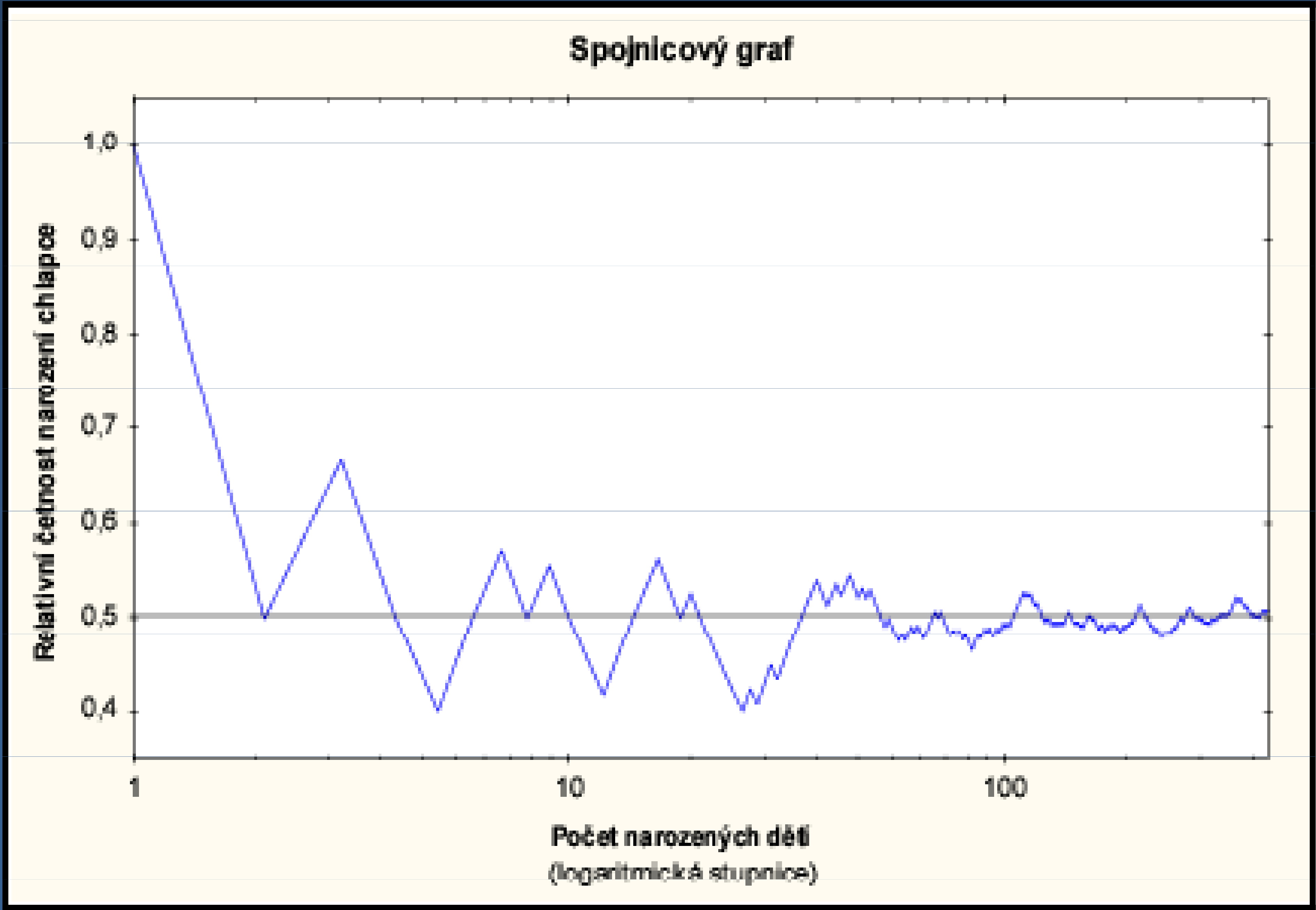
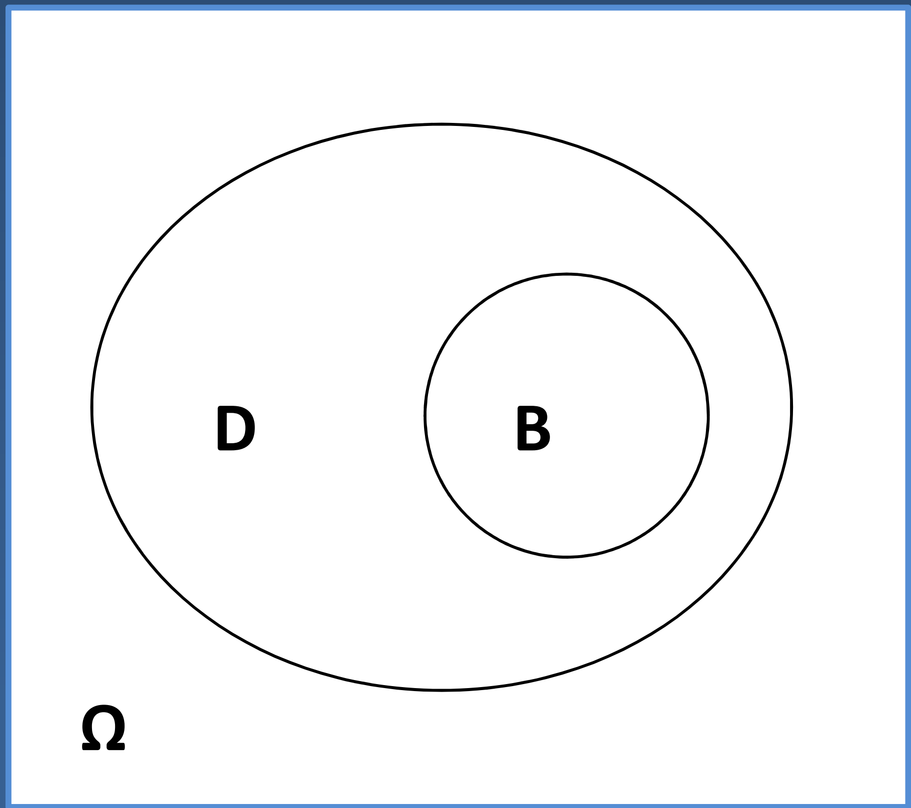
5. Pravděpodobnost

