

# Informatika 0

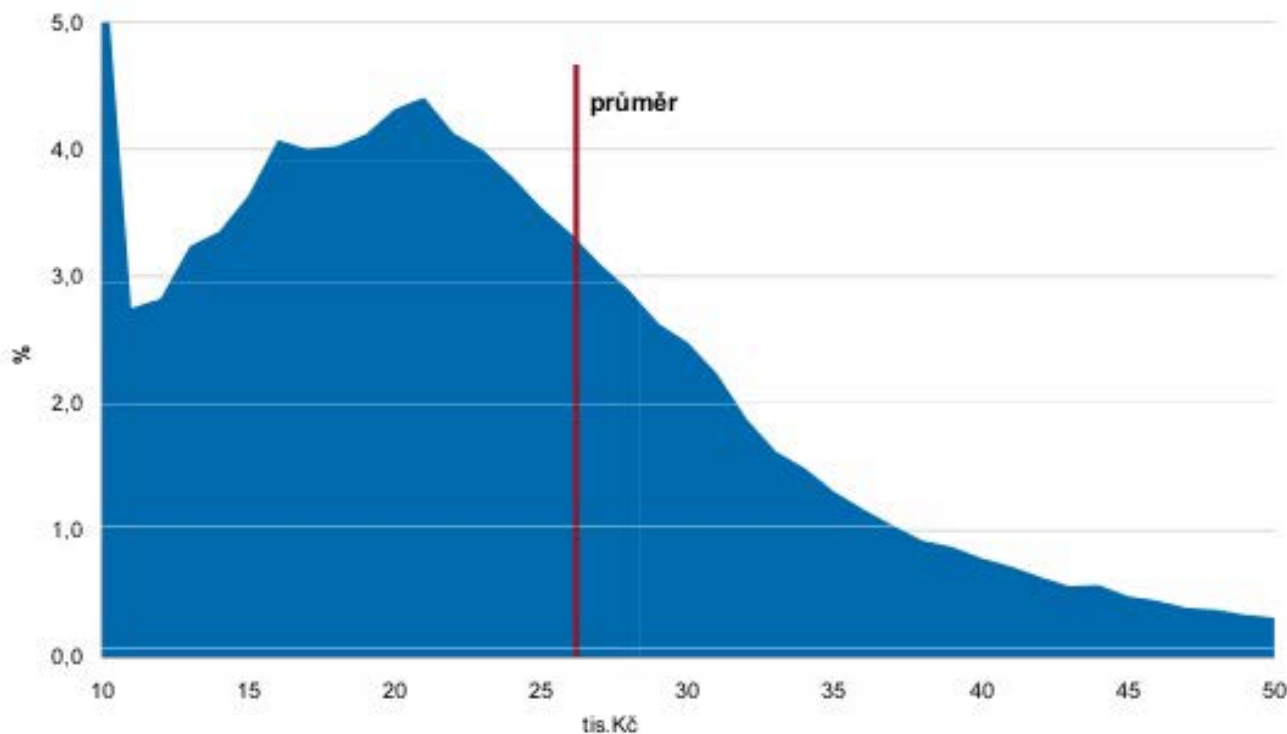
J. Blažej

7. část

# Statistické grafy

Zobrazení souboru dat (obvykle početnějšího), kdy nejde o reprezentaci každého bodu, ale zobrazení některých statistických vlastností celku

# Rozdělení pravděpodobnosti



Zdroj: ISPV

Dvě třetiny zaměstnanců jsou pod průměrnou mzdou

# Rozdělení pravděpodobnosti

- spojitá
  - normální (Gaussovo)
  - rovnoměrné
  - exponenciální
  - Maxwelllovo
  - ...
- diskrétní
  - Poissonovo
  - rovnoměrné
  - binomické
  - ...

# Popisná statistika

- soubor dat
- počet prvků
- minimum a maximum
  
- průměr a rozptyl
  - obecně míry centrální tendence (polohy)
  - a míry disperze (rozptýlenosti)
  
- modus a kvantily  
(medián, kvartily, decily, percentily)

(míry disperze) **CZ – en** (types of dispersions)

