

# Rozpoznávanie obrazcov

## Štatistika 2

17.3.2014

# Číselné charakteristiky náhodnej premennej

- Momenty
- Kvantily

# Momenty

- Začiatočný moment r-tého rádu

$$\mu_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx.$$

- centrálné momenty, kde od hodnôt pôvodných dát odčítame ich priemer

$$\mu'_r = E((X - E(X))^r) = \int_{-\infty}^{\infty} (x - E(X))^r f(x) dx.$$

- centrálné momenty druhého, tretieho a štvrtého rádu

# Stredná hodnota

- hodnota, „okolo“ ktorej sa sústred'ujú všetky hodnoty náhodnej premennej
- Priemer

$$\mu = \int_{-\infty}^{\infty} x f(x) dx.$$

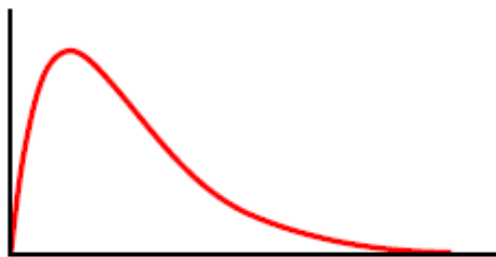
- $A = [0 \ 1 \ 1; 2 \ 3 \ 2; 1 \ 3 \ 2; 4 \ 2 \ 2]$
- $M = \text{mean}(A)$
- $M = \text{mean}(A, 2)$

# Rozptyl

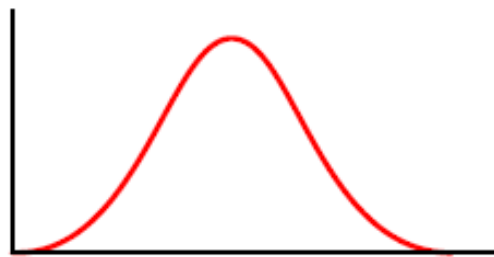
- variancia, disperzia, stredná kvadratická odchýlka
  - centrálny moment druhého rádu
  - variabilita hodnôt dát okolo strednej hodnoty
  - najčastejšie používaná miera variability
  - určuje strednú kvadratickú odchylku jednotlivých nameraných hodnôt od výberového priemeru
- 
- $u = [6 \ 1 \ 3 \ -7 \ 2 \ 5 \ 8 \ 0 \ -1]$
  - $y = \text{var}(u)$

# Šikmosť

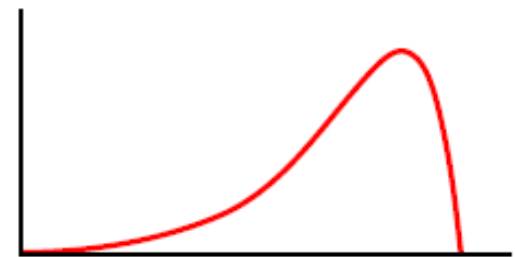
- podiel tretieho centrálneho momentu a tretej mocniny smerodajnej odchýlky
  - miera symetrie rozdelenia pravdepodobnosti vzhľadom na normované (štandardné) normálne rozdelenie  $N(0,1)$



(a) Pozitívne zošikmené rozdelenie.



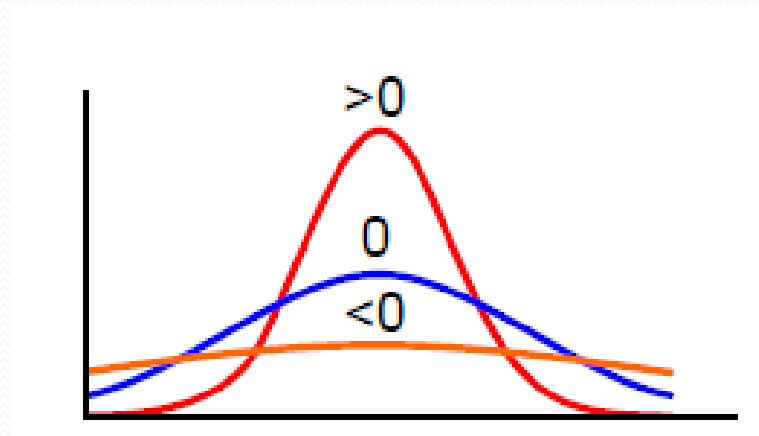
(b) Symetrické rozdelenie.