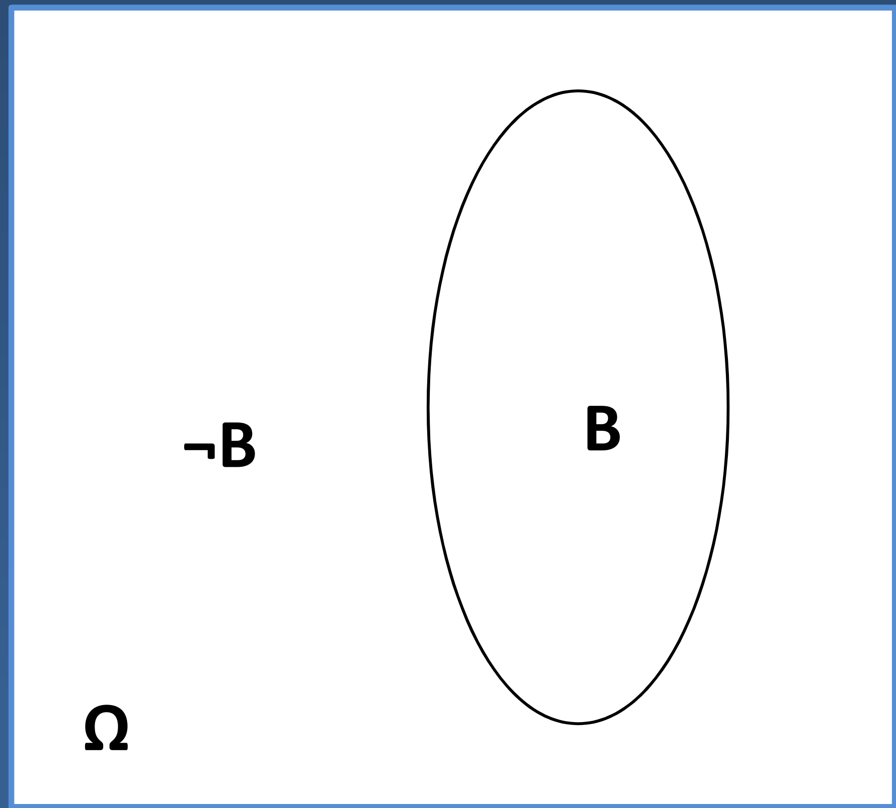
- Náhodný pokus – výsledek není jednoznačně předurčen podmínkami, mnohokrát opakovatelné (hod kostkou, mincí, tahání losů)
- Náhodný jev – tvrzení o výsledku, po provedení pokusu lze rozhodnout, zda platí nebo ne (*A* – „narození chlapce“, ¬*A* – „narození dívky“)
- Četnost *r/n* – *r*- počet narozených chlapců, *n* – rozsah výběru
- r – absolutní četnost
- *r/n* – relativní četnost výskytu náhodného jevu *A* ve výběru o rozsahu *n*