rozptyl

variance

střední kvadratická  
odchylka (fluktuace)  $\sigma^2$

směrodatná  
odchylka

standard deviation  
 $\sigma$  (rms, jen ve fyzice)  
jitter

# v ruštině je to jinak!

míra disperze = мера разброса

$\sigma^2$  – дисперсия

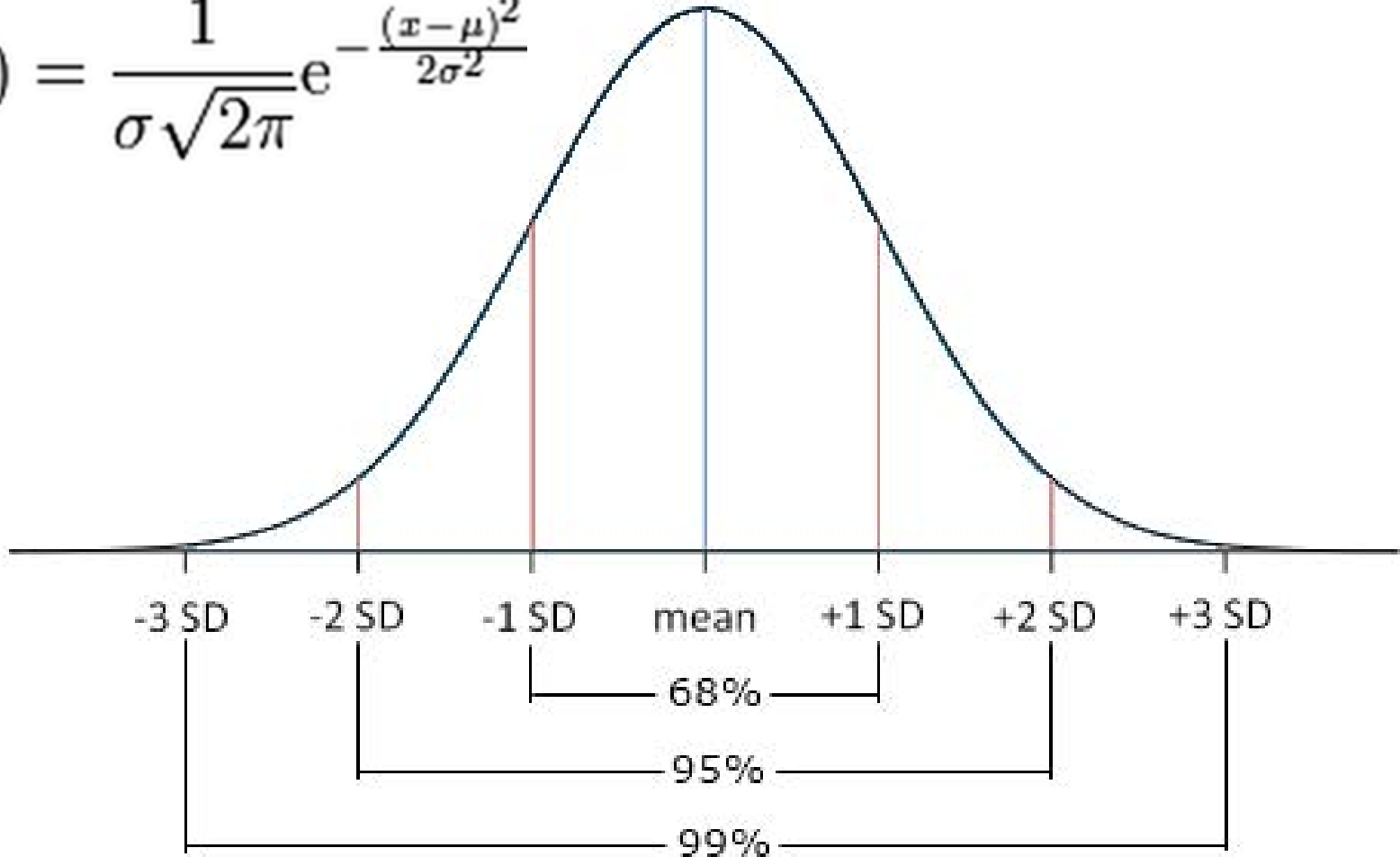
$\sigma$  – среднеквадратическое отклонение

výběrová směrodatná odchylka = стандартное отклонение

Tedy jsou používány velmi podobné termíny,  
ale v posunutých nebo zaměněných významech!

