Analýza a klasifikace dat – přednáška 7 – doplnění

RNDr. Eva Koriťáková

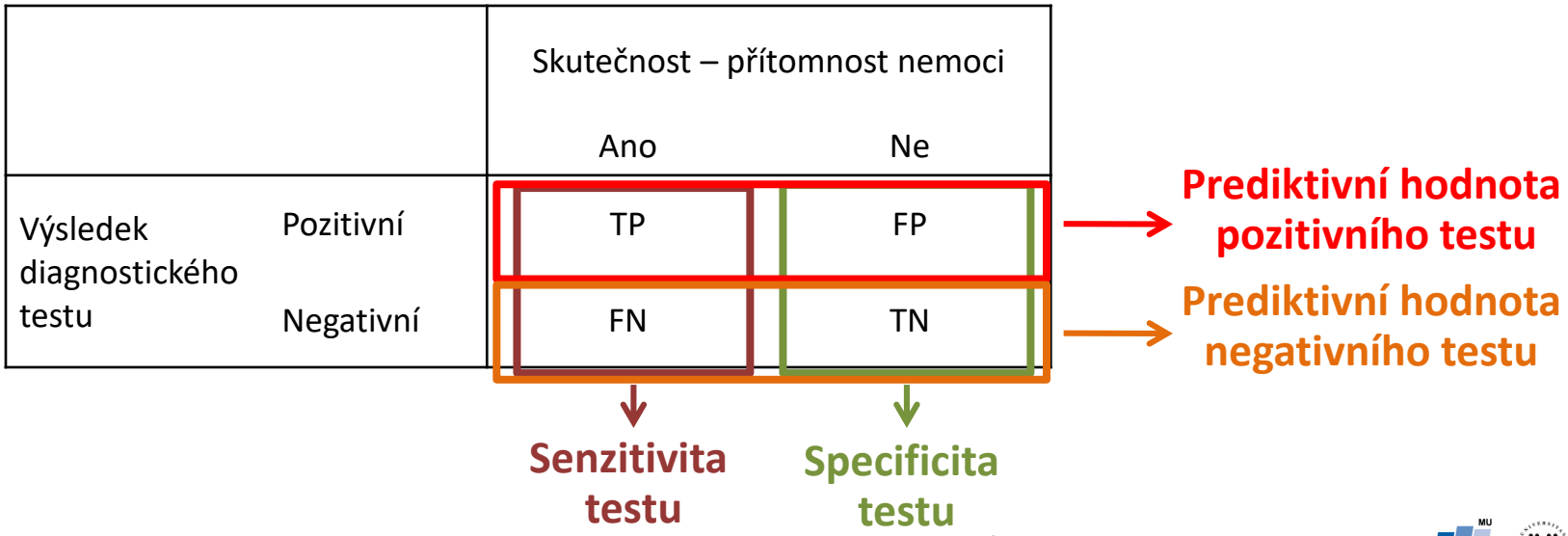
Podzim 2016

Hledání diagnostického cut-off pomocí ROC křivek

Diagnostické testy

- Příklady: hodnocení úspěšnosti diagnostiky pomocí neuropsychologických testů, hodnocení úspěšnosti klasifikace pacientů s Alzheimerovou chorobou a kontrolních subjektů.

- Diagnostický test u dané osoby indikuje přítomnost nebo nepřítomnost sledovaného onemocnění.
- Osoba ve skutečnosti má nebo nemá sledované onemocnění.
  → **Zajímají nás diagnostické schopnosti testu.**