Spojnicový graf
Relativní četnost narození chlapce
1,0
0,9
0,8
0,7
0,6
0,5
0,4
1
10
100
Počet narozených dětí
(logaritmická stupnice)

$$B \subset D \Rightarrow P(B) \leq P(D)$$

$$P(\neg B) = 1 - P(B)$$





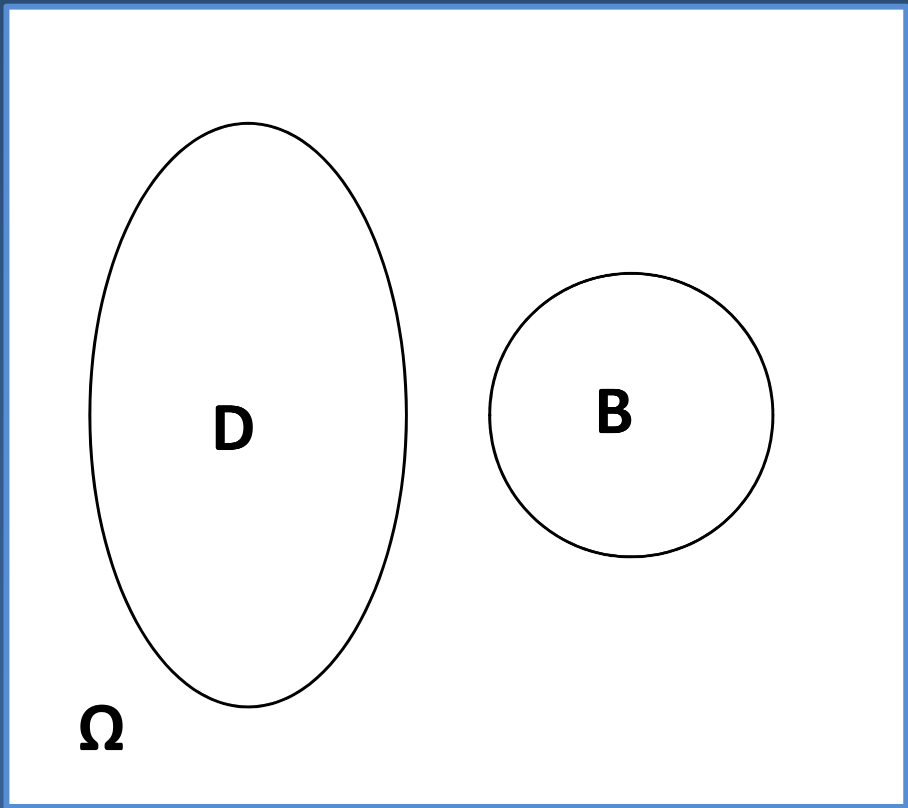
Obecný vzorec:

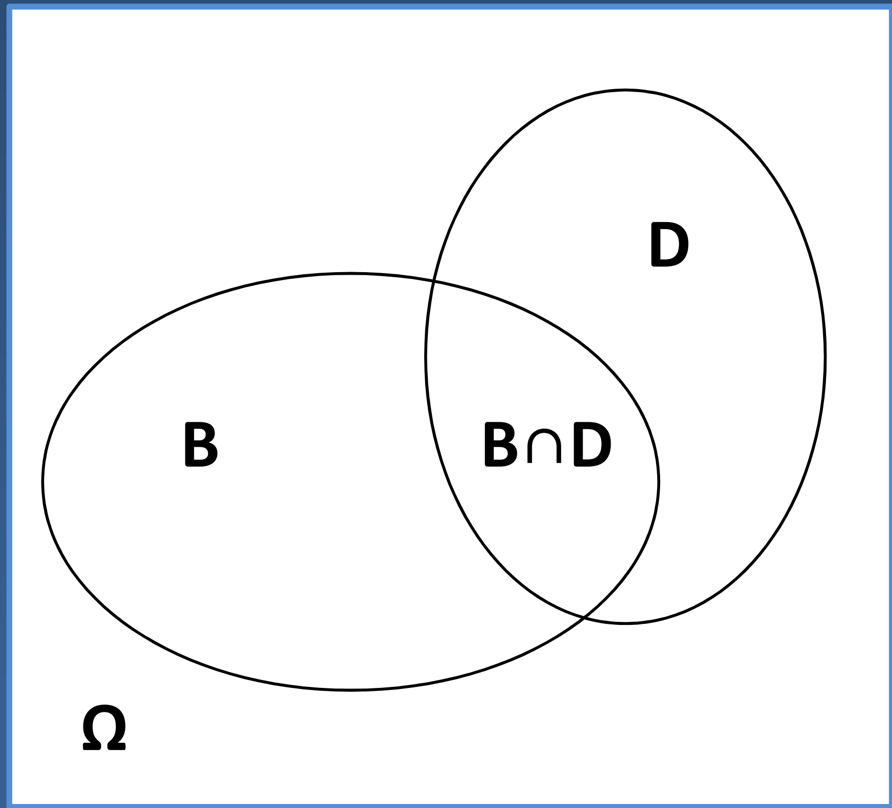
$$B \cap D = 0 \Rightarrow$$

$$P(B \cup D) = P(B) + P(D)$$

$$P(B \cup D) =$$

$$P(B) + P(D) - P(B \cap D)$$





Definice pravděpodobnosti

- Jev A je charakterizován číslem P(A) = pst náhodného jevu A (míra častosti výskytu tohoto jevu)
- Vlastnosti psti:
- 0 ≤ P(A) ≤ 1
- P(Ω) = 1 => jistý jev
- P(0) = 0=> nemožný jev
- P(¬A) = 1 – P(A)

$$B \cap D = 0 \Rightarrow P(B \cup D) = P(B) + P(D)$$

$$P(B \cup D) = P(B) + P(D) - P(B \cap D)$$

$B \subset D \Rightarrow P(B) \leq P(D)$ podjev

- Při mnohonásobném opakování se relativní četnost A jen nepatrně liší od P(A)