# Normální rozdělení

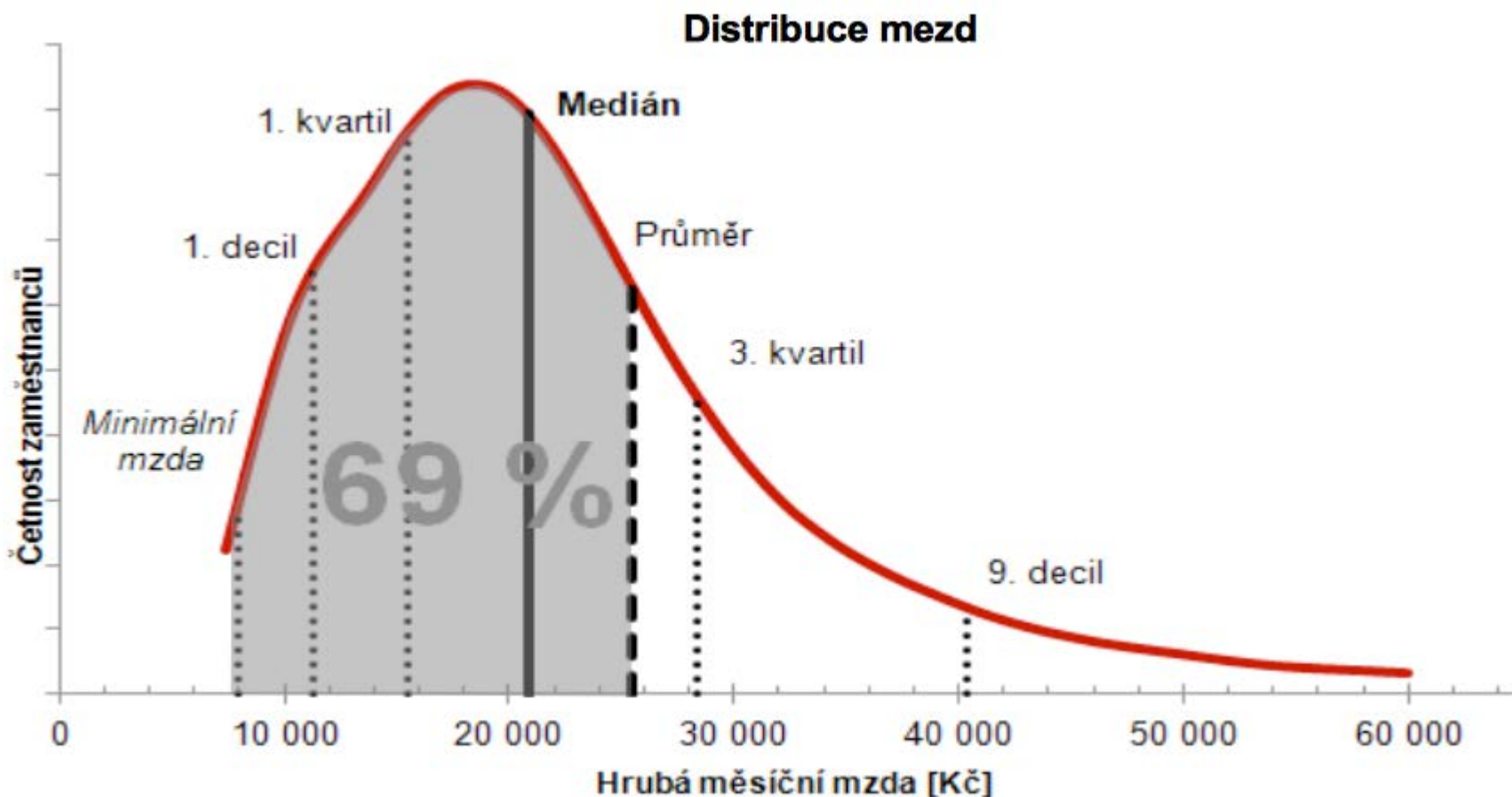
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