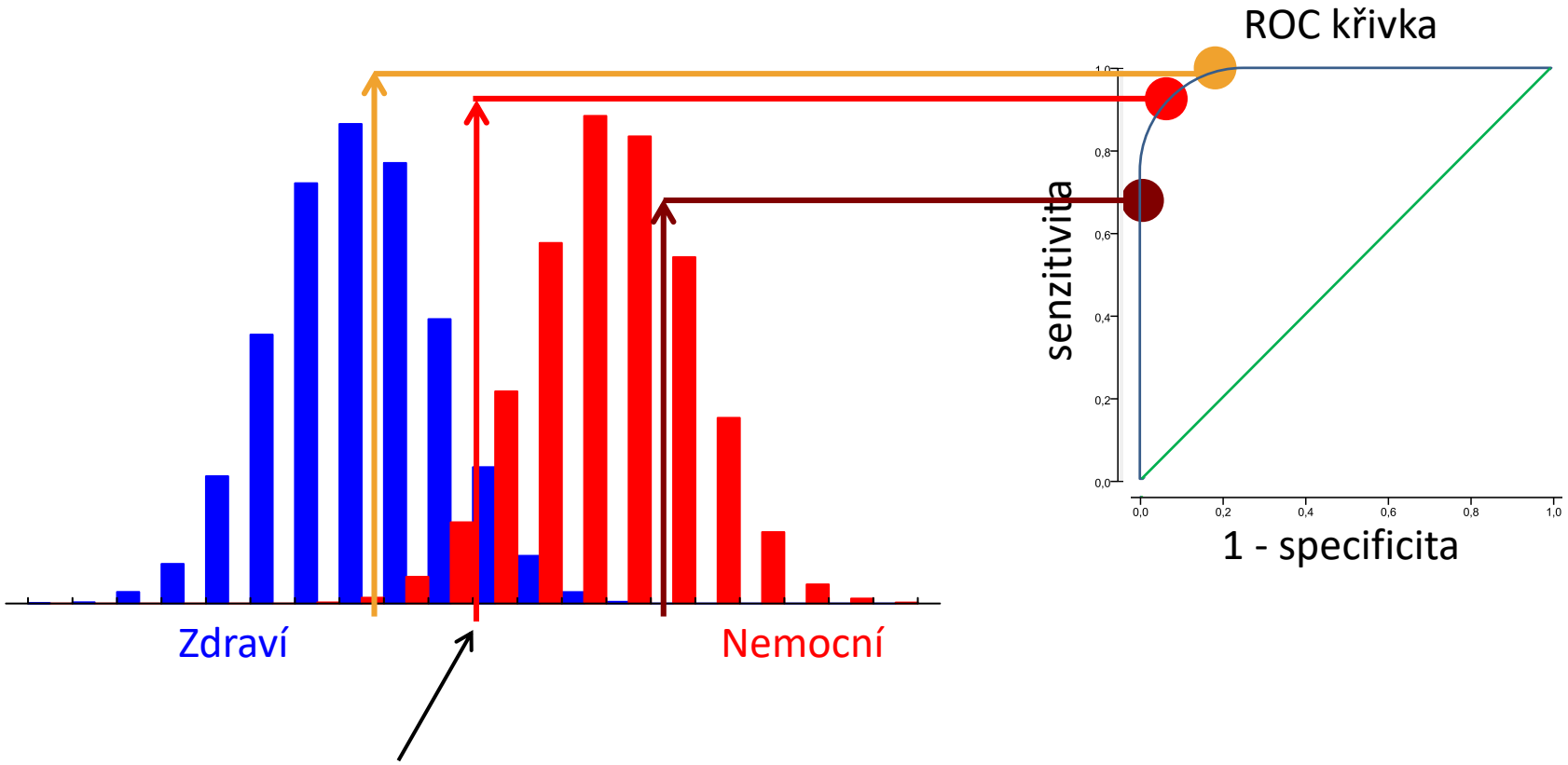
ROC analýza – motivace

- Výše zmíněné ukazatele diagnostické síly testů (senzitivita, specificita apod.) **nelze použít u diagnostických testů, jejichž výstupem je spojitá (kvantitativní) proměnná** (např. koncentrace analytu v krevním séru, systolický krevní tlak).
- Výhoda, pokud na základě předchozích výzkumů známe dělicí body, které odlišují normální a patologické hodnoty spojité proměnné, pomocí nichž můžeme spojitou proměnnou binarizovat – tzn. vytvoření dvou kategorií „pozitivní“ / „negativní“ (např. „pod normou“ / „v normě“).
- Pokud dělicí body nejsou známy předem, můžeme se je snažit nalézt pomocí **ROC („Receiver Operating Characteristic“) křivky**.

