



STATISTIKA

© Biochemický ústav LF MU (V.P.) 2007

**Účelem není znát vzorce a výpočty
(to by bylo zcela zbytečné),
ale vědět, co to znamená !**

**The purpose is not the knowledge of formulas
and calculations (it would be useless),
however to know what it means !**

Některé technické jevy → konstantnost → „jistota“



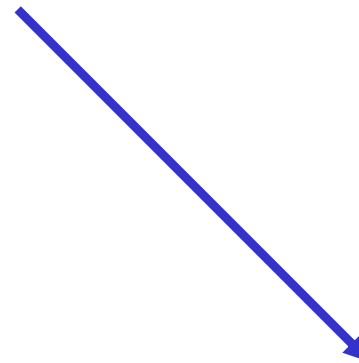
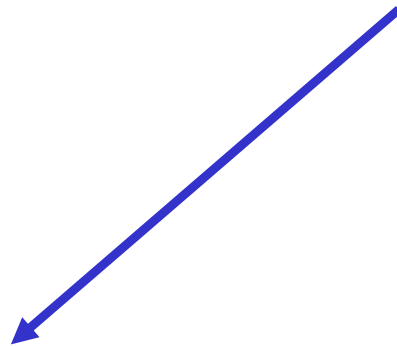
Biologické jevy → variabilita → pravděpodobnost



živý organismus → biologická variabilita



rozdělení četnosti biologických jevů



**symetrické,
„normální rozdělení“
„Gaussovské rozdělení“**

asymetrické

Carl Friedrich Gauß (1777 – 1855)

německý matematik

the German mathematician



Gaussova křivka

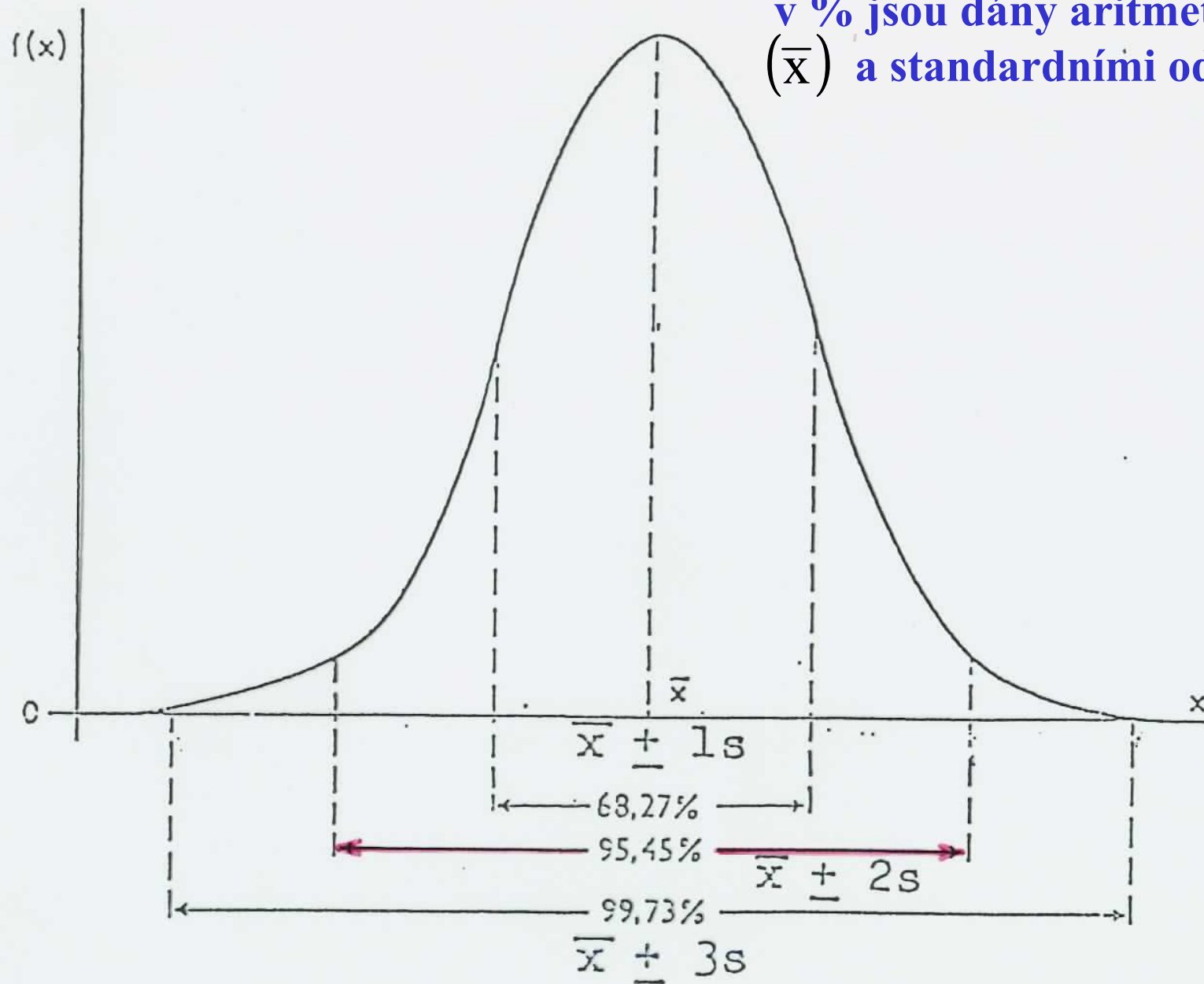
the Gaussian curve



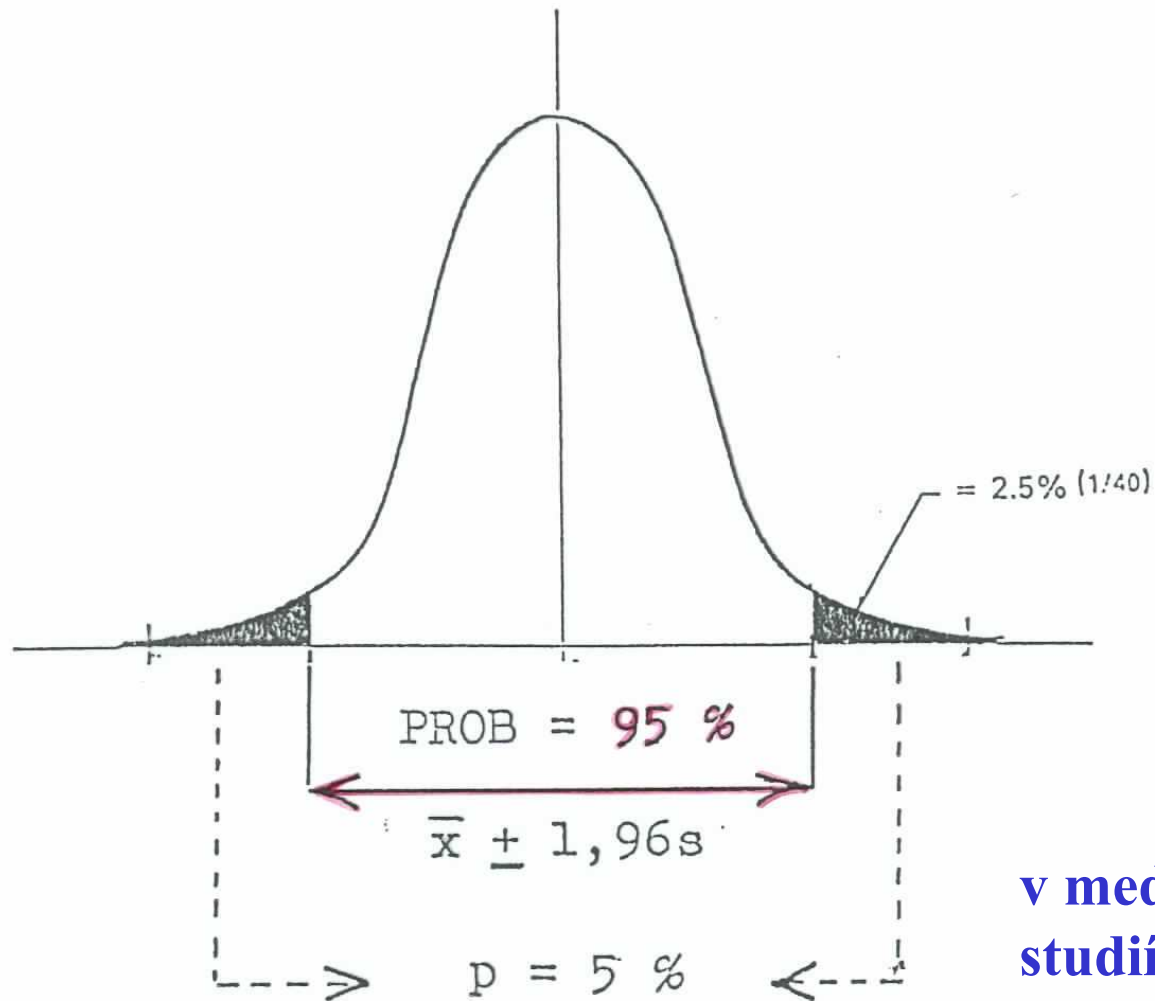
**Křivka rozložení funkce
pravděpodobnosti je symetrická
a zvonovitého tvaru
(= „normální rozdělení“,
„Gaussovské rozdělení“)**

**The curve of the probability
density function is symmetrical
and bell-shaped
(= „normal distribution“,
„the Gaussian distribution“)**

Intervaly pravděpodobnosti rozložení
v % jsou dány aritmetickým průměrem
(\bar{x}) a standardními odchylkami (s) .



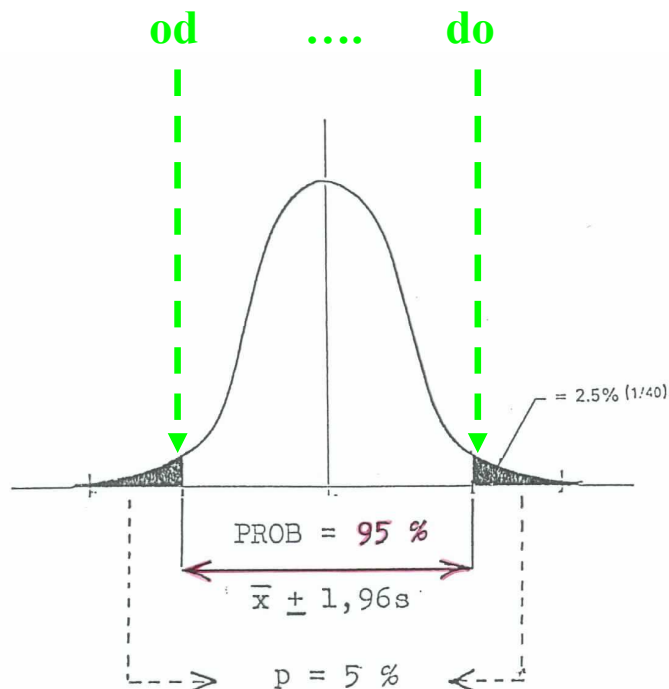
„Normální hodnoty“ :



v medicínských a biologických studiích je to dohodnuté rozpětí hodnot vymezené 95 % intervalem spolehlivosti

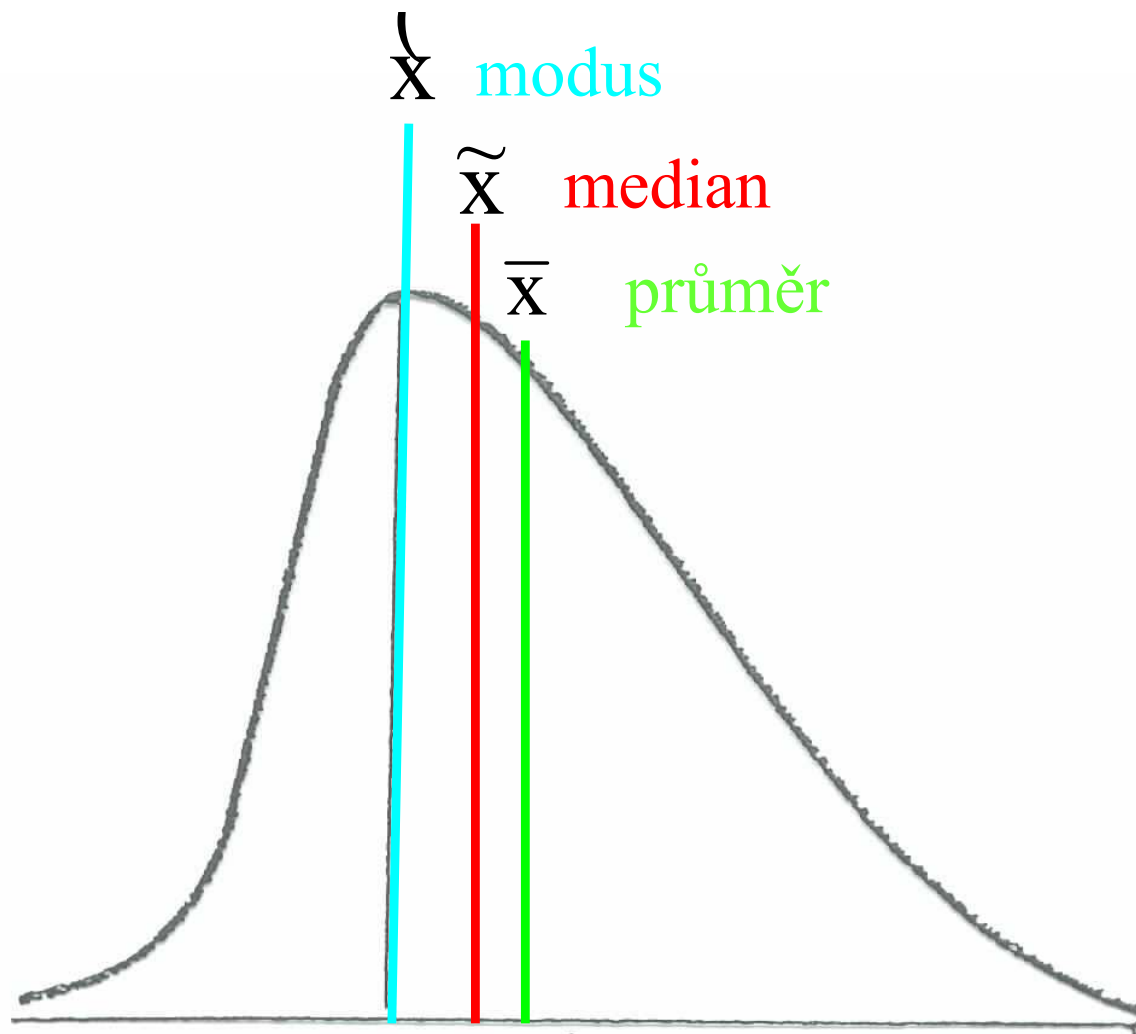
(oboustranně ohraničený interval spolehlivosti)

„Normální hodnoty“ („referenční hodnoty“):



Podle běžné konvence referenční hodnoty zahrnují celou populaci. Interval je však ohraničen oboustranně 2,5 % pásmem očekávaných hodnot. Ve skutečnosti tedy 5 % výsledků „normální“ zdravé populace bude ležet mimo referenční hodnoty.