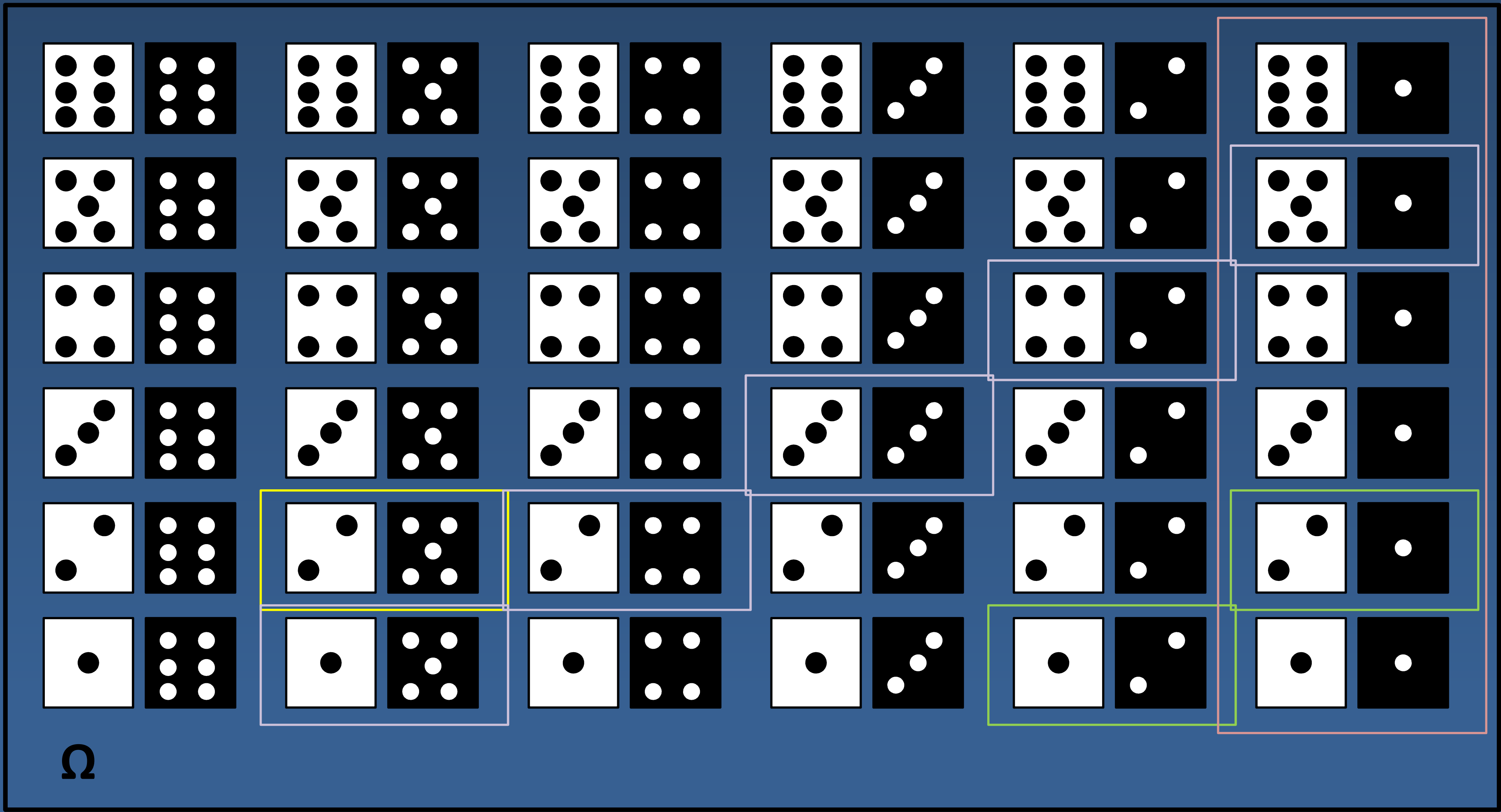
Podmíněná pravděpodobnost

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Podmíněná pravděpodobnost jevu *A* vzhledem k jevu *B*
- Pst současného výskytu jevu *A a B:*

$$P(A \cap B) = P(A|B)P(B)$$

Příklad 1: házení 2 vyváženými kostkami



P(černá 5, bílá 2) = 1/36

P(na černé padne 1) = 6/36

P(součet = 3) = 2/36

P(součet = 6) = 5/36

Příklad 2: rodina; 3 sourozenci

ω_i	D	B	$B \cap D$	$B \cup D$	C
(m,m,m)					
(f,m,m)					
(m,f,m)					
(f,f,m)					
(f,f,f)					
(m,f,f)					
(f,m,f)					
(m,m,f)					