- **Cíle ROC analýzy:**
  1. Určit, zda je spojitá proměnná vhodná pro diagnostické odlišování zdravých a nemocných jedinců.
  2. Nalezení dělicího bodu („cut-off point“) na škále hodnot spojité proměnné, který nejlépe odlišuje zdravé a nemocné jedince.

ROC analýza

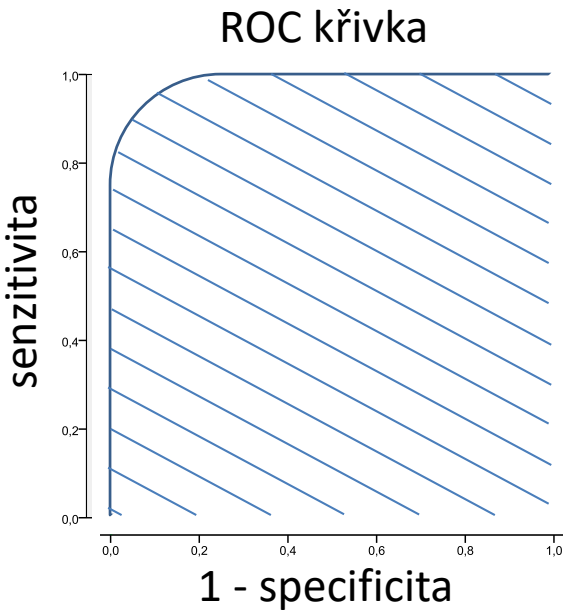
- Princip: Jakákoli hodnota spojité proměnné nějak rozlišuje zdravé a nemocné jedince, tzn. je spojena s nějakou senzitivitou a specificitou.



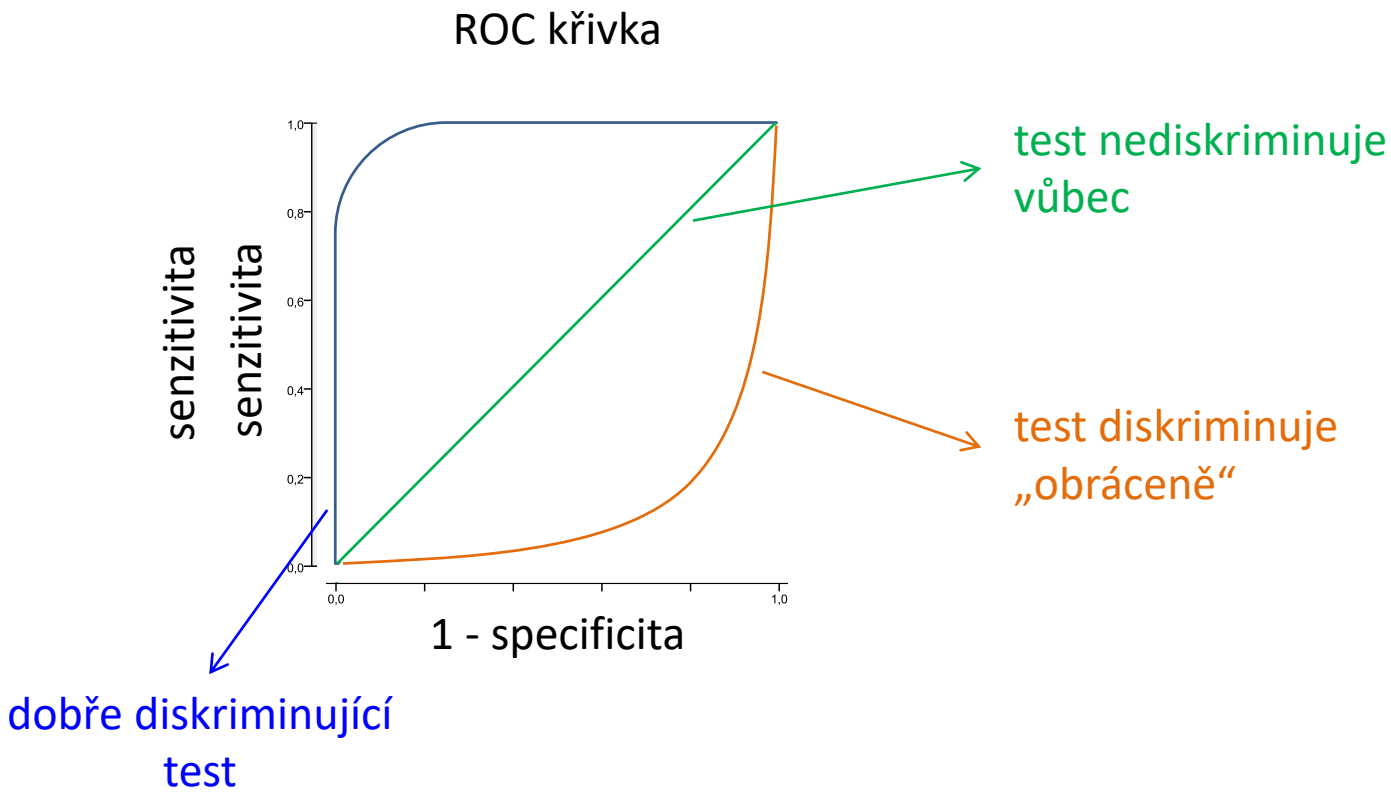
Nejlepší dělící bod („cut-off“) – nejvyšší sensitivita a specificita pro odlišení skupin – tzn. maximální součet hodnot senzitivity a specificity.