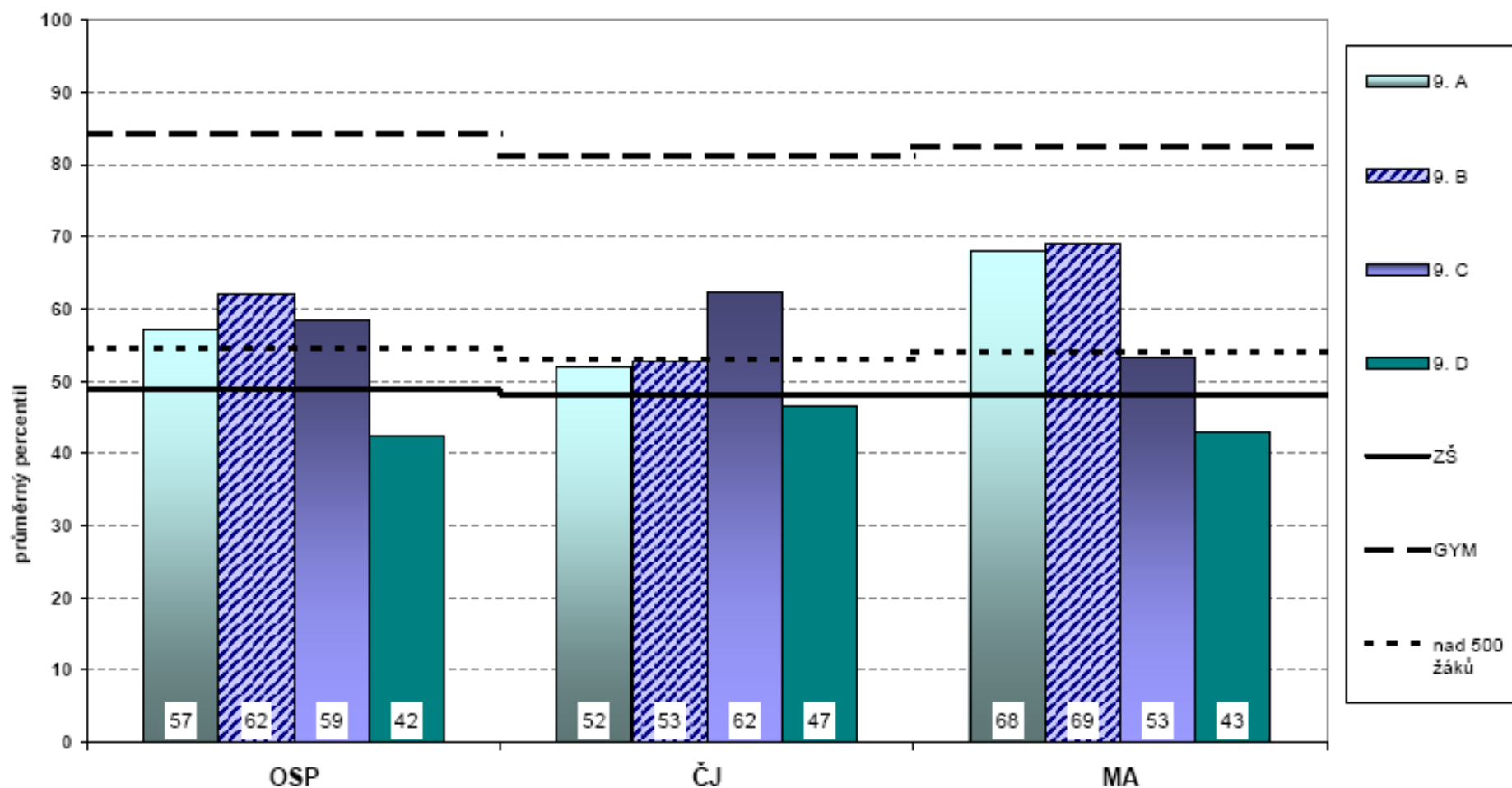
používanou mírou rozptýlenosti je také FWHM