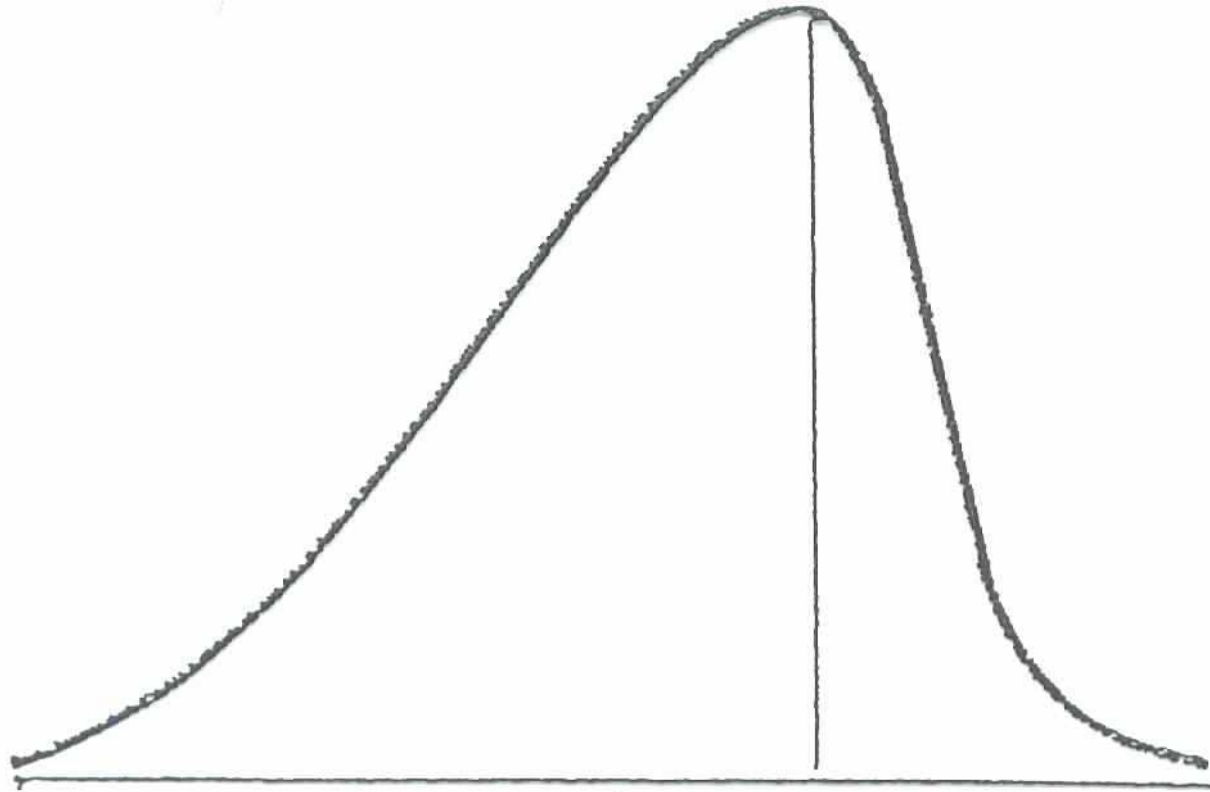
Nesymetrické rozdělení :



levostranně asymetrické rozdělení,
„logaritmicko-normální“ (lognormální) rozdělení

Nesymetrické rozdělení :

$$\overset{\curvearrowright}{\bar{x}} > \tilde{x} > \bar{x}$$

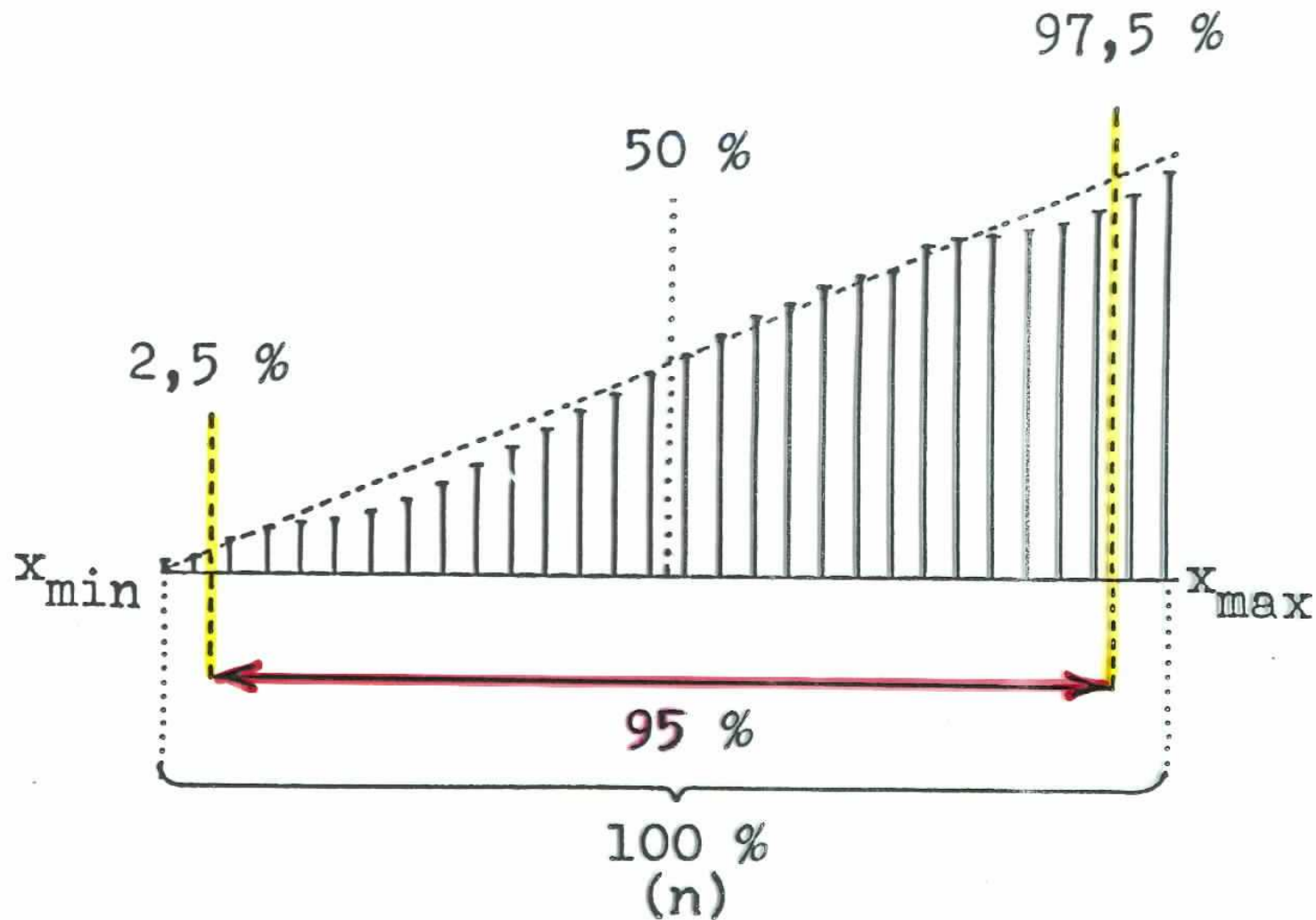


pravostranně asymetrické rozdělení,
„logaritmicko-normální“ (lognormální) rozdělení

u symetrického rozdělení: $\overset{\curvearrowright}{\bar{x}} = \tilde{x} = \bar{x}$

95 % interval spolehlivosti

odvozený z jednotlivých hodnot, seřazených podle své velikosti



Dohodnutá symbolika / conventional symbolism :

	základní soubor population	výběrový soubor sample
průměr mean, average	μ	\bar{X}
směrodatná odchylna standard deviation	σ_n	σ_{n-1}
	$\sqrt{\frac{\dots}{n}}$	$\sqrt{\frac{\dots}{n-1}}$

Směrodatná odchylka :

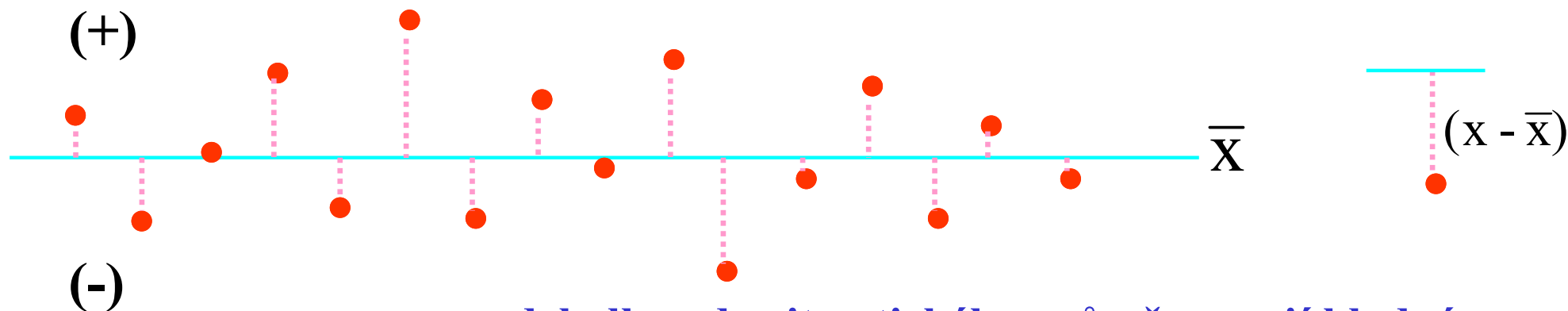
$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \approx \sqrt{\frac{\sum x^2 - n \bar{x}^2}{n - 1}}$$

↑
výukový vzorec

↑
praktický vzorec

„S“ nebo „ σ_{n-1} “ = pro výběrový soubor,
v angl. literatuře také SD (standard deviation)

Statistika potřebuje nenulové hodnoty :



odchylky od aritmetického průměru mají kladné a záporné hodnoty

$$\sum (x - \bar{X}) = 0$$

součet prostých odchylek jednotlivých hodnot od aritmetického průměru je nulový

$$\sum (x - \bar{X})^2 \neq 0$$

součet čtverců odchylek se nerovná nule, proto je tento typ součtu používán pro výpočty

Jak získá statistika nenulové hodnoty ?

1/ druhé mocniny všech čísel (kladných nebo záporných) jsou kladné, proto vidíte:

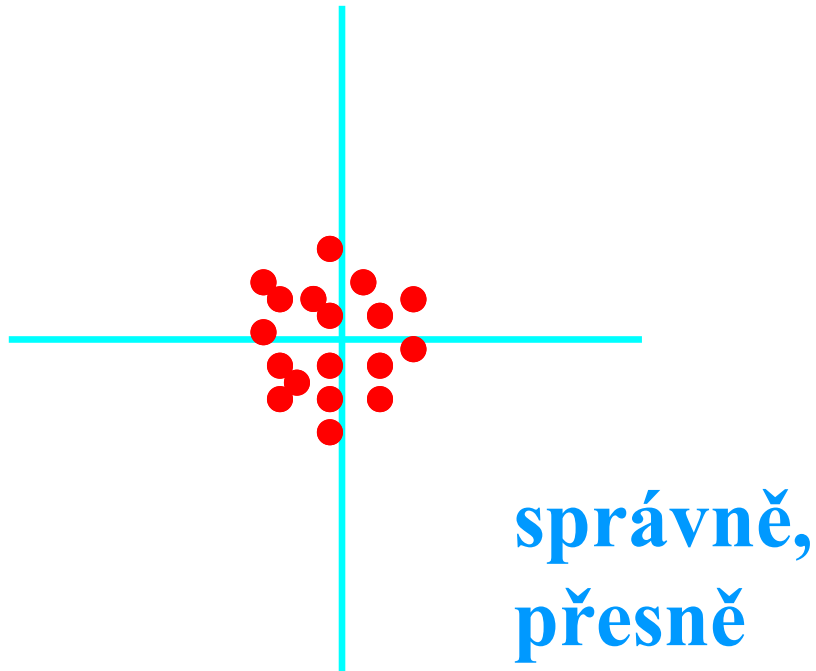
$$\sum (x - \bar{x})^2 \quad \text{nebo} \quad VK_A^2 + VK_B^2 \quad \dots\dots$$

2/ po provedených výpočtech jsou druhé mocniny „vráceny“ do původního stavu druhou odmocninou, proto vzorce:

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad \sqrt{\frac{\sum x^2 - n \bar{x}^2}{n - 1}}$$

$$K \sqrt{VK_A^2 + VK_B^2} \quad \dots\dots$$

Spolehlivost = správnost + přesnost

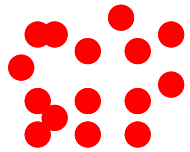


Charakteristikou přesnosti je variabilita.