D nejmladší je dívka

B v rodině je jediná dívka

$B \cap D$ jediná dívka je nejmladší

$B \cup D$

C nejstarší je hoch

Kvalita screeningového testu

- Jevy D^+ osoba nemoc má, D^- osoba nemoc nemá
- Pozitivní výsledek testu T^+, negativní výsledek T^-
- P(D^+) - prevalence

	Nemoc		
Výsledek testu	Přítomna (D^+)	Nepřítomna (D^-)	Celkem
T^+	a	b	$a+b$
T^-	c	d	$c+d$
Celkem	$a+c$	$b+d$	n

Kvalita screeningového testu

- Senzitivita (*SE*) = $P(T^{+} | D^{+})$ $$SE = \frac{a}{a+c}$$
- Specificita (*SP*) = $P(T^{-} | D^{-})$ $$SP = \frac{d}{b+d}$$
- Falešná pozitivita (*FP*) = $\frac{b}{b+d}$
- Falešná negativita (*FN*) = $\frac{c}{a+c}$
- *SE* + *FN* = 1, *SP* + *FP* = 1

Kvalita screeningového testu

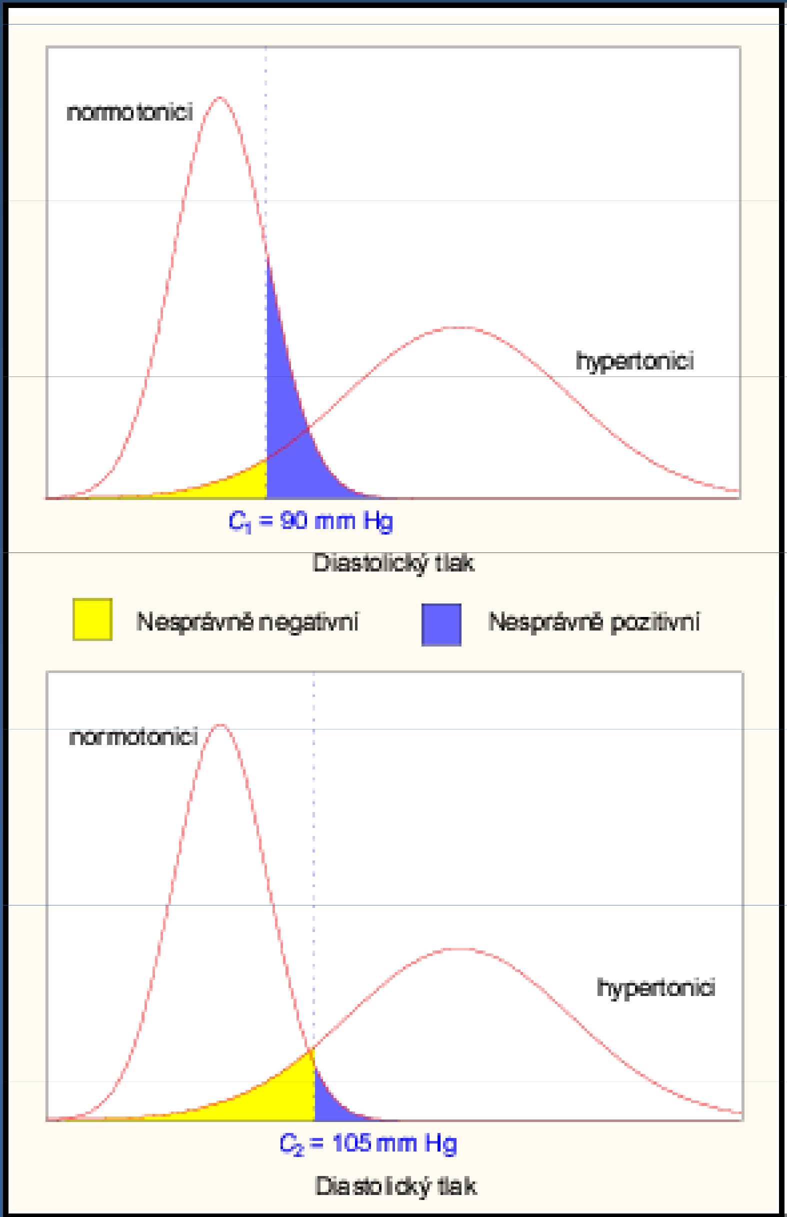
- Pozitivní prediktivní hodnota (*PV*⁺) $P(D^+ | T^+)$