# Popisná statistika - příklad

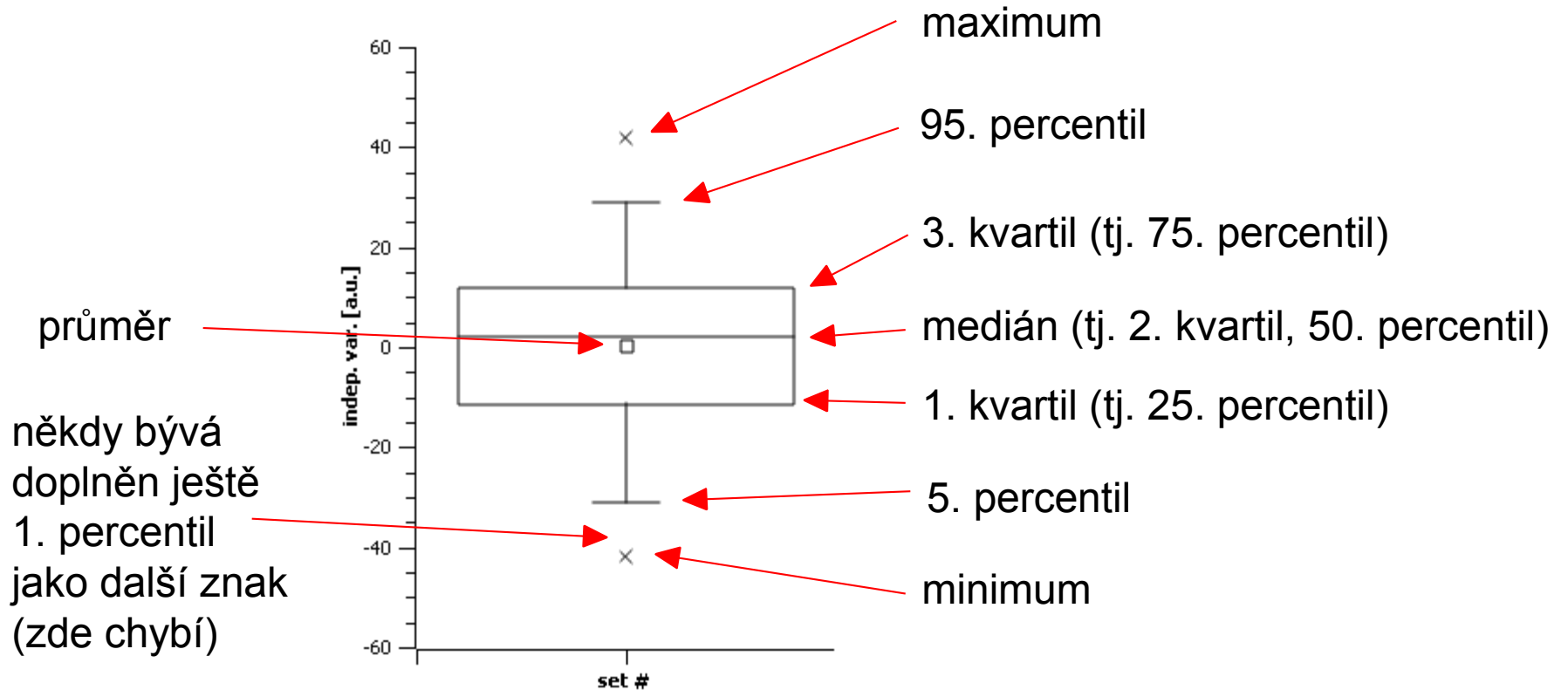


## Průměrný celkový percentil po jednotlivých třídách - 9. ročníky



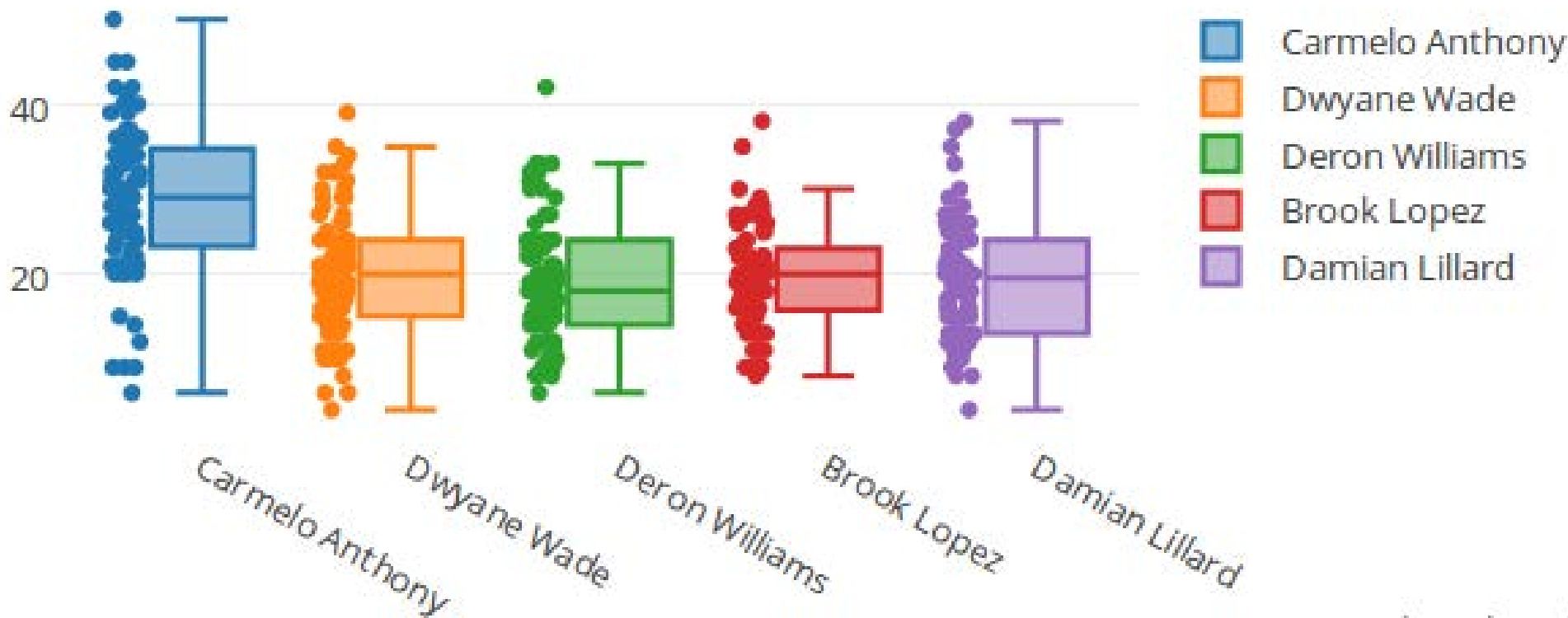
Graf znázorňuje průměrné celkové percentily všech tříd 9. ročníku vaší školy. Zároveň je zde pro porovnání vyznačen průměrný percentil všech žáků 9. ročníku ZŠ a GYM a dále průměrný percentil žáků škol odpovídající velikosti.