Mírou variability je např. rozptyl (s^2) nebo variační koeficient (VK).

$$VK = \frac{s}{\bar{x}} \cdot 100 \quad (\%)$$

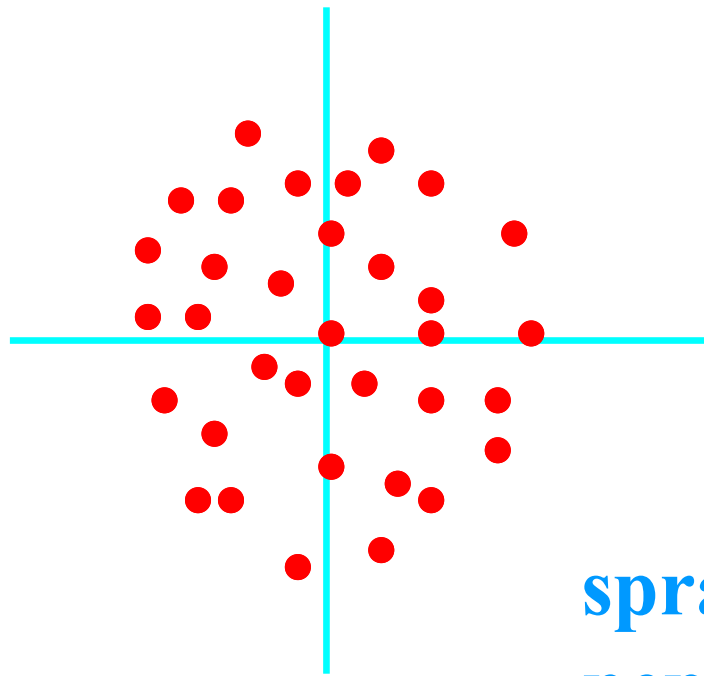
Spolehlivost = správnost + přesnost



Nesprávné výsledky jsou dány
odchylkou od správné hodnoty
(nenáhodná, systematická chyba)

přesně,
nesprávně

Spolehlivost = správnost + přesnost



správně,
nepřesně

biologické jevy



variabilita



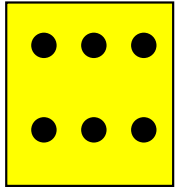
pravděpodobnost :

1/ prostá (nepodmíněná)

2/ podmíněná

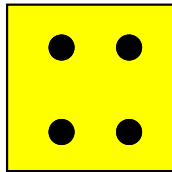
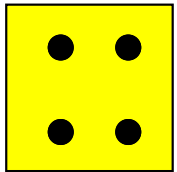
Pravděpodobnost prostá (nepodmíněná) :

(hra v kostky)



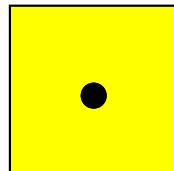
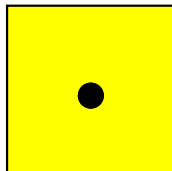
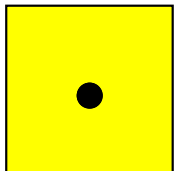
$$P = \frac{1}{6} = 0,166 = 16,6\%$$

Počet hodů: $100 / 16,66 = 6$



$$P = \frac{1}{6} \cdot \frac{1}{6} = 0,166 \cdot 0,166 = 0,166^2 = 0,027 = 2,7\%$$

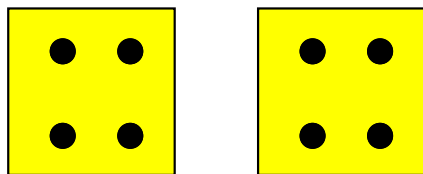
Počet hodů: $100 / 2,7 = 37$



$$P = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = 0,166^3 = 0,004 = 0,4\%$$

Počet hodů: $100 / 0,4 = 250$

Pravděpodobnost prostá (nepodmíněná) :



navzájem nezávislé jevy

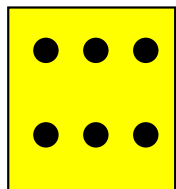
$$P = \frac{1}{6} \cdot \frac{1}{6} = 0,166 \cdot 0,166 = 0,166^2 = 0,027 = 2,7\%$$



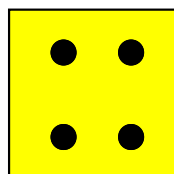
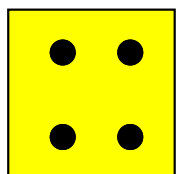
součin pravděpodobností
(nikoliv součet !)

Počet hodů: $100 / 2,7 = 37$

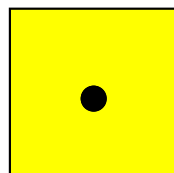
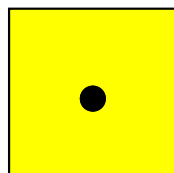
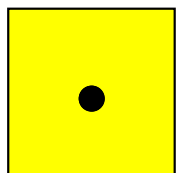
Pravděpodobnost prostá (nepodmíněná) :



$$P = 16,6\%$$



$$P = 2,7\%$$



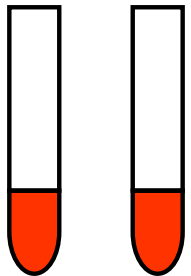
$$P = 0,4\%$$

**Čím více jevů,
tím nižší celková pravděpodobnost !**

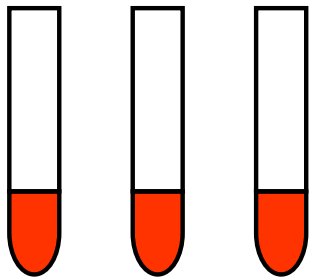
Pravděpodobnost prostá (nepodmíněná) :



$$P = 0,95^1 = 95 \%$$



$$P = 0,95^2 = 90,25 \%$$



$$P = 0,95^3 = 85,74 \%$$

Čím více současně požadovaných laboratorních stanovení, tím menší pravděpodobnost, že výsledek jediného z nich bude ležet uvnitř 95 % intervalu spolehlivosti (u zdravého jedince).

Pravděpodobnost prostá (nepodmíněná) :

pravděpodobnost výskytu 1 výsledku (zdravý jedinec) :

- uvnitř

- mimo



95 % interval
spolehlivosti



obecně :

$$P = 0,95^n$$

$$p = (1 - 0,95)^n$$

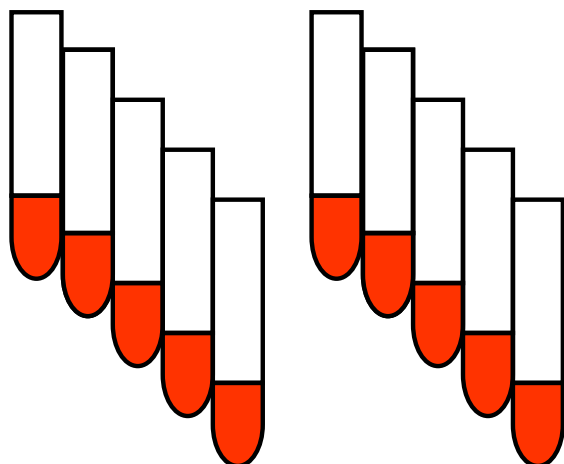


n = 1



$$P = 0,95^1 \\ = 95 \%$$

$$p = (1 - 0,95)^1 \\ = 5 \%$$



n = 10



$$P = 0,95^{10} \\ = 0,5987 \\ \sim 60 \%$$

$$p = (1 - 0,95)^{10} \\ = 0,4013 \\ \sim 40 \%$$

Podmíněná pravděpodobnost :

$P(T/D)$ = pravděpodobnost jevu „T“
za podmínky „D“

podmínka „D“ je vždy přítomna

T = test

T+ = pozitivní test

T- = negativní test

D = diagnóza, nemoc (disease)

D+ = daná diagnóza je přítomna,
nemoc je přítomna

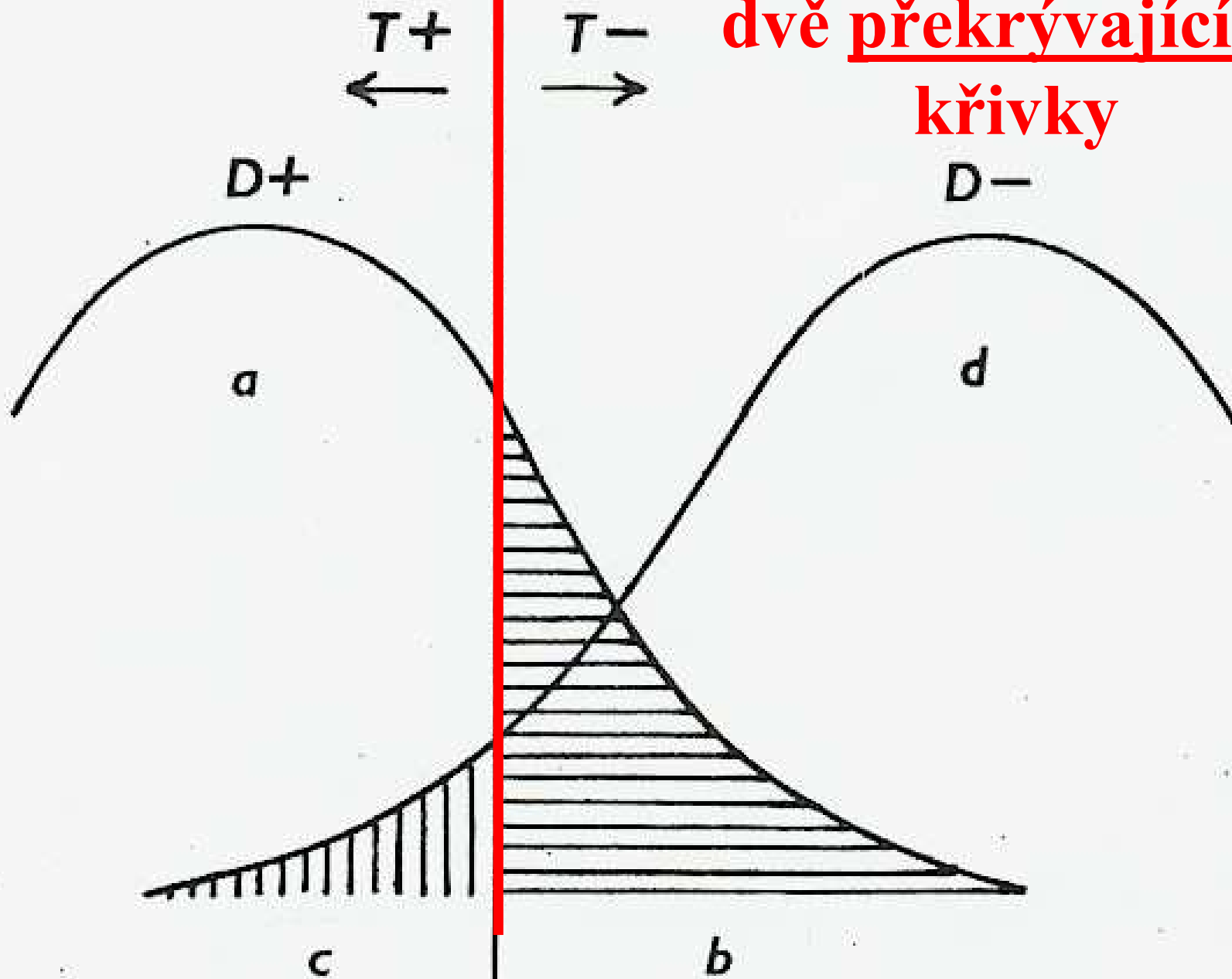
D- = daná diagnóza není přítomna,
nemoc není přítomna

VALIDITU DIAGNOSTICKÝCH TESTŮ,

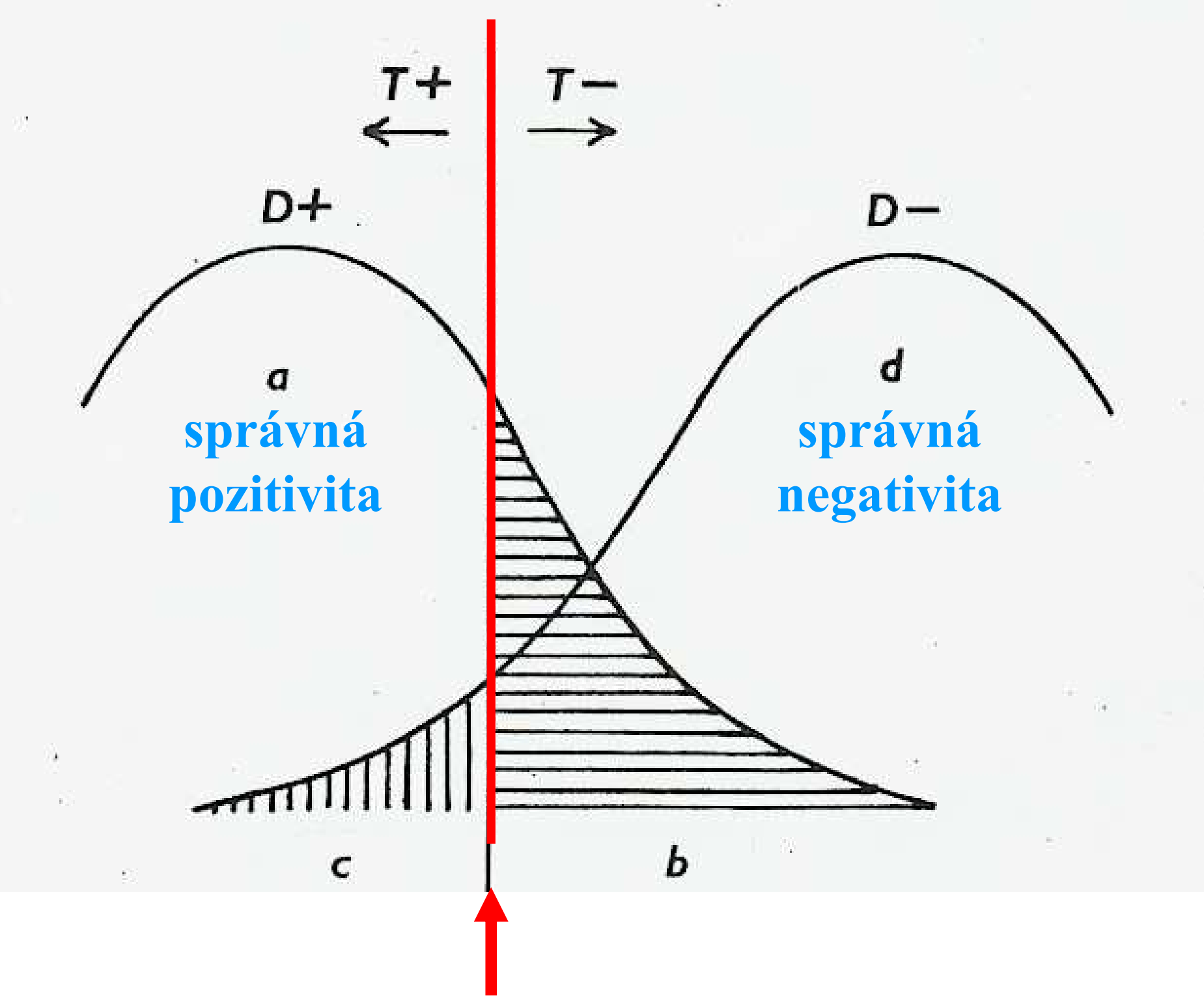
kteřé dávají pouze dva druhy odpovědi
(test negativní a test pozitivní),