ROC analýza – plocha pod ROC křivkou

- Plocha pod ROC křivkou = „Area Under the Curve“ (AUC).
- Nabývá hodnot od 0 do 1.
- Slouží k vyjádření diagnostické síly (efektivity) testu.
- Čím větší hodnota AUC, tím lepší diagnostický test je (hodnota AUC nad 0,75 většinou poukazuje na uspokojivou diskriminační schopnost testu).

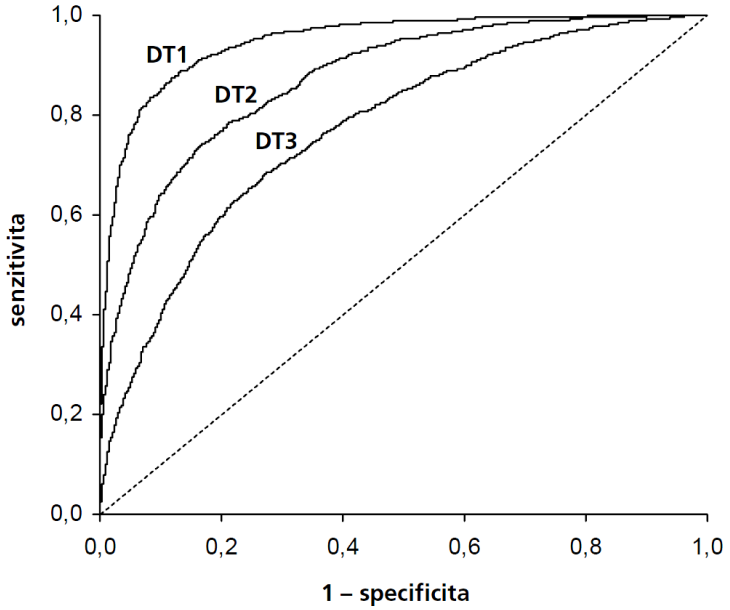


ROC analýza – srovnání diagnostické síly různých testů



ROC analýza – srovnání diagnostické síly různých testů

- Lze srovnat i velmi rozdílné testy (např. testy založené na různých proměnných).