# kvantilový graf

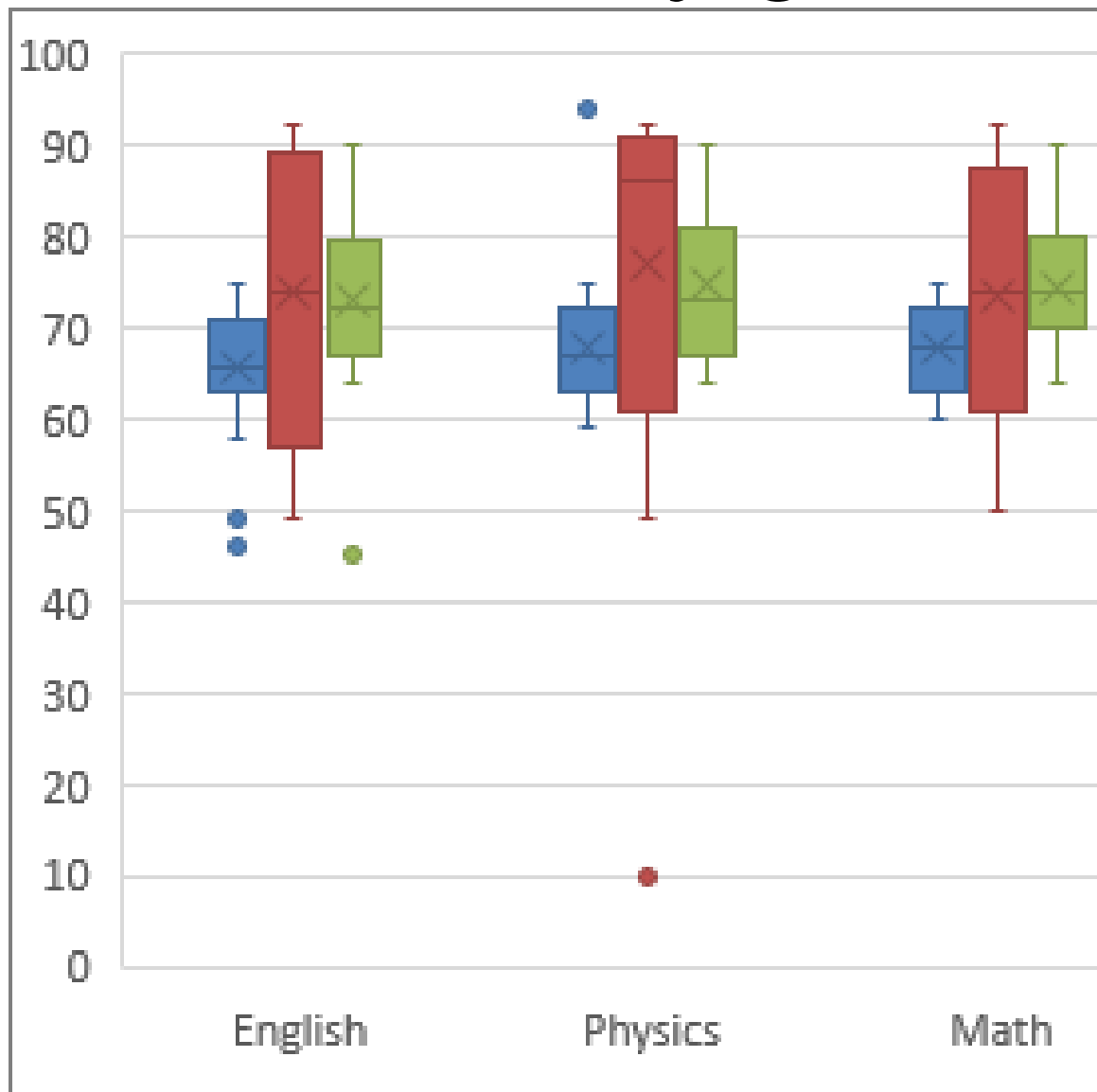


anglicky „boxplot“, odtud někdy česky „krabicový“ nebo „krabičkový“ graf

# kvantilový graf



# kvantilový graf



# histogram

- graf četnosti
- nezávislá proměnná je odvozena od intervalu hodnot původní závislé proměnné
- forma závisí na datech, velikosti buňky a počáteční hodnotě

# histogram

nejednoznačnost v názvosloví

- méně často může označovat jakýkoliv sloupcový graf
- v rámci tohoto předmětu je to vždy graf četnosti, ať je zobrazen v jakékoliv formě (nejčastěji právě jako sloupcový graf)