(c) Negatívne zošikmené rozdelenie.

# Špicatost'

- Podiel štvrtého centrálneho momentu a štvrtej mocniny smerodajnej odchýlky
  - relatívna plochosť rozdelenia pravdepodobnosti vzhľadom na normované (štandardné) normálne rozdelenie  $N(0,1)$



# Kvantily

- Kvantil  $Q_p$  je hodnota, pre ktorú platí, že pravdepodobnosť, že náhodná veličina  $X$  má hodnotu menšiu ako  $Q_p$ , je  $100p\%$

$$P(X < Q_p) = p$$

$$F(Q_p) = p$$

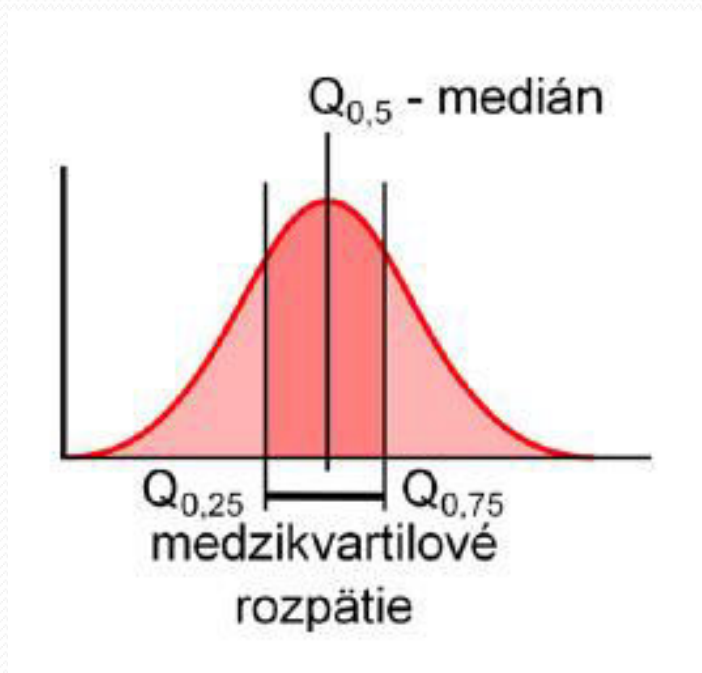


# Medián

- kvantil deliaci súbor na dve polovice
- $A = [0 \ 1 \ 1; 2 \ 3 \ 2; 1 \ 3 \ 2; 4 \ 2 \ 2]$
- $M = \text{median}(A)$
- $M = \text{median}(A, 2)$

# Kvartil

- Kvartily delia súbor na štvrtiny.
- $Q_{0,25}$  (dolný kvartil),  $Q_{0,5}$  a  $Q_{0,75}$  (horný kvartil)



# Decil

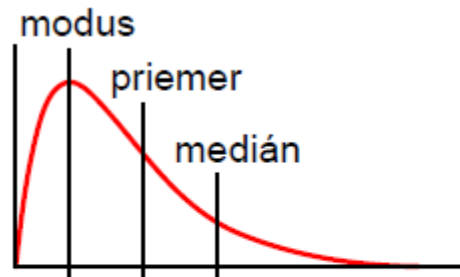
- Decily delia súbor na desatiny
- $Q_{0,1}, Q_{0,2}, \dots, Q_{0,9}$