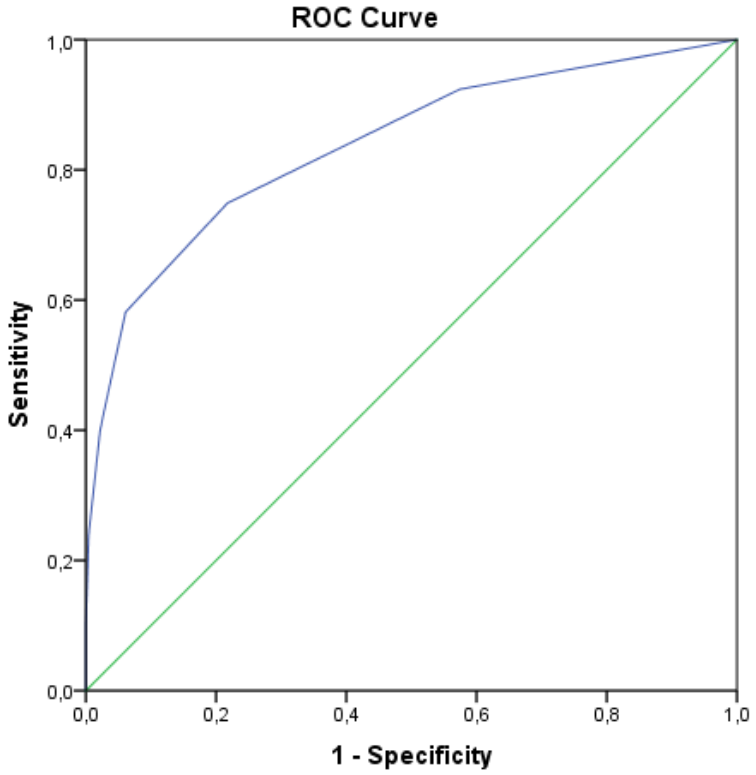


Diagnostický test	AUC	
DT1	0,949	→ nejlepší
DT2	0,872	
DT3	0,770	→ nejhorší

Zdroj: Dušek, Pavlík, Jarkovský, Koptíková, Analýza dat v Neurologii, Cesk Slov Neurol N 2011; 74/ 107(4)

ROC analýza – příklad

Příklad: Zjistěte, zda je MMSE skóre vhodné na diagnostiku mírné kognitivní poruchy (MCI). Najděte dělící bod (cut-off), který nejlépe odlišuje pacienty s MCI od kontrolních subjektů.



Area Under the Curve

Test Result Variable(s): MMSE

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,838	,016	,000	,807	,868

Coordinates of the Curve

Test Result Variable(s):