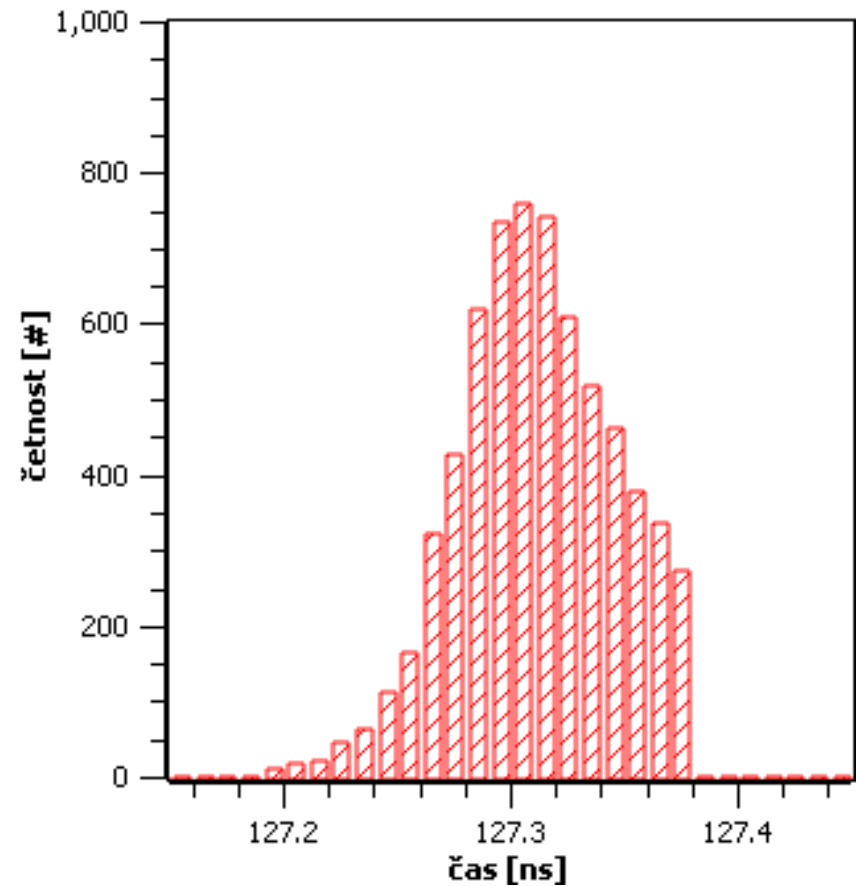
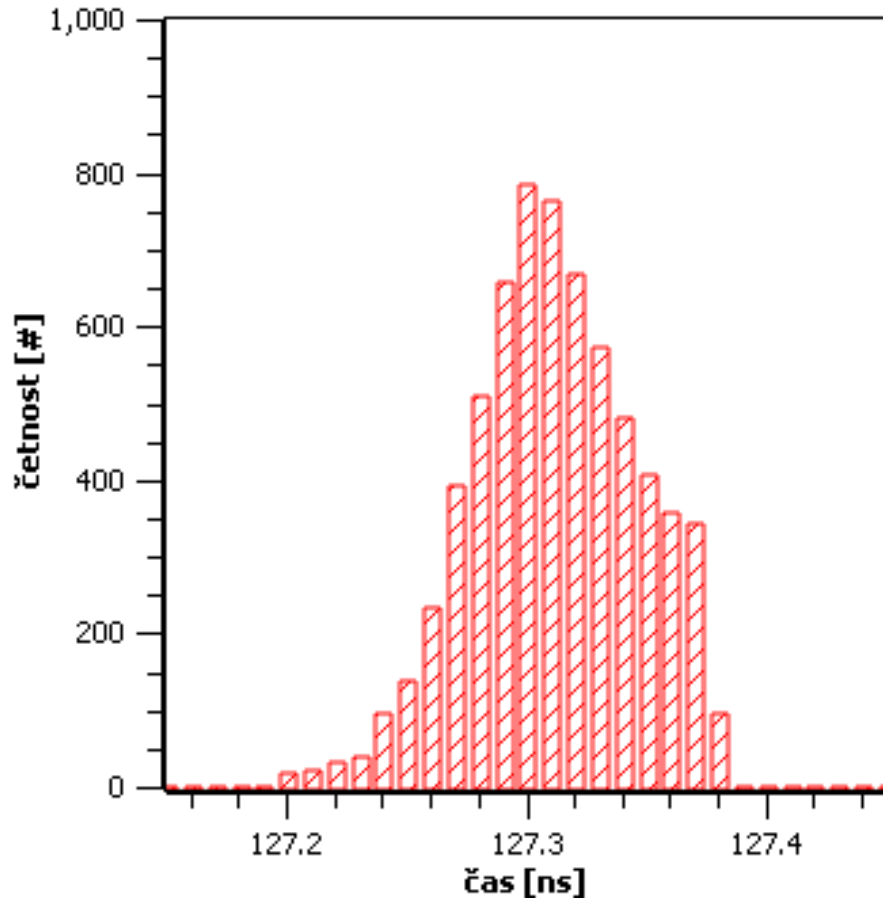
# histogram

- je pouze zobrazením existujících dat, změna formy histogramu nemění vlastnosti souboru dat