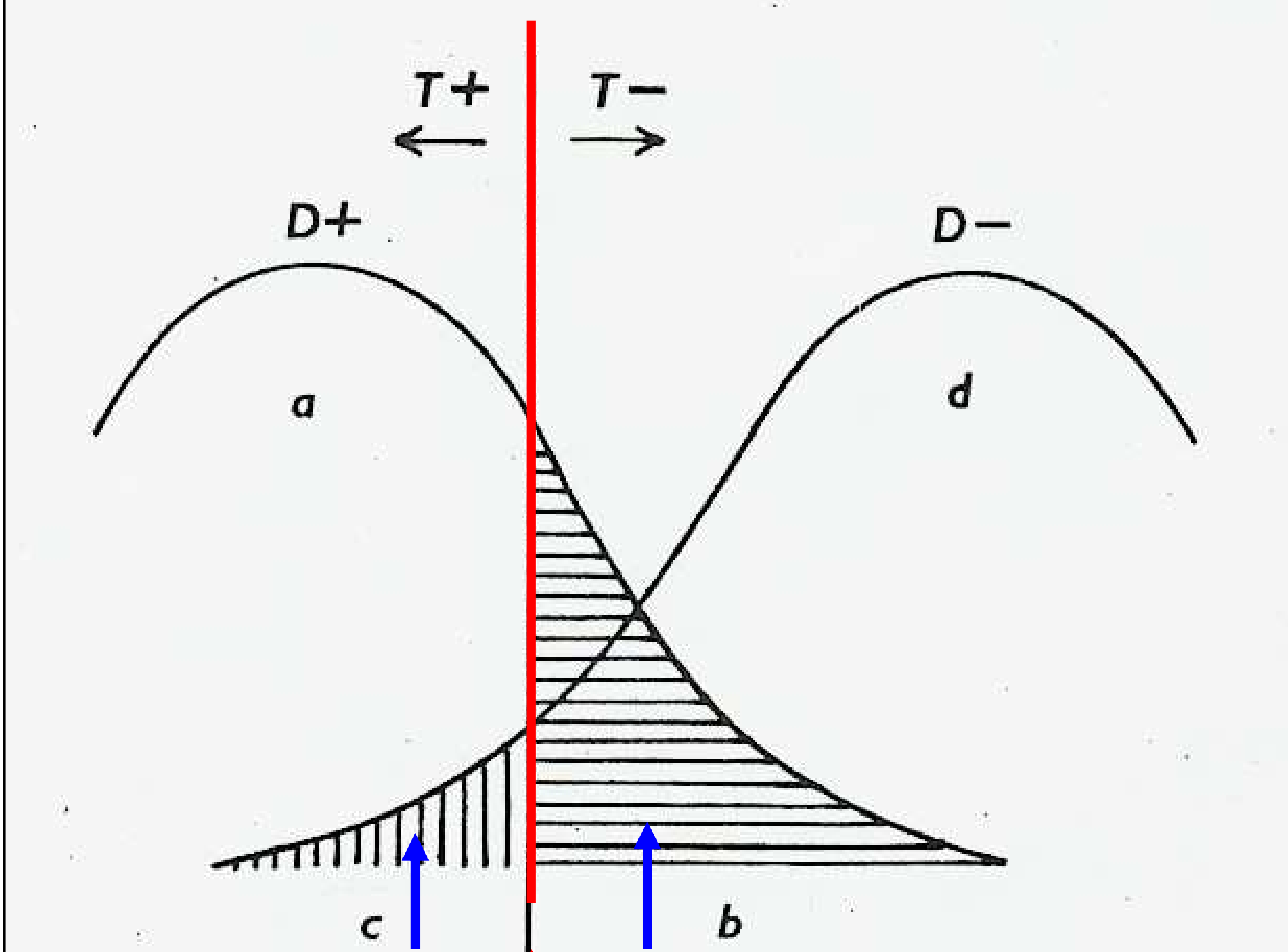
lze vyjádřit dvěma základními poměrnými ukazateli – senzitivitou a specificitou.

dvě překrývající se
křivky



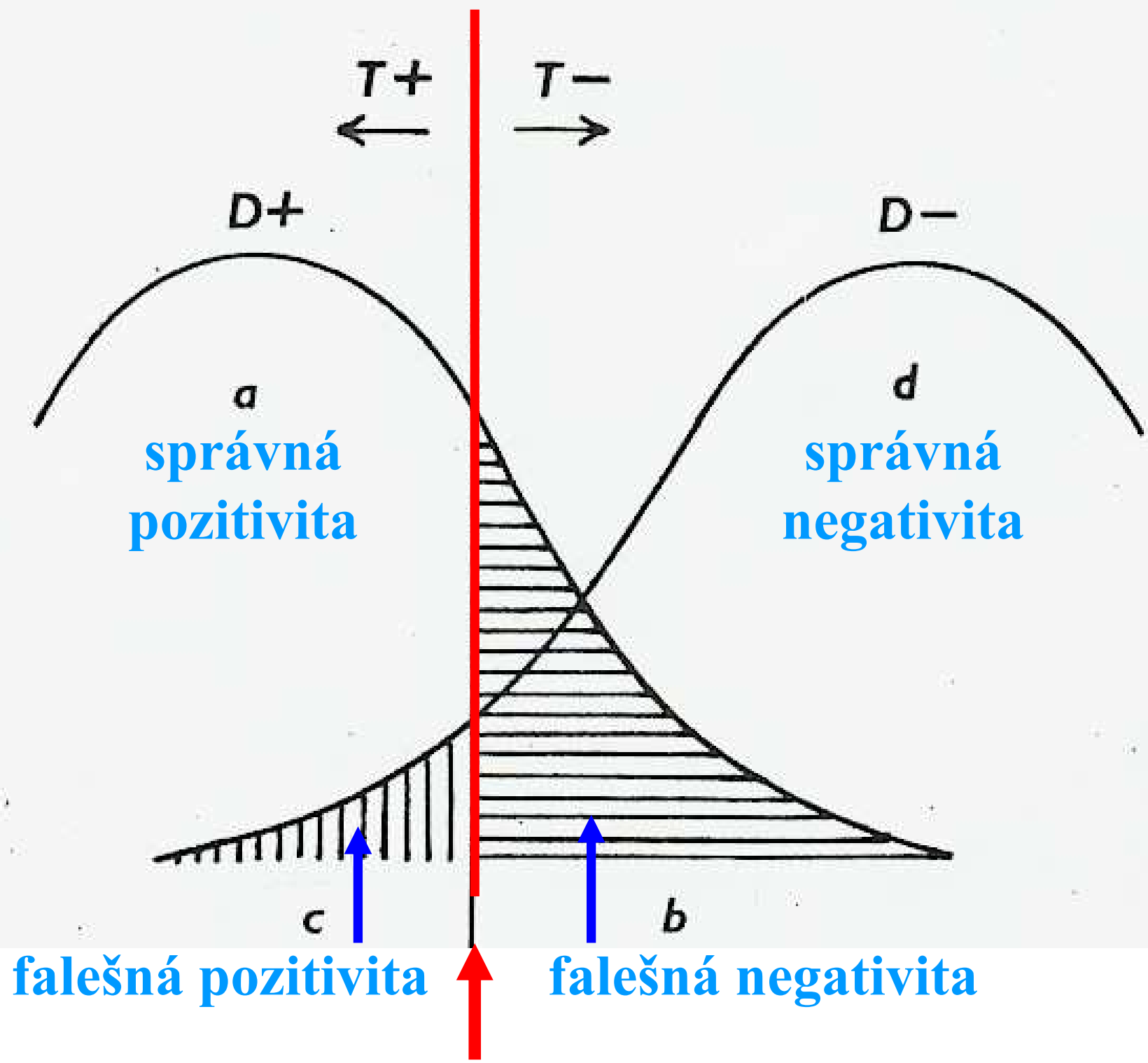
kritická hodnota pro diferenciaci





falešná pozitivita

falešná negativita



	pacient s nemocí (D+)	pacient bez nemoci (D-)
test je pozitivní (T+)	a	c
test je negativní (T-)	b	d

- a = správná pozitivita
- b = falešná negativita
- c = falešná pozitivita
- d = správná negativita

$$ST = a / (a + b)$$

$$SF = d / (c + d)$$

Podmíněná pravděpodobnost :

**P (T-/D-) = správná negativita,
specifičnost, specificita (SF)**

SF = 0,7 → použitelná metoda

SF > 0,95 → velmi dobrá metoda

**P (T+/D+) = správná pozitivita,
senzitivita (ST)**

**efektivita (vydatnost) = nejvyšší pravděpodobnost
shody testu s diagnózou**

Senzitivita = podíl správné positivity testu

„senzitivita“

ST = P (T+/D+)

~ „pozitivita“

= podmíněná pravděpodobnost P (T+/D+),
že pacient s hledanou nemocí (D+)
má pozitivní výsledek testu (T+)

= pravděpodobnost pozitivního testu (T+)
u pacientů s nemocí (D+)

= schopnost testu dát pozitivní odpověď (T+)
v případě, že vyšetřovaná osoba trpí danou
nemocí (D+)

