

Histogram jako nástroj pro vizualizaci distribuce

Histogram je grafické znázornění rozložení souboru dat. Vytváří se rozdělením rozsahu hodnot v datech do řady intervalů (binů) a následným vynesemím počtu datových bodů v každém binu.

Histogramy jsou užitečným nástrojem pro vizualizaci tvaru distribuce, protože mohou pomoci identifikovat vzory a trendy v datech, jako je šikmost nebo multimodalita. Histogram může například ukazovat, že rozdělení je zkosené doleva nebo doprava, což naznačuje, že na jedné straně rozdělení je více datových bodů než na druhé.

Histogramy lze také použít k odhadu typu distribuce souboru dat, i když přesná forma distribuce není známa. To lze provést fitováním teoretického rozdělení, jako je normální rozdělení, do histogramu a porovnáním shody rozdělení s daty. Pokud je shoda dobrá, naznačuje to, že základní rozdělení může být podobné teoretickému rozdělení. Celkově jsou histogramy užitečným nástrojem pro pochopení tvaru a distribuce datové sady a mohou pomoci při další analýze a rozhodování. V rámci aplikace

Aplikace studentům umožňuje prozkoumávat histogramy. Generují vzorky buď z normálního, log-normálního, exponenciálního nebo jednotného rozdělení a vizualizují histogram. Mohou upravit velikost vzorku a velikost binů. Příliš široké biny poskytnou nedostatečné rozlišení; příliš úzké data neseskupí. V tomto appletu se podívejte jak vypadá histogram pro čtyři různé typy rozdělení. Cílem je najít vhodnou šířku binu pro vámi zvolený počet vzorků a typ rozdělení.

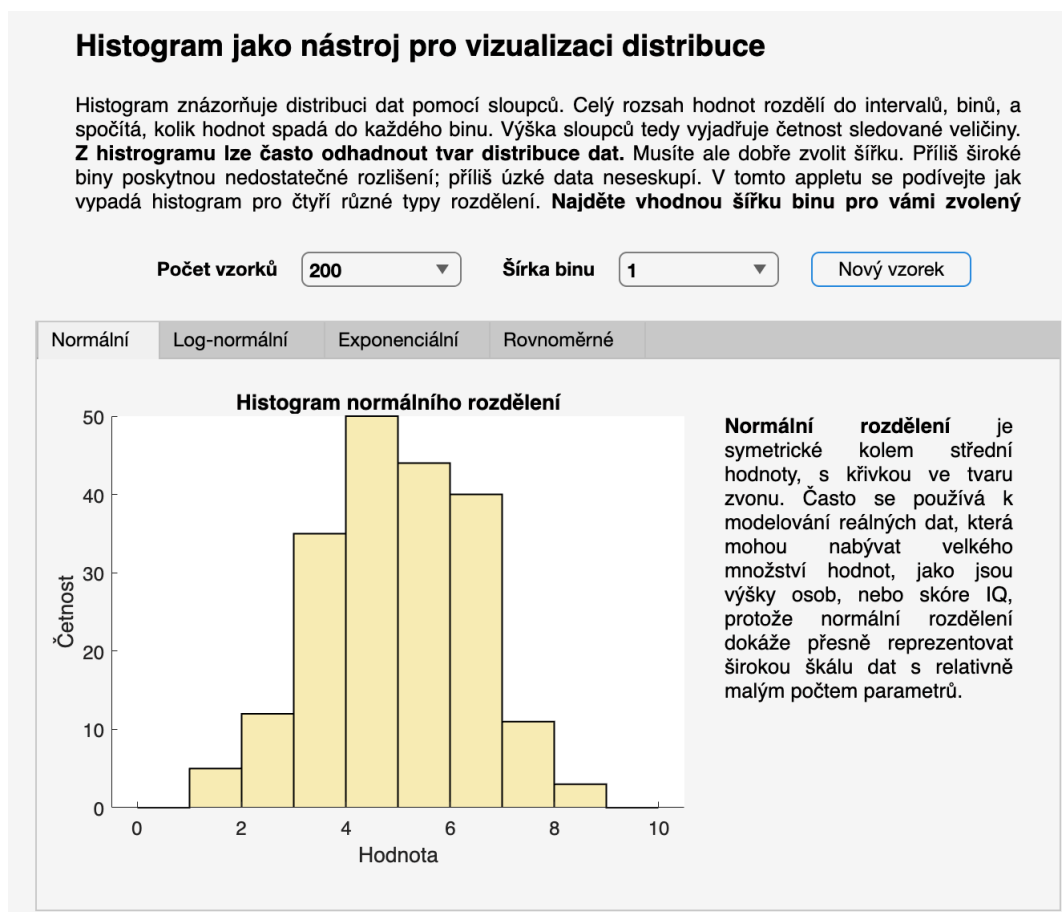


Figure 5 Příklad vizualizace normálního rozdělení v aplikaci.