# Percentil

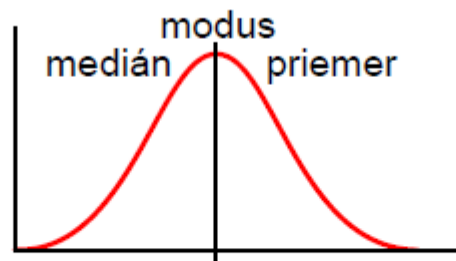
- Percentily delia súbor na stotiny
- $Q_{0,01}, Q_{0,02}, \dots, Q_{0,99}$

# Modus

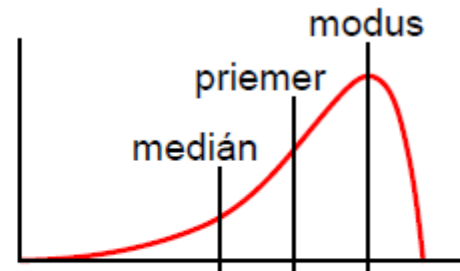
- najčastejšie sa opakujúca hodnota v súbore
- $X = [3\ 3\ 1\ 4; 0\ 0\ 1\ 1; 0\ 1\ 2\ 4]$
- $M = \text{mode}(X)$
- $M = \text{mode}(X,2)$



(a) Pozitívne zošikmené rozdelenie.



(b) Symetrické rozdelenie.



(c) Negatívne zošikmené rozdelenie.

# Kovariancia

- Kovariancia
  - závislosť medzi dvomi náhodnými veličinami
  - $\text{cov}(x)$ 
    - Vracia variancie, ak  $x$  je vektor

# Kovariančná matica

- Symetrická
- Diagonála – variancie

$$\Sigma_x = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{bmatrix}$$

- $A = [-1 \ 1 \ 2 ; -2 \ 3 \ 1 ; 4 \ 0 \ 3]$
- $\text{cov}(A)$
- $\text{diag}(\text{cov}(A))$  vráti variancie
- $\text{sqrt}(\text{diag}(\text{cov}(X)))$  vracia vektor štandardných odchýlok

# Korelácia

- závislosť medzi dvomi alebo viacerými znakmi v štatistickom súbore alebo medzi dvoma alebo viacerými náhodnými veličinami

$$\rho(X, Y) = \text{cov}(\text{norm } X, \text{norm } Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

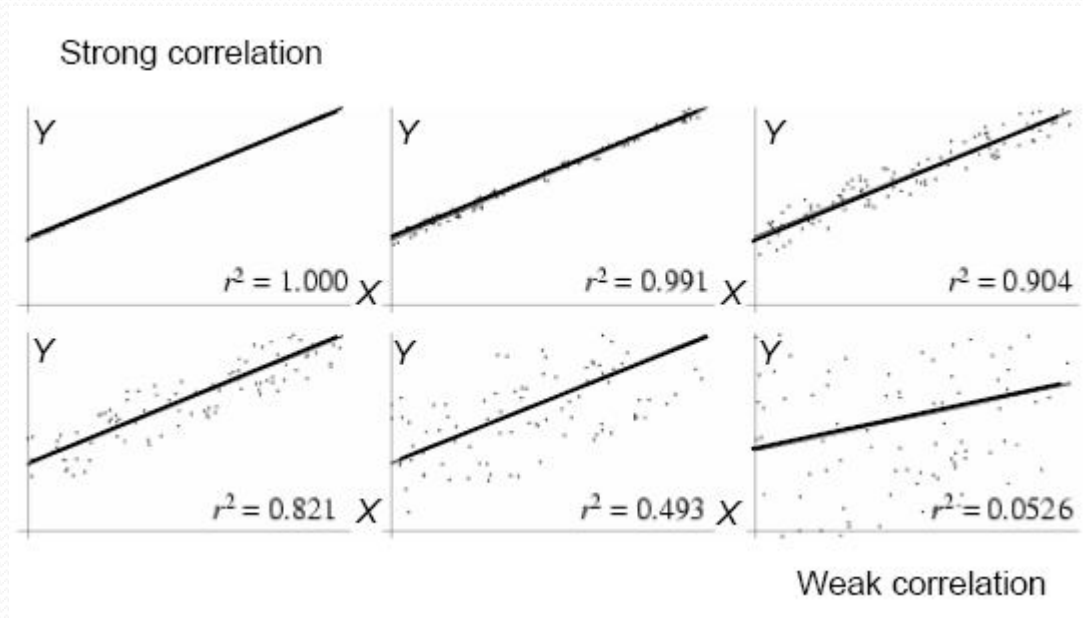
- $\text{RHO} = \text{corr}(X)$ 
  - Vracia  $p \times p$  maticu, ktorá obsahuje lineárne korelačné koeficienty medzi každým párom stĺpcov  $n \times p$  vstupnej matice  $X$