Specificita = podíl správné negativity testu

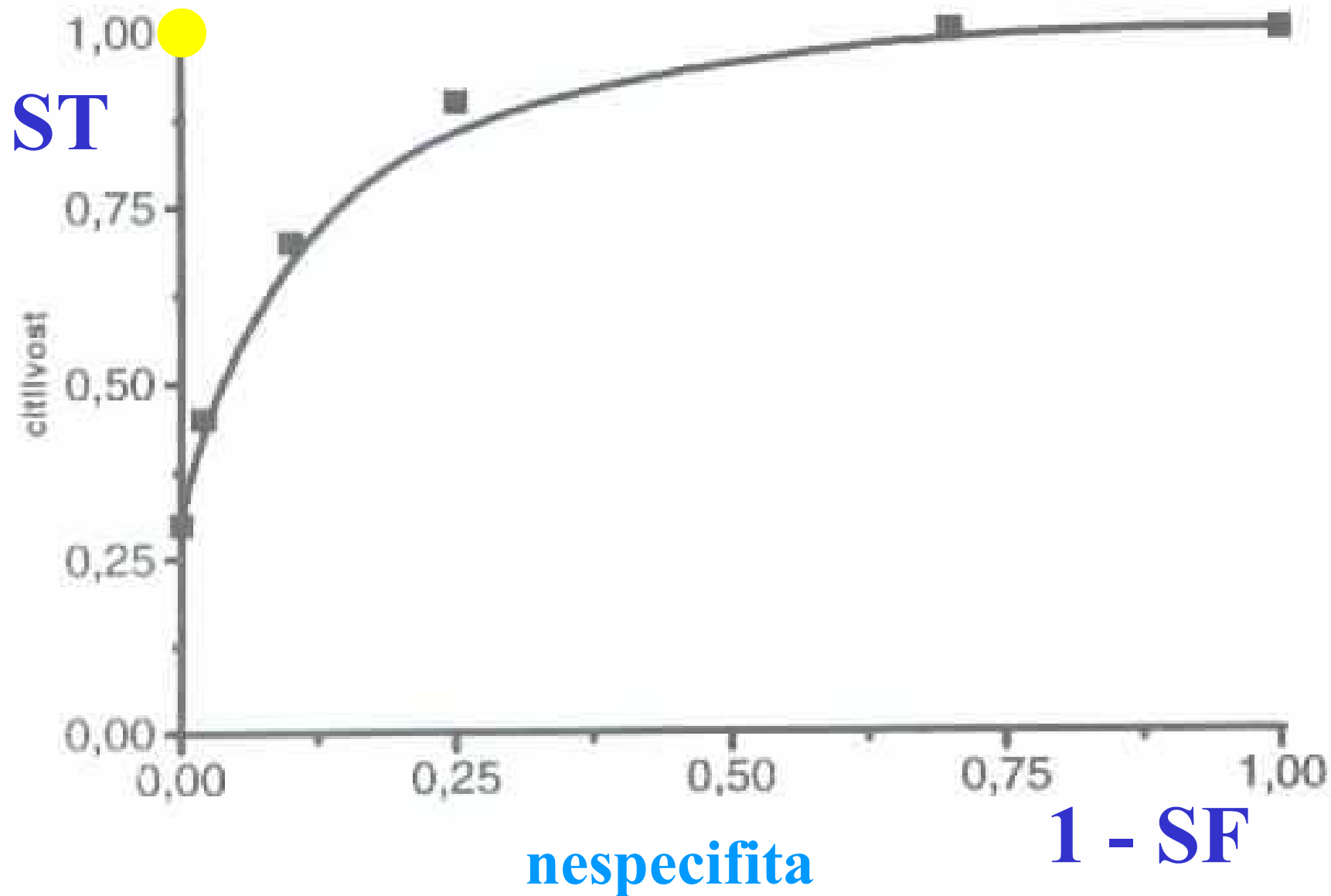
$$\mathbf{SF} = \mathbf{P (T-/D-)}$$

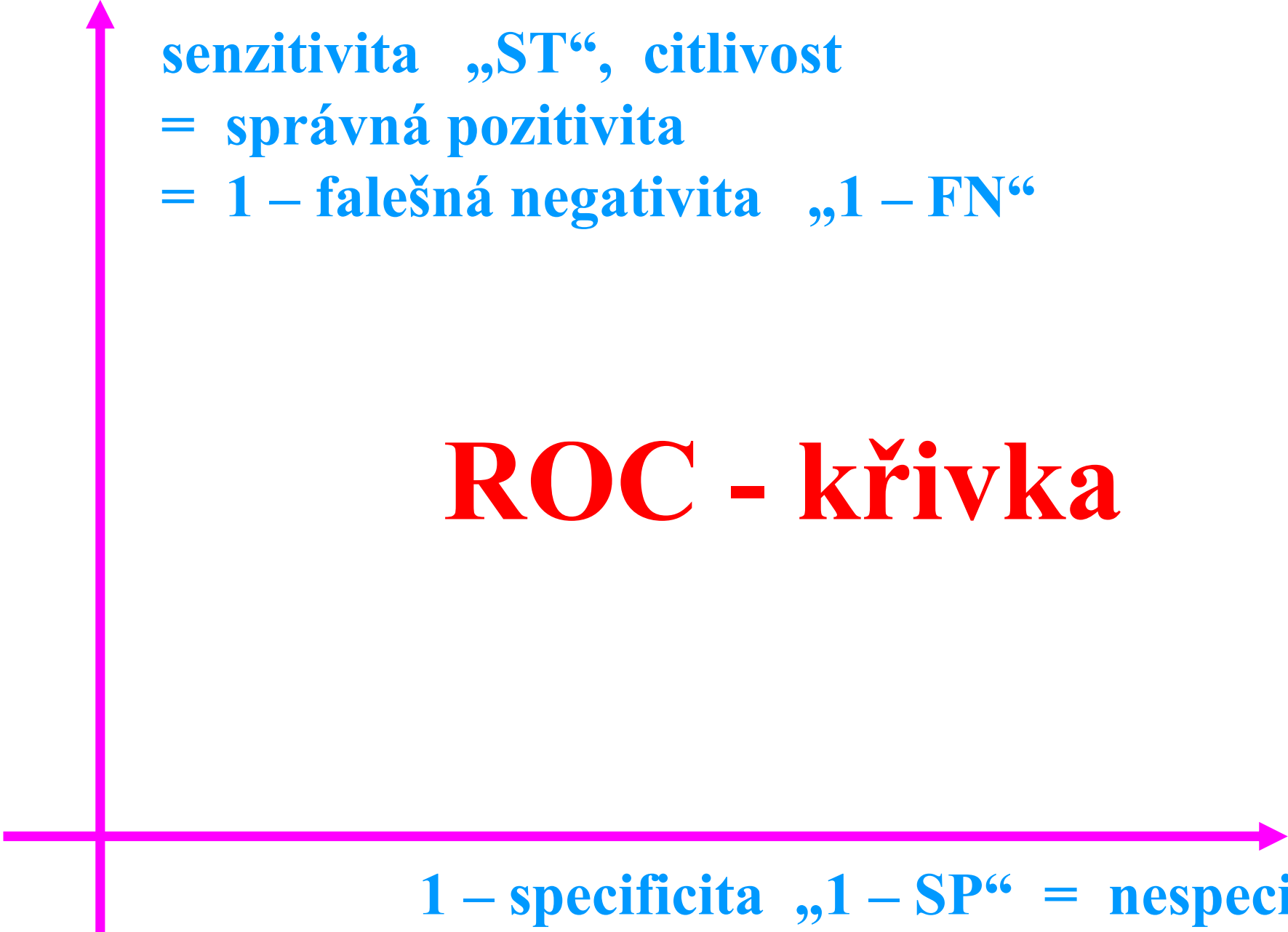
= podmíněná pravděpodobnost $\mathbf{P (T-/D-)}$,
že pacient bez hledané nemoci ($\mathbf{D-}$)
má negativní výsledek testu ($\mathbf{T-}$)

= pravděpodobnost negativního testu ($\mathbf{T-}$)
u pacientů bez nemoci ($\mathbf{D-}$)

= schopnost testu dát zápornou odpověď ($\mathbf{T-}$)
v případě, že vyšetřovaná osoba nemá danou nemoc
($\mathbf{D-}$)

Operativní charakteristická křivka :

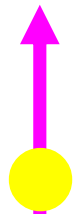




senzitivita „ST“, citlivost
= správná pozitivita
= 1 – falešná negativita „1 – FN“

ROC - křivka

1 – specificita „1 – SP“ = nespecificita
= falešná pozitivita „FP“

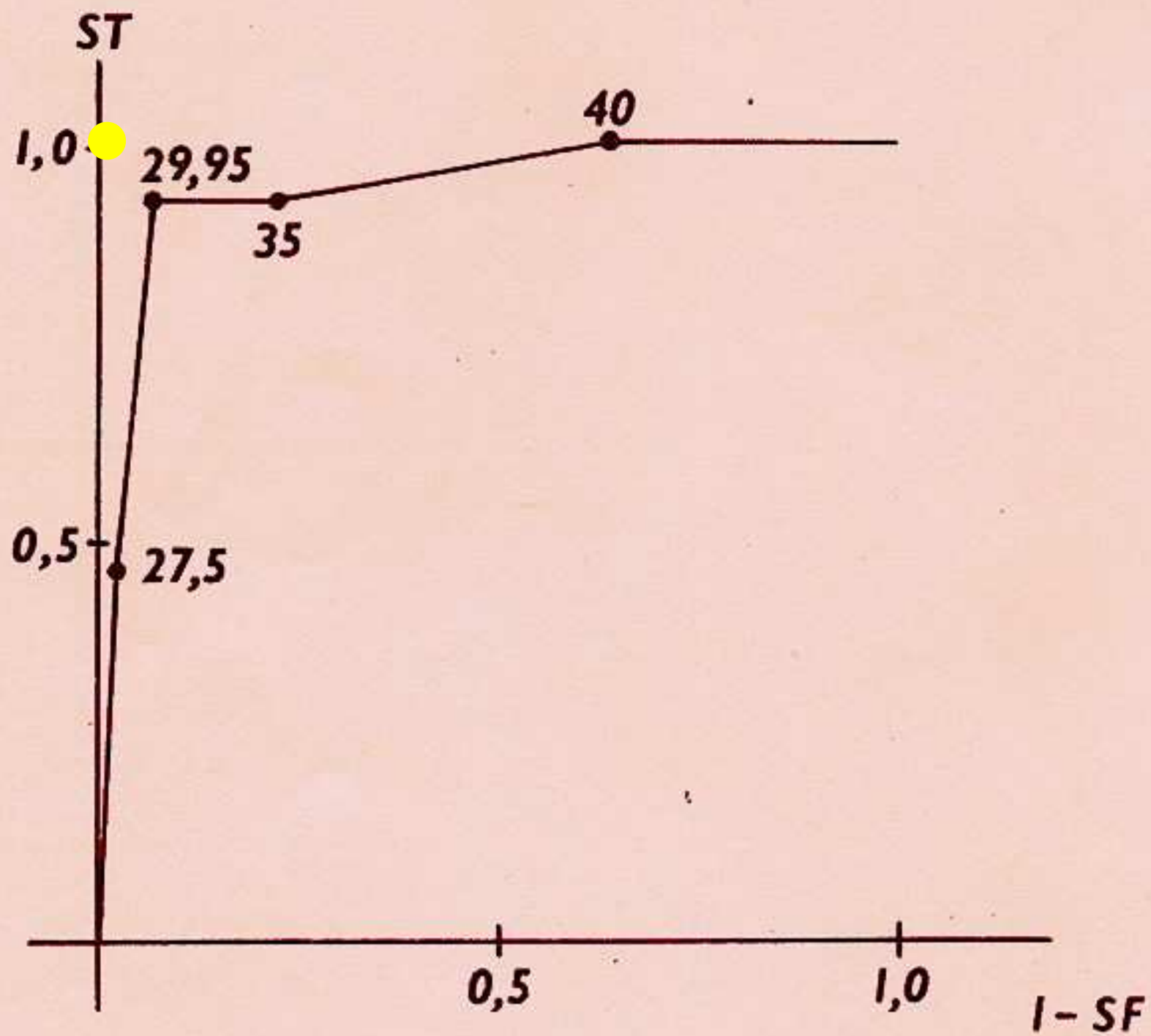


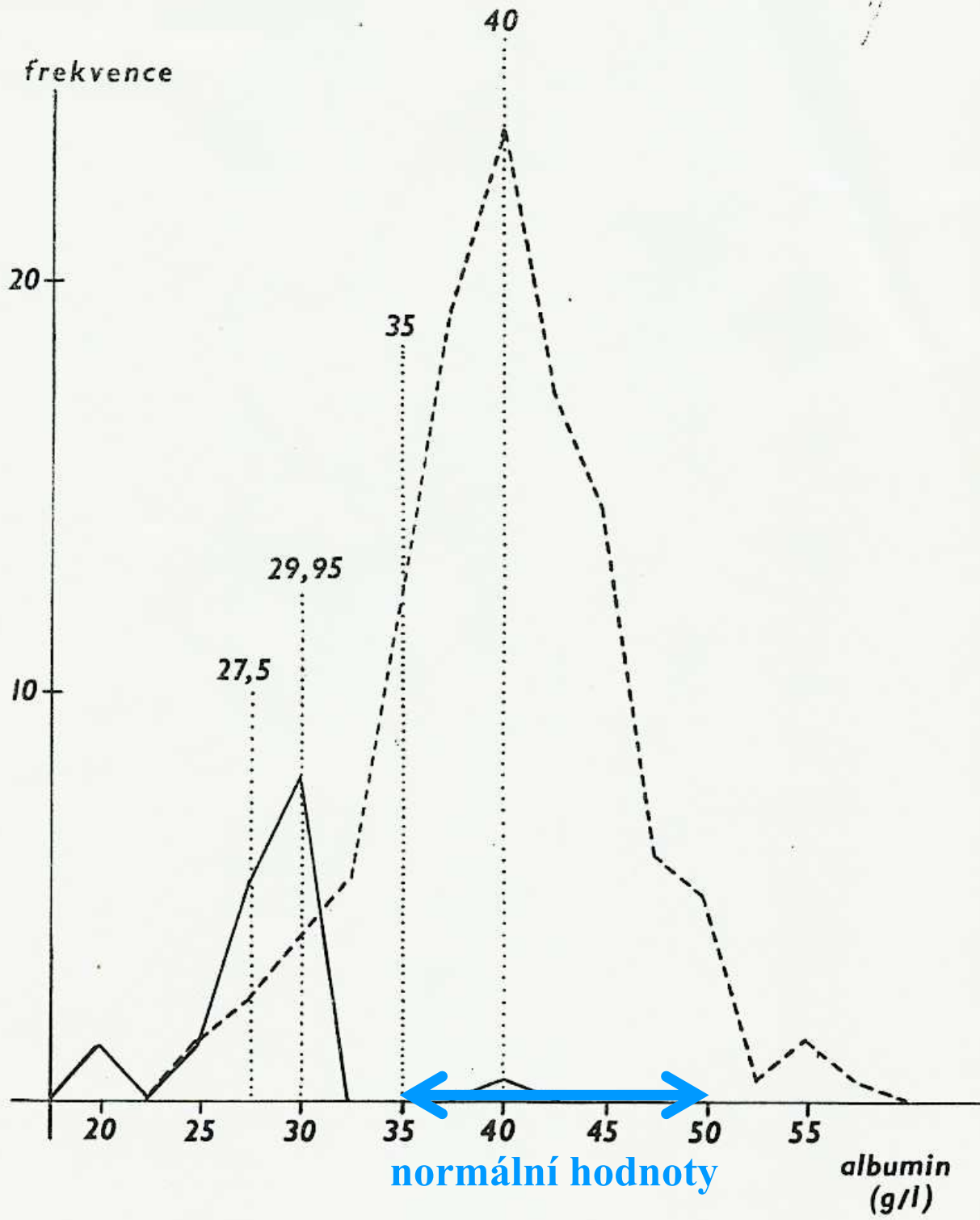
ideální bod (nedosažitelný)