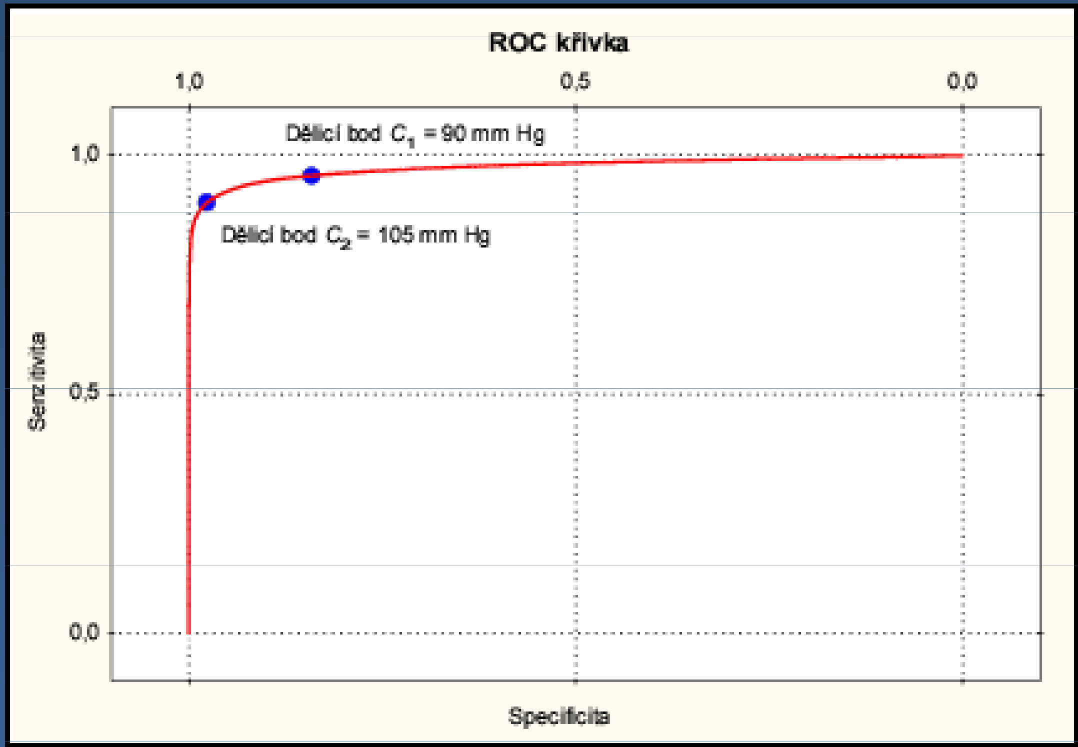
$$PV^+ = \frac{a}{a+b}$$

- Negativní prediktivní hodnota (*PV*⁻) $P(D^- | T^-)$

$$PV^- = \frac{d}{c+d}$$

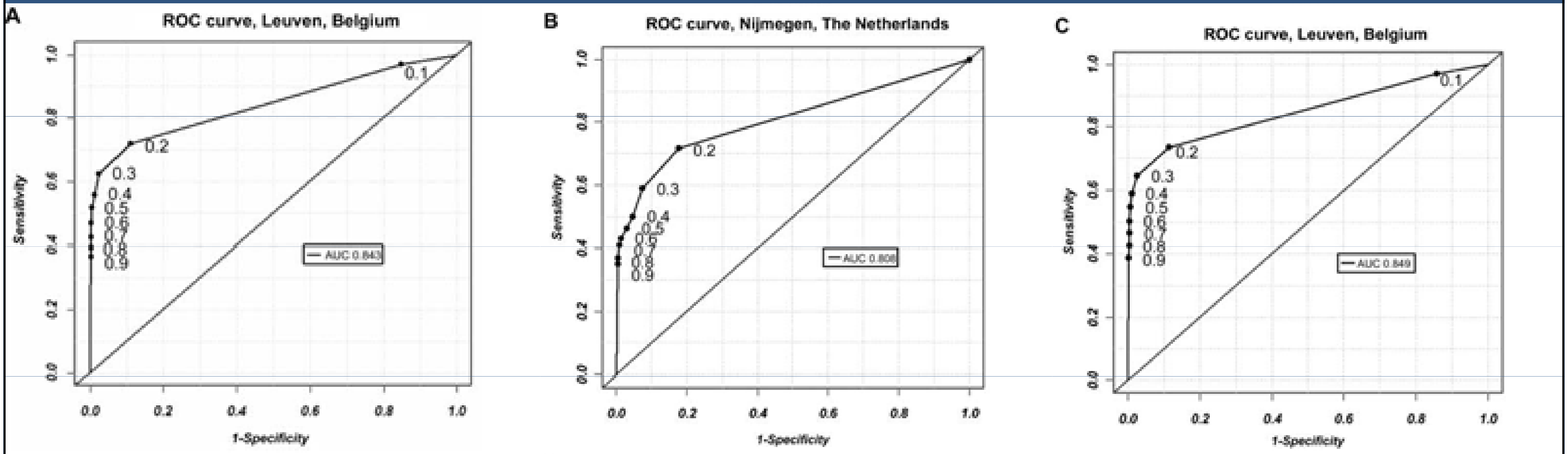
ROC (Receiver Operating Charakteristic) křivka





ROC v praxi

- Detekce aspergilózy
- Měření OD (optical density)



Příklad

- Vzácným onemocněním onemocní 1 osoba z 1000
- Existuje test s vlastnostmi: u nemocného je výsledek testu pozitivní v 99% případů (SE = 99%)
- 2% zdravých má výsledek testu pozitivní (FP = 2%)
- Jaká je pst, že osoba s pozitivním výsledkem testu je skutečně nakažená zkoumanou nemocí?

Příklad

- $P(D^+)=0,001$ – 1 pacient z 1000 trpí onemocněním
- $P(T^+|D^+)=0,99$ – pst pozitivního testu u nemocného
- $P(T^+|D^-)=0,02$ – pst falešně pozitivního výsledku u zdravého
- $P(D^+|T^+)=a/(a+b)$
- $P(D^+ \cap T^+)=P(T^+|D^+)P(D^+)=0,99x0,001=0,00099$
- $P(D^- \cap T^+)=P(T^+|D^-)P(D^-)=0,02x0,999=0,01998$

	D^+	**D^-**		
T^+	0,00099	0,01998	0,02097	$P(T^+)$
T^-	0,00001	0,97902	0,97903	$P(T^-)$
	0,001	0,999	1	
	P(D+)	P(D-)		

	D^+	**D^-**	
T^+	1	20	21
T^-	0	979	979
	1	999	1000

a/(a+b)=0,00099/0,02097=0,0472