# Štatistická závislosť

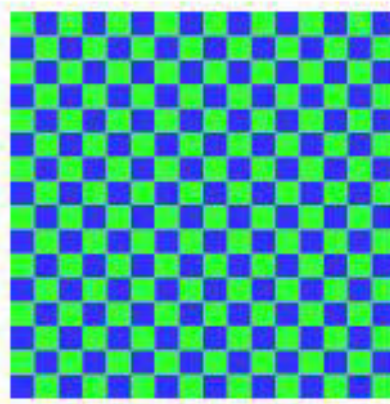
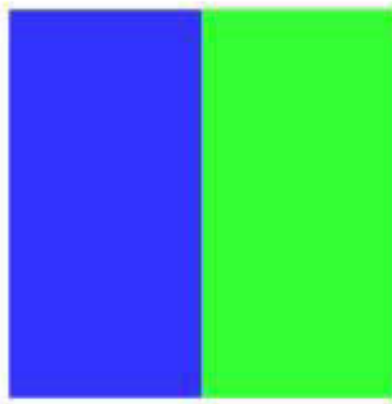
- Korelačný koeficient
  - hodnoty medzi -1 a 1
  - MATLAB

$$\text{CorrCoef}\{X, Y\} = \frac{1}{N} \sum_{i=1}^N \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_x \sigma_Y}$$



# Entropia

- Štatistika
  - Miera neistoty výsledku
- Spracovanie obrazu
  - Miera množstva informácie



# Entropia

- $H(Y) = -\sum \log(P[Y = y])P[Y=y]$
- $P[Y=y]$  – pravdepodobnosť
  
- Podmienená entropia
  - $H(Y|X)$

# Entrópia - príklad

	Druh stromu X	Nakazený hnilobou koreňov
1	Jabloň	Áno
2	Orech	Nie
3	Slivka	Áno
4	Jabloň	Nie
5	Jabloň	Nie
6	Slivka	Áno
7	Orech	Nie
8	Jabloň	áno

# Entrópia - MATLAB

- $E = \text{entropy}(I)$ 
  - vracia hodnotu reprezentujúcu entrópiu v grayscale obraze
  - Ak  $I$  je RGB/HSV/... Vracia hodnoty ako pre multidim. grayscale obrázkov
  - výpočet:
    - $-\text{sum}(p.*\log_2(p))$
    - $p$  contains the histogram counts returned from `imhist`

```
I = imread('circuit.tif');
```

```
imshow(I);
```

```
J = entropy(I);
```

# Vzájomná informácia

- Udáva ako sa zníži neurčitost' premennej  $Y$ , ak poznáme premennú  $X$
- $I(Y,X) = H(Y) - H(Y|X) = H(X) - H(X|Y)$
- $X, Y$  nezávislé –  $I(Y,X) = 0$
- $I(Y,Y) = H(Y)$

# MI matlab

- <http://www.mathworks.com/matlabcentral/fileexchange/13289-fast-mutual-information-of-two-images-or-signals>
- <http://www.mathworks.com/matlabcentral/fileexchange/36538-very-fast-mutual-information-between-two-images>