~ 100 % správné positivity
a 0 % falešné positivity

ROC - křivka



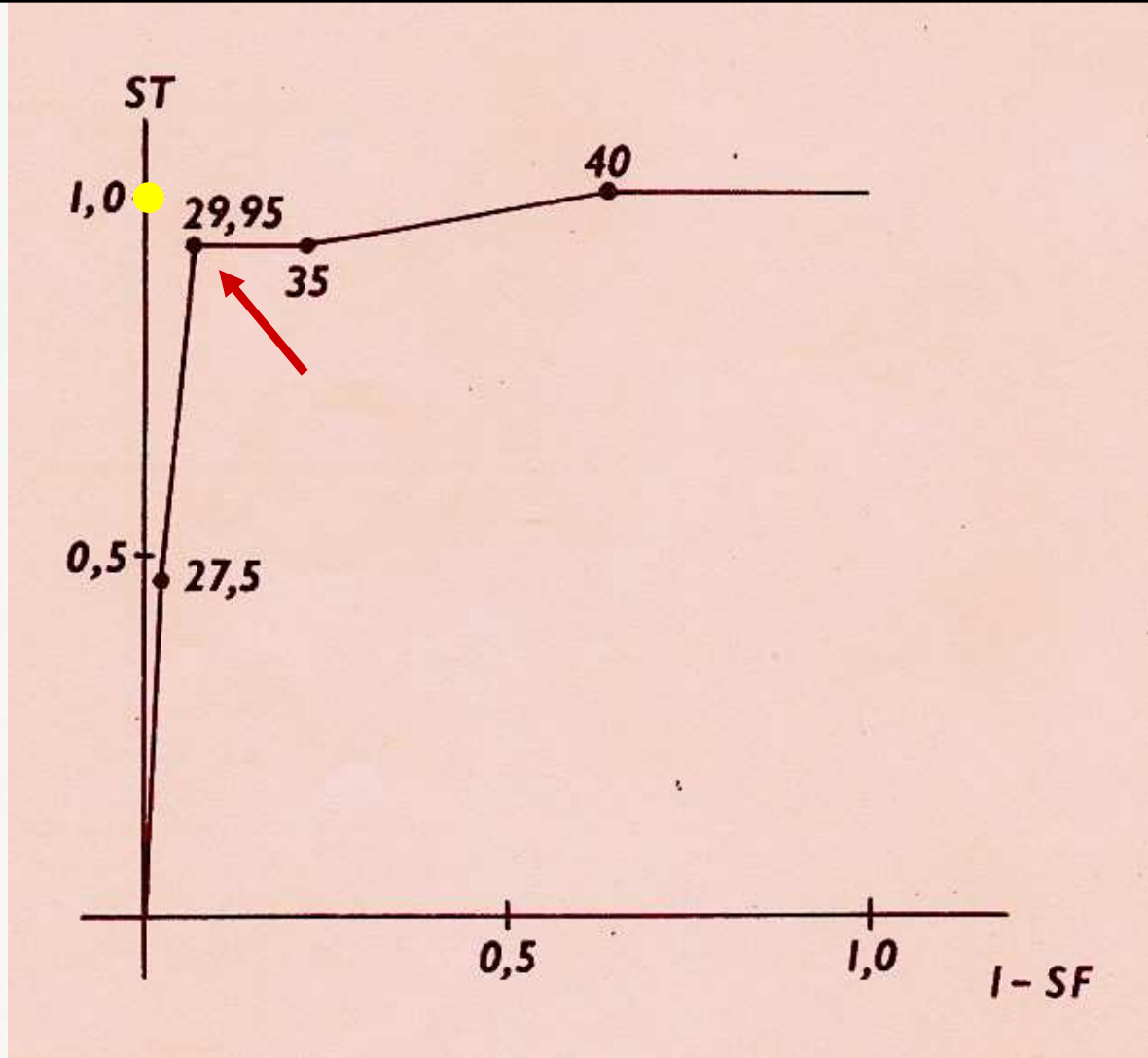
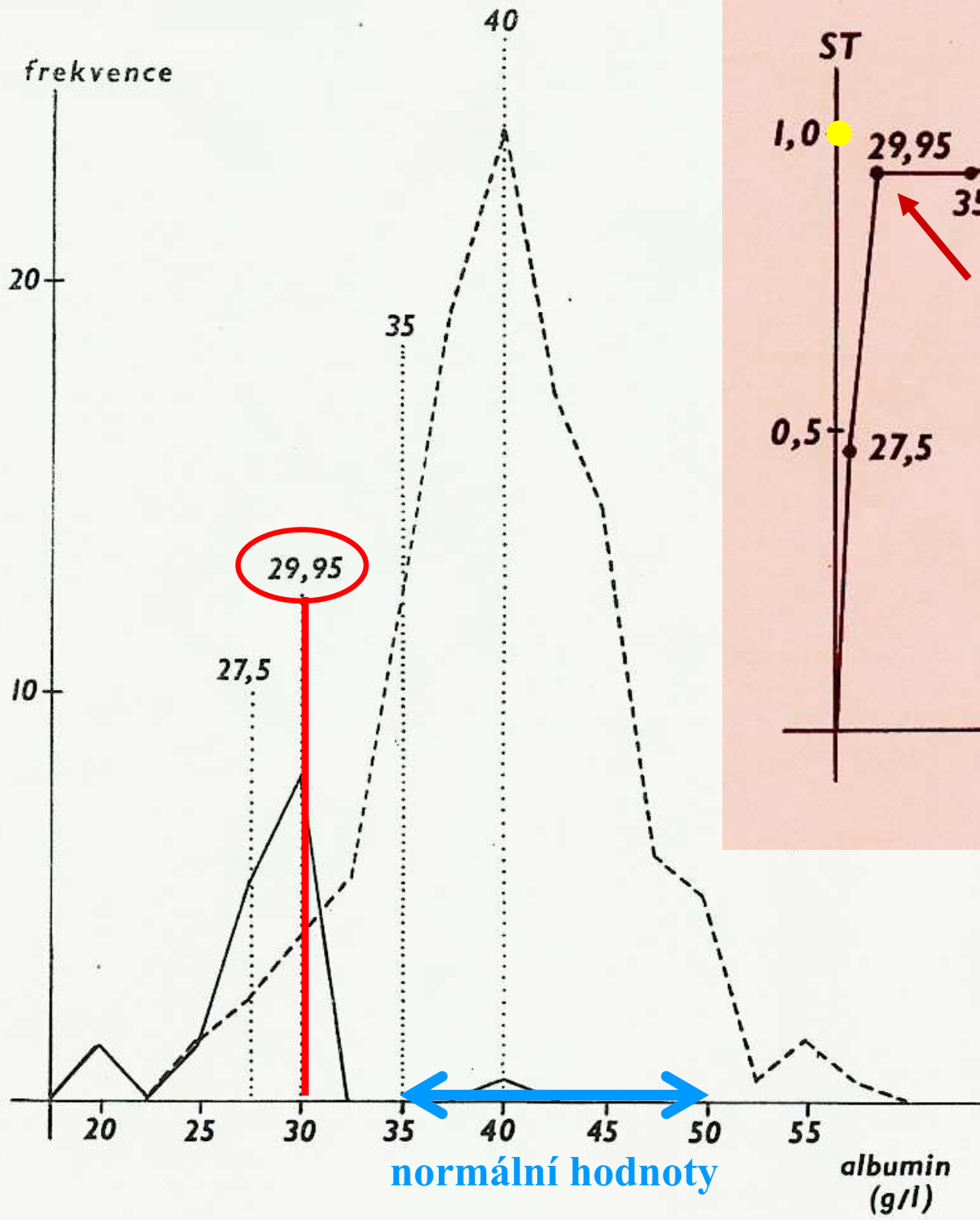




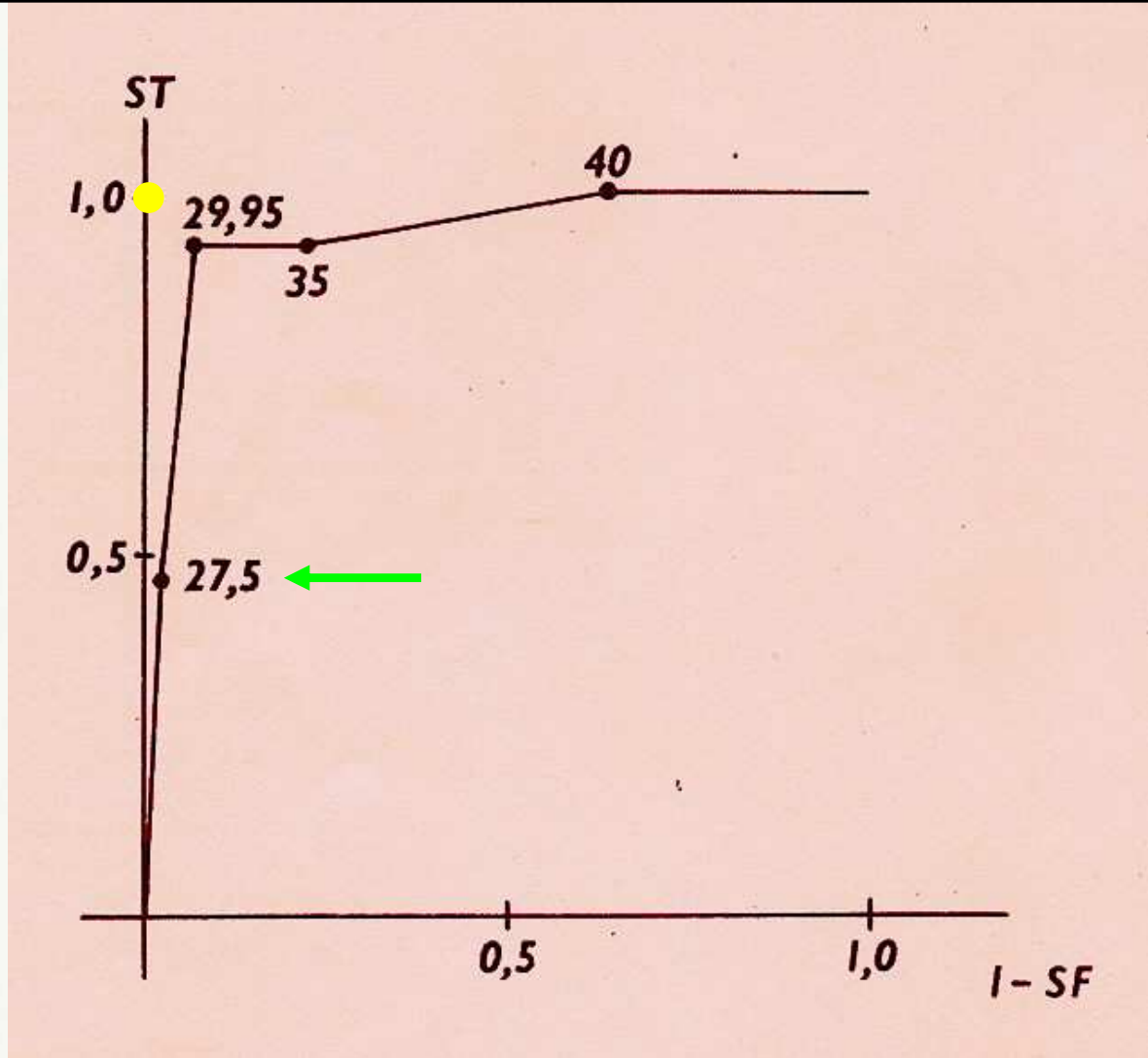
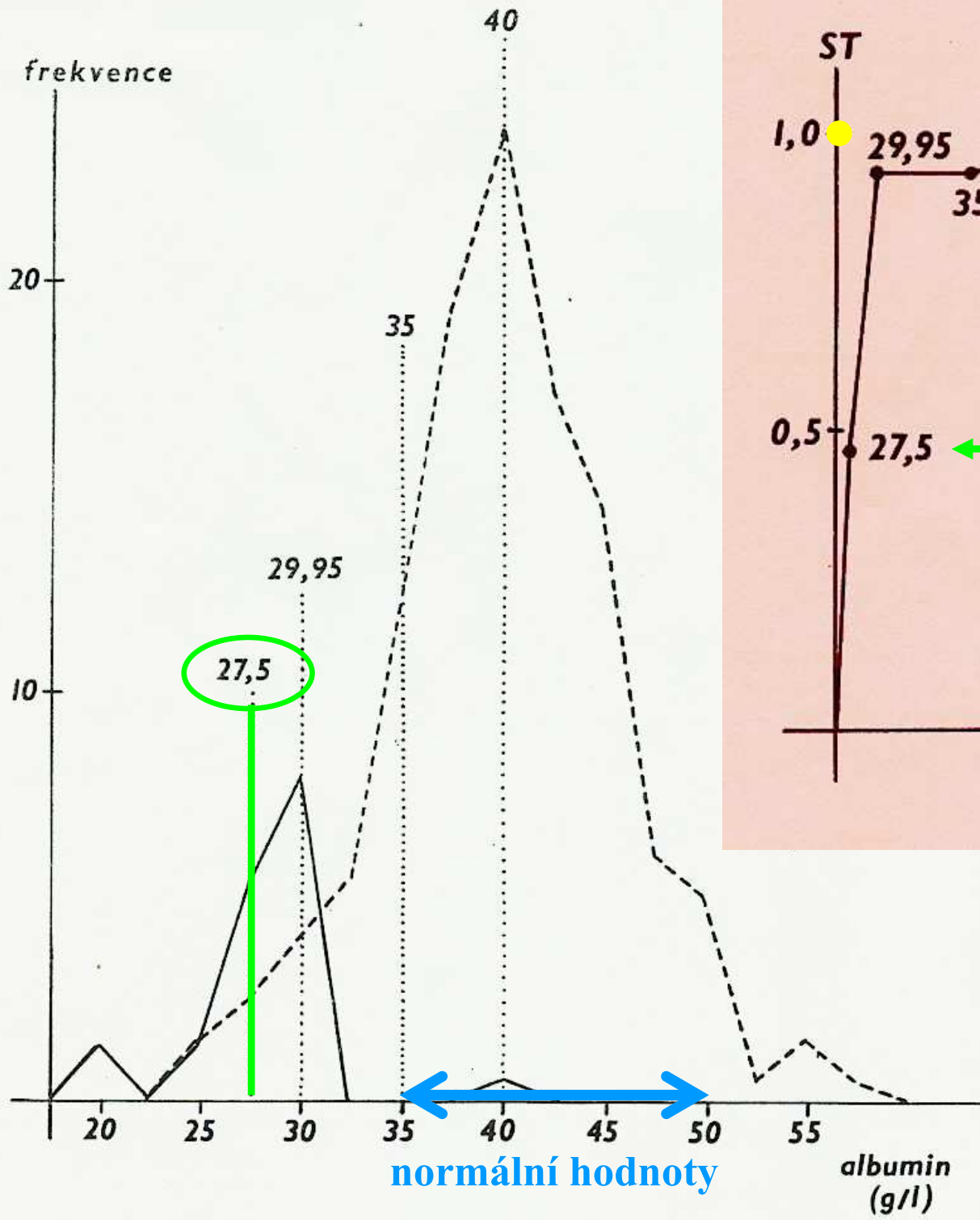
Dva soubory
S-albuminu :

