.
- 95 % případů v intervalu průměr \pm 2 sm.odch.
- 99 % případů v intervalu průměr \pm 3 sm.odch.



Odpovídá rozdělení proměnné normálnímu rozdělení

- Ano - > můžeme používat řadu statistických nástrojů bez problémů
- Ne -> transformování proměnné
 - > rekódování proměnné
 - > přizpůsobení se situaci (záleží na požadované přesnosti)

Šikmost a špičatost

- šikmost (skewness), špičatost (kurtosis)
- Šikmost – průměr \neq medián (většina hodnot je pod nebo nad průměrem)
- Špičatost – hodnoty jsou příliš koncentrovány nebo rozptýleny okolo středu
- výsledek šikmosti nebo špičatosti podělený jejich standardní chybou (Std. Error) v absolutní hodnotě $> 1,96/2,58 \Rightarrow$ rozdělení proměnné je výrazně špičaté nebo šikmé (tohle má smysl pro vzorky do cca 500 případů)
- analyze – descriptive statistics - explore

KS test a SW test

- „exaktní“ zhodnocení normality
- Liší se rozložení proměnné signifikantně od normálního rozložení?
- Vhodné spíše pro malé vzorky
- analyze – descriptive statistics – explore - plots