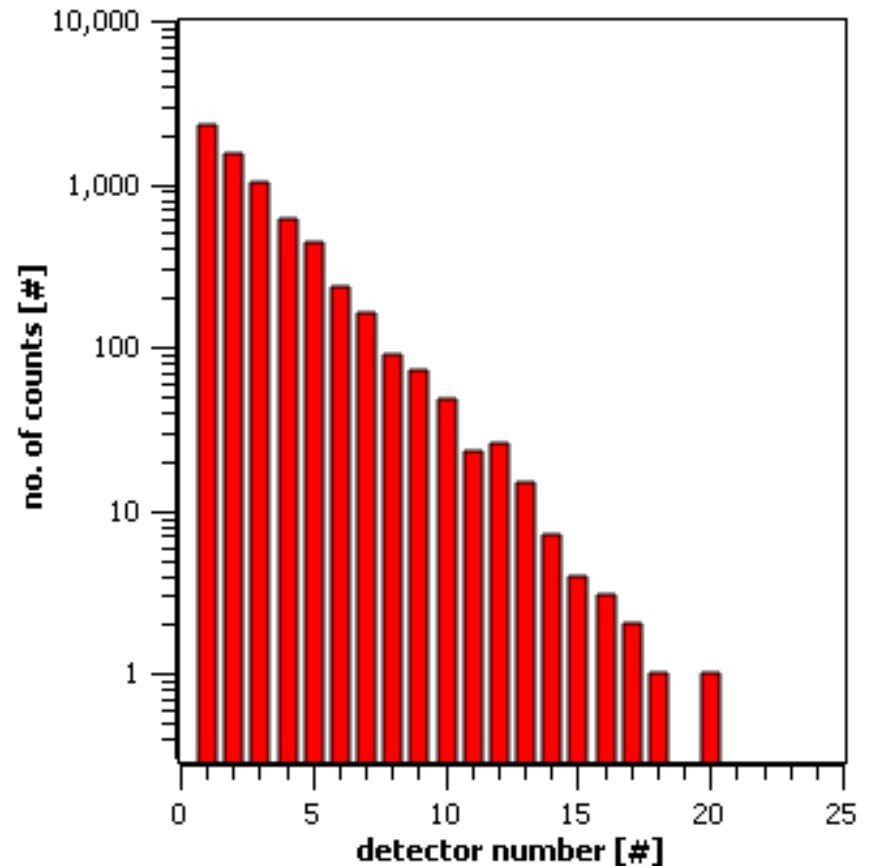
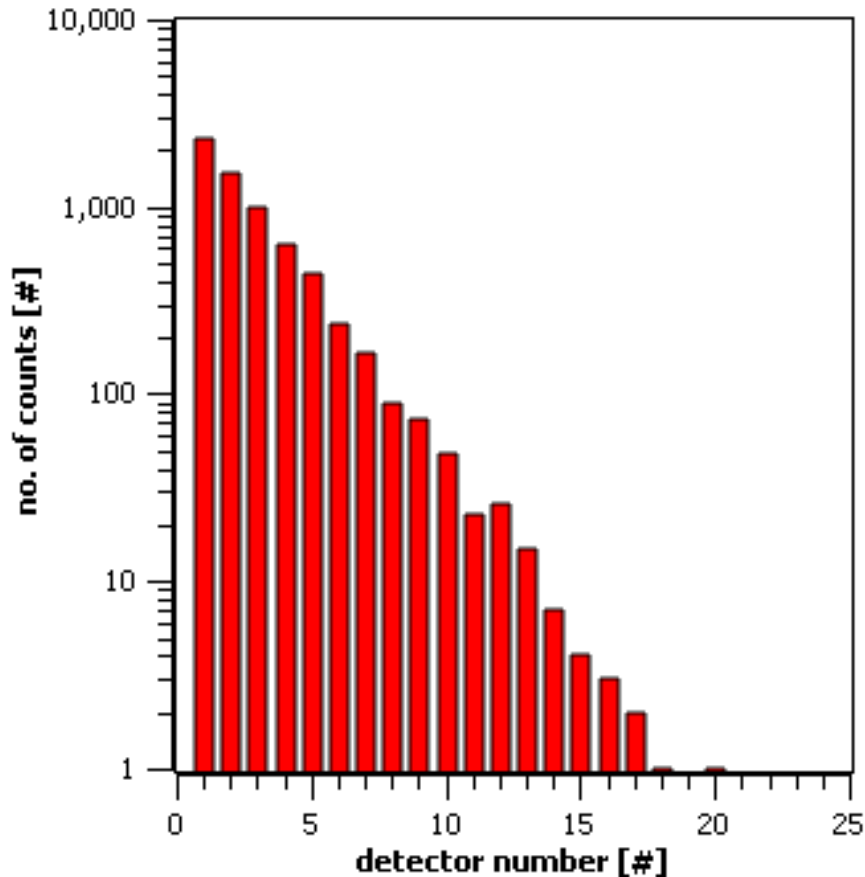
# histogram

- efekt posunutí počátku buňky



# histogram

- v semilogaritmickém zobrazení je nutné dbát na rozlišení stavů 0 a 1



# densitogram

- graf hustoty
- 2D analogie histogramu

# grafy z „hustých“ dat

- jde o situaci, kdy velikost značky spolu s množstvím dat a velikostí oblasti, ve které zobrazujeme neumožňuje najít kompromis, kdy by byl každý bod zobrazen

# grafy z „hustých“ dat

plovoucí průměr z 100 (zeleně) a 1000 (červeně) sousedních dat