- 1 D+**
- - - 2 D-**

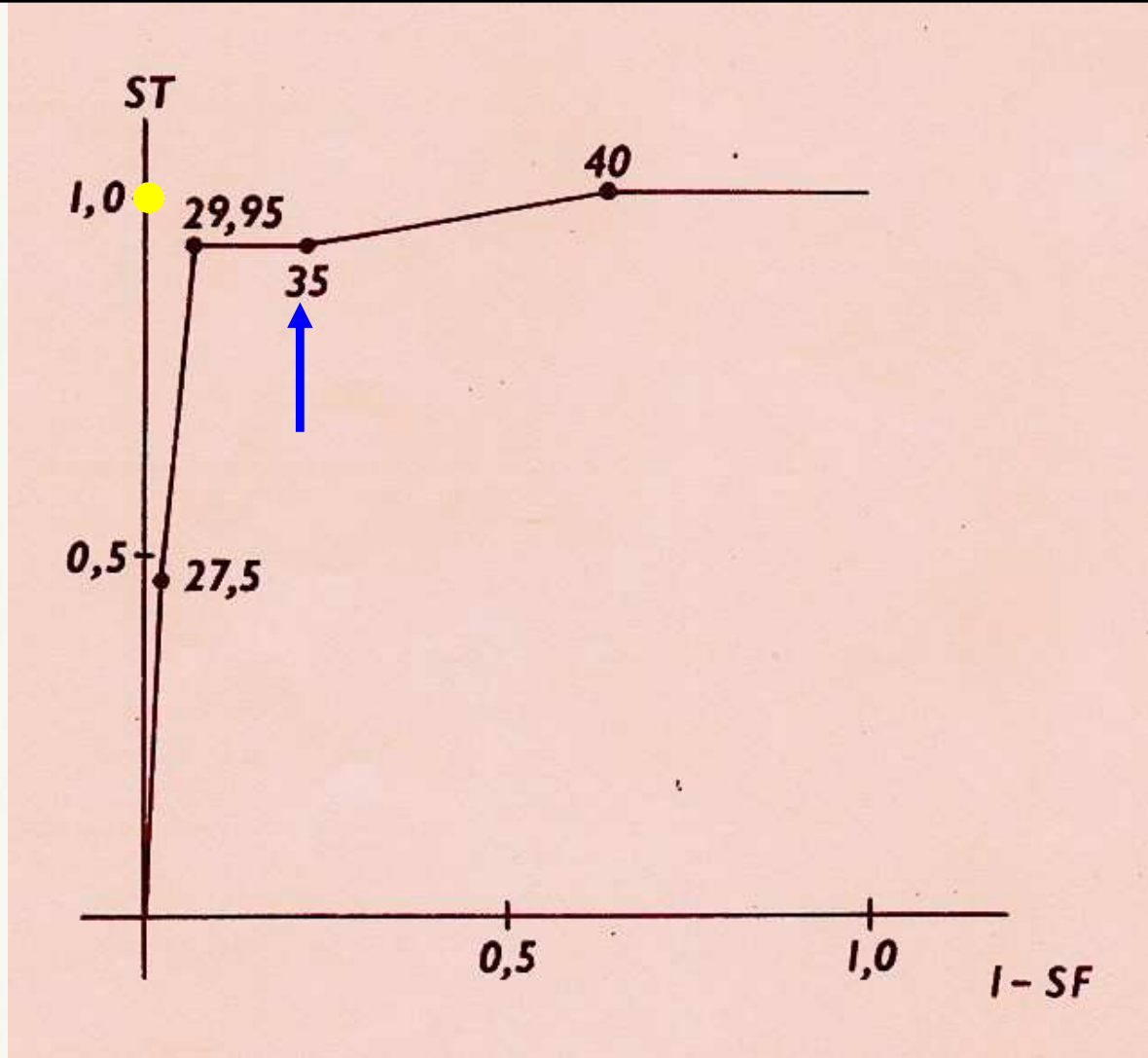
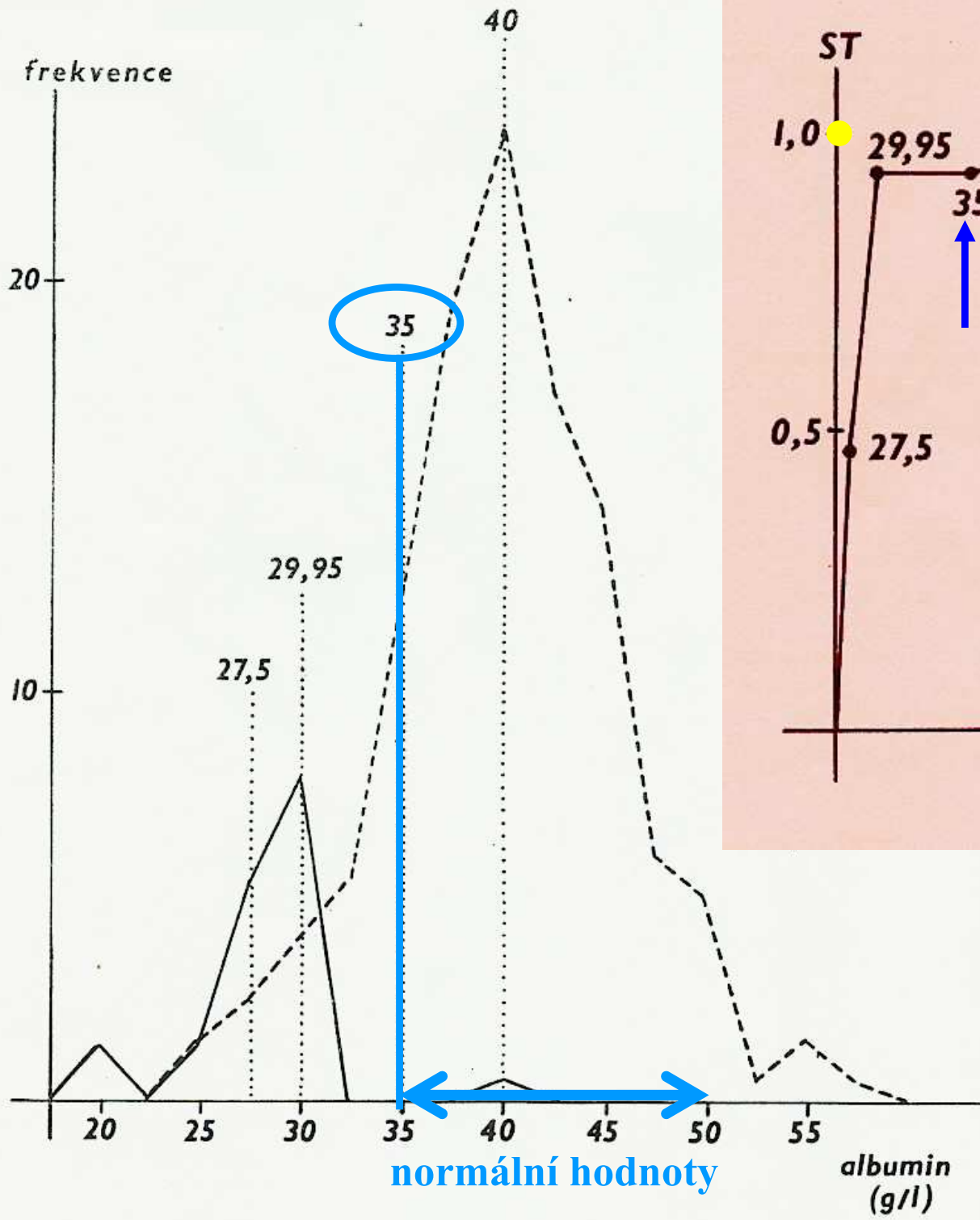
D+ :
 hnísání rány,
 dehiscence rány,
 bronchopneumonie,
 sepse,
 tvorba dekubitů



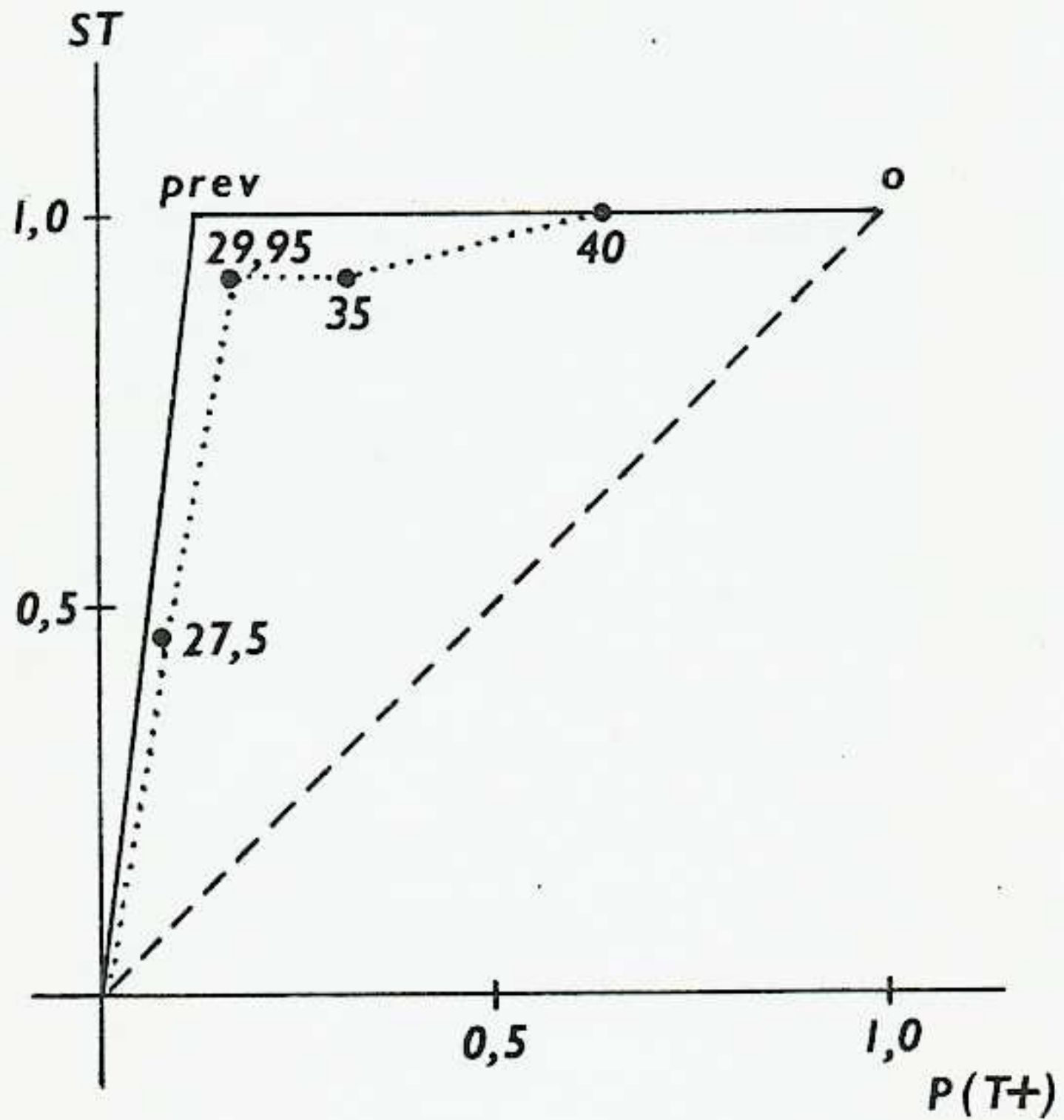
Kritická hodnota
pro diferenciaci
a ROC křivka (1) :



Kritická hodnota
pro diferenciaci
a ROC křivka (2) :



Kritická hodnota
pro diferenciaci
a ROC křivka (3) :



Kritický rozdíl je veličina k posouzení statisticky významné odlišnosti dvou srovnávaných hodnot u stejného pacienta („sám sobě kontrolou“)

$$\text{kritický rozdíl} = K \cdot \sqrt{VK_A^2 + VK_B^2}$$

A = analytická variabilita (reprodukovatelnost po dnech)

B = biologická variabilita (proměnlivost u daného jedince)

$$K = 2,77 = 1,96 \cdot \sqrt{2}$$



2 srovnávané hodnoty

$$\bar{x} \pm 1,96 s \quad (95 \% \text{ interval spolehlivosti})$$

Kritický rozdíl :

Cholesterol:	minulé stanovení	8,0 mmol/l (~ 100 %)
		↓ - 25 % (pokles)
	dnešní stanovení	6,0 mmol/l (~ 75 %)

Kritický rozdíl :

Cholesterol:	minulé stanovení	8,0 mmol/l (~ 100 %)	
			↓ - 25 % (pokles)
	dnešní stanovení	6,0 mmol/l (~ 75 %)	

Laboratoří sdělený kritický rozdíl pro cholesterol byl 19 % .

Rozdíl u našeho pacienta (25 %) je větší než kritický rozdíl.

Rozdíl u pacienta je tedy statisticky významný, (je větší než součet analytické a biologické variability).

Pouhou náhodou může být způsoben jen výjimečně
(s pravděpodobností < 5 %)

Referenční hodnoty („normální hodnoty“):



referenční jedinec → NE :

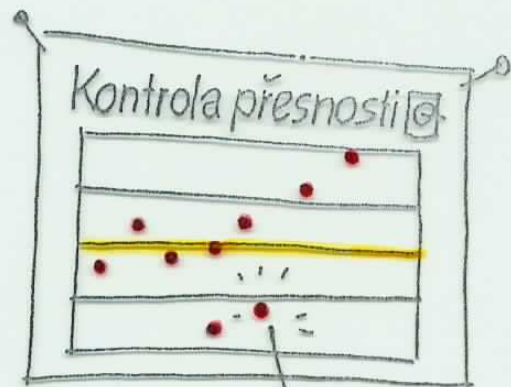
1. těhotné ženy
2. dítě / věk ? muž / žena ?
3. jedinci po fyzické námaze / po vystavení stresu
4. po požití potravy (výjimka: zátěžové stavy)
5. po podání léků
6. nemoc / rizikové faktory
7. „zdraví dárci krve“

Preanalytická variabilita :

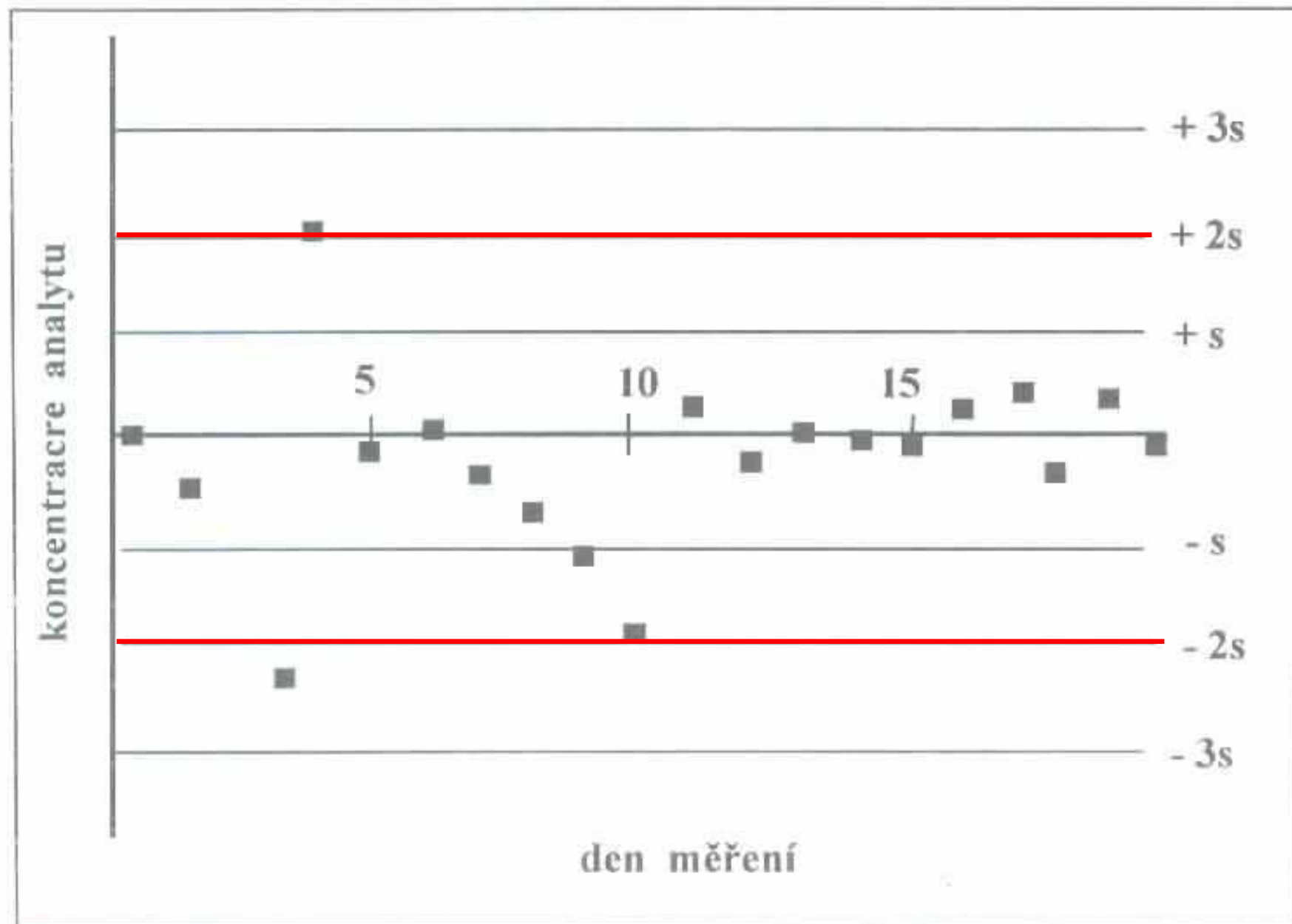
Standardní odběr :

- **poloha (vleže vs. ambulantně)**
- **denní doba**
- **nalačno**
- **kompresie žíly / prstu**
- **doba a způsob srážení (druh a koncentrace antikoagulantu)**
- **skladování (teplota, UV, ...)**
- **doba dodání do laboratoře**

KONTROLA PROVOZU

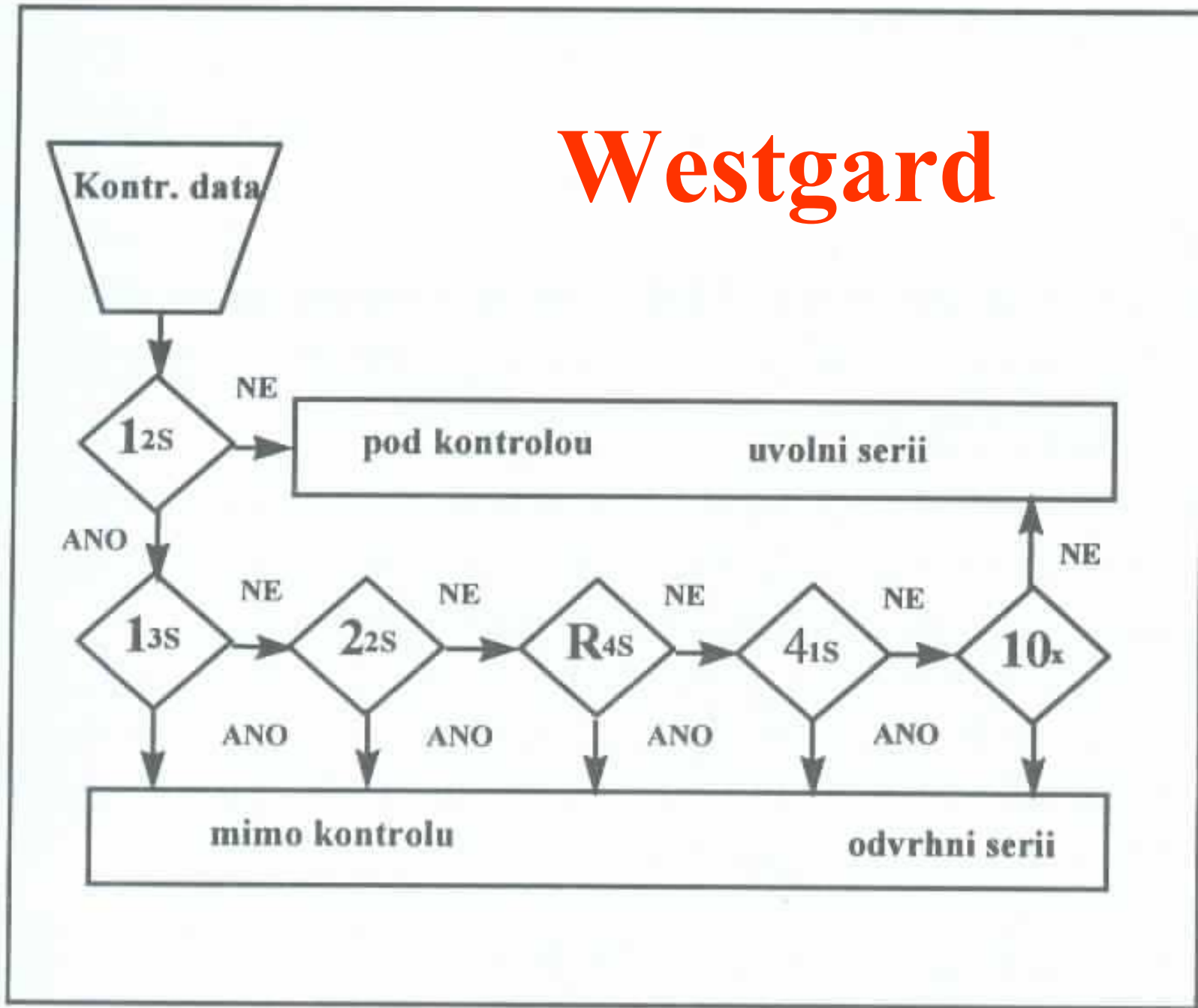


Regulační diagram

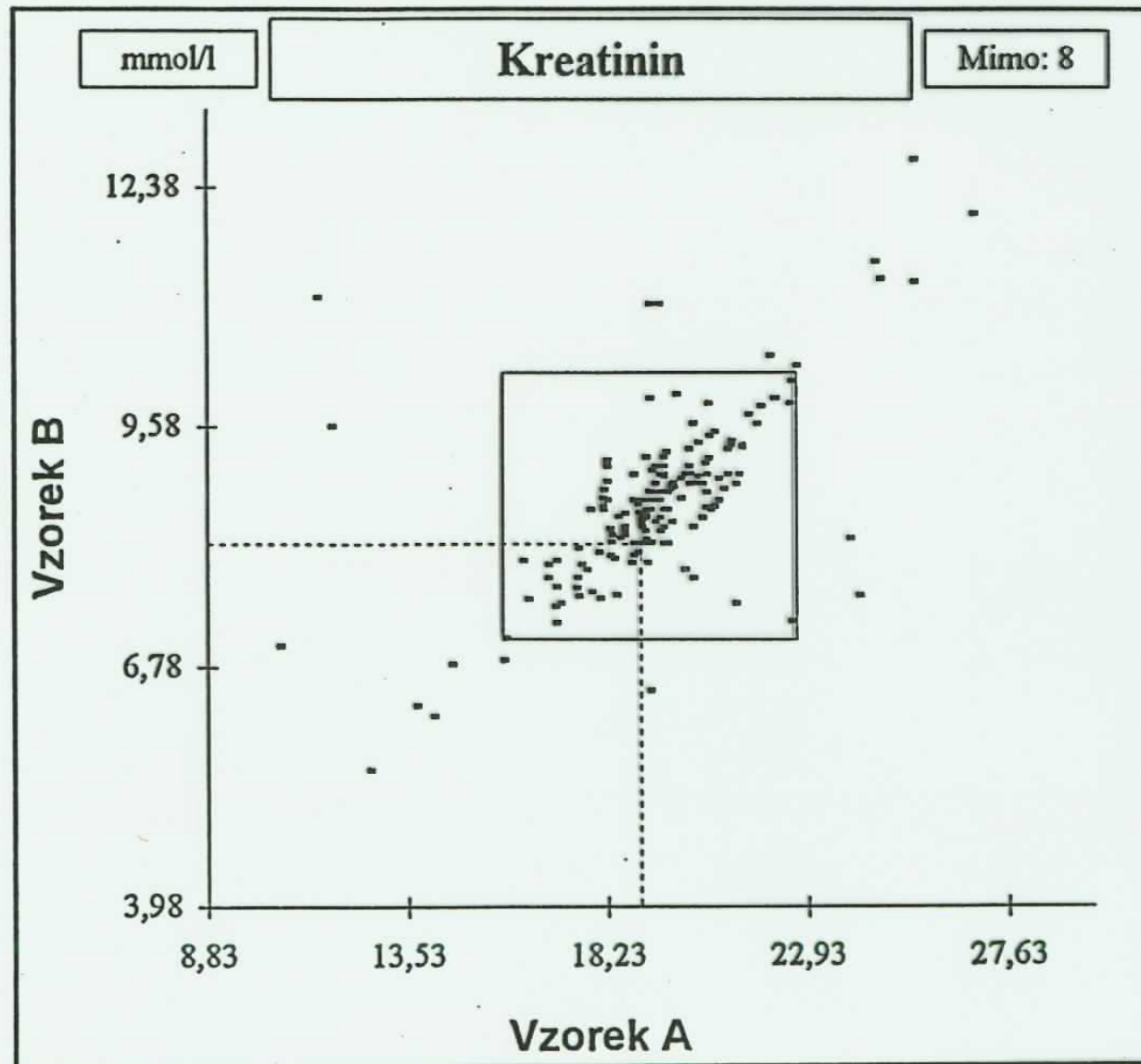


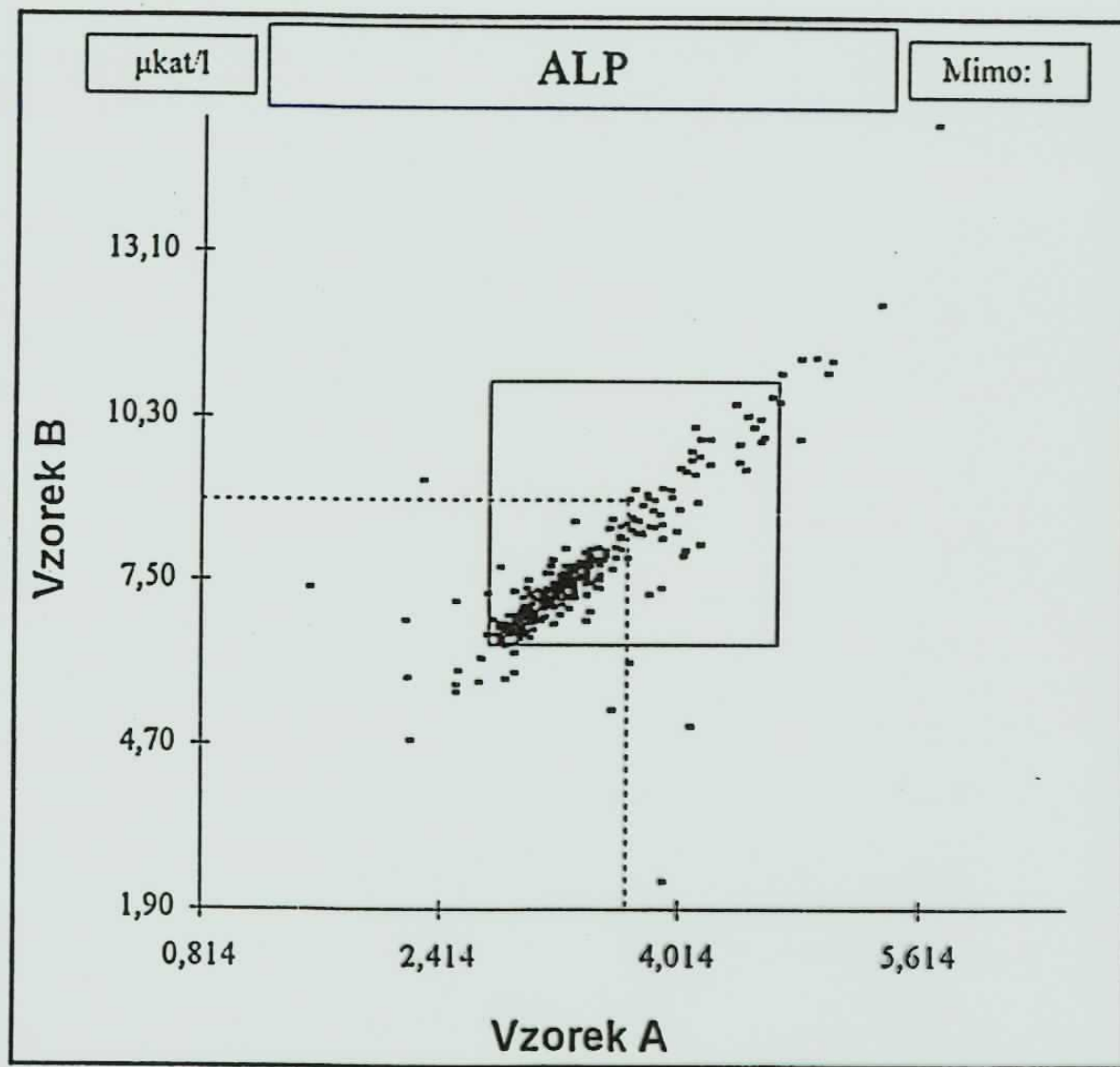


Westgard



1_{2s}	kontroluje se, jestli alespoň jeden výsledek kontroly překročil $\pm 2s$
1_{3s}	kontroluje se, jestli alespoň jeden výsledek kontroly překročil $\pm 3s$
2_{2s}	kontroluje se, jestli dva po sobě jdoucí výsledky kontroly překročily buď $2s$ nebo $-2s$
R_{4s}	kontroluje se, jestli rozdíl (variační rozpětí) mezi dvěma výsledky kontrolního vzorku $> 4s$
4_{1s}	kontroluje se, jestli 4 po sobě jdoucí výsledky téhož kontrolního vzorku přesahují buď $4 \times 1s$ nebo $4 \times (-1s)$
10_x	kontroluje se, zda 10 po sobě jdoucích výsledků stanovení téhož kontrolního vzorku je na jedné nebo druhé straně průměru







beista

Nedorozumění laboratorních a klinických oddělení :

- 1/ 95 % interval spolehlivosti**
- 2/ „kvalita samotného laboratorního výsledku nikdy nemůže být lepší než kvalita dodaného vzorku !“**
- 3/ biochemický „make up“**