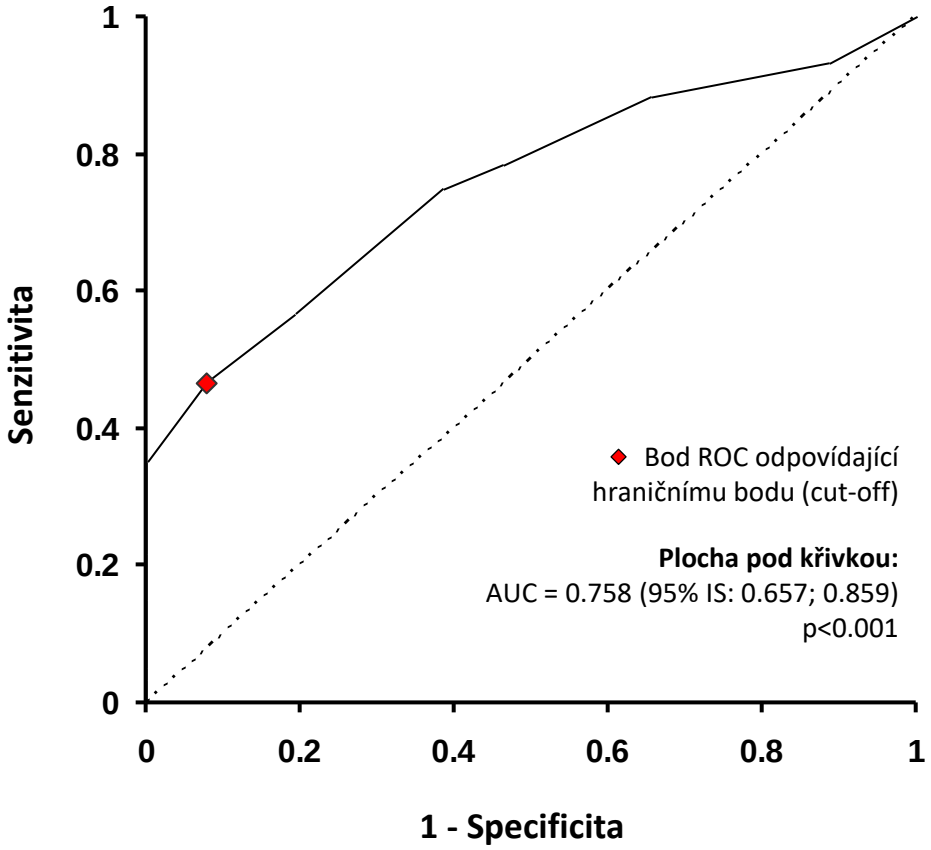
Positive if Less Than or Equal To[a]	Sensitivity	1 - Specificity	Specificity	Sensitivity + Specificity
22.00	0.000	0.000	1.000	1.000
23.50	0.002	0.000	1.000	1.002
24.50	0.101	0.000	1.000	1.101
25.50	0.239	0.004	0.996	1.235
26.50	0.399	0.022	0.978	1.377
27.50	0.581	0.061	0.939	1.520
28.50	0.749	0.217	0.783	1.531
29.50	0.924	0.574	0.426	1.350
31.00	1.000	1.000	0.000	1.000

ROC analýza – řešení v softwaru SPSS

- Analyze – ROC Curve – zadat Test Variable a State Variable (jako Value of State Variable zadat rizikovou kategorii)
- na záložce Options lze zvolit, zda „Larger test result indicates more positive test“ nebo „Smaller test result indicates more positive test“ – Continue
- zatržení „Standard error and confidence interval“ umožní k AUC vypočítat intervaly spolehlivosti a p-hodnotu
- zatržení „Coordinate points of the ROC Curve“ umožní získat tabulku se senzitivitou a 1-specificitou pro jednotlivé cut-off body (po zkopírování této tabulku do Excelu je možno vypočítat specificitu a nalézt nejlepší cut-off)

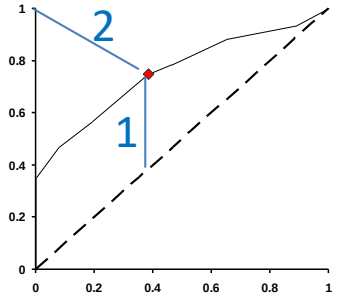
Hledání cut-off – doplnění

Příklad:



Sens	Spec	Sens+Spec
1.000	0.000	1.000
0.933	0.115	1.049
0.883	0.346	1.229
0.783	0.538	1.322
0.750	0.615	1.365
0.567	0.808	1.374
0.467	0.923	1.390
0.350	1.000	1.350
0.217	1.000	1.217
0.150	1.000	1.150
0.050	1.000	1.050
0.033	1.000	1.033
0.000	1.000	1.000

Hledání cut-off – kritéria



Kritérium	Vzoreček	Reference
1. Youdenova J statistika [1] – maximalizace vzdálenosti od diagonály	$\max(se + sp)$	• W. J. Youden (1950) "Index for rating diagnostic tests". Cancer, 3, 32–35. • R-kový balík pROC • http://www.medicalbiostatistics.com/roccurve.pdf
2. Nejbližší bod levému hornímu rohu grafu	$\min((1-se)^2+(1-sp)^2)$	• R-kový balík pROC • http://www.medicalbiostatistics.com/roccurve.pdf
3. Maximalizace součinu senzitivity a specificity	$\max(se * sp)$	• R-kový balík OptimalCutpoints • dr. Budíková používá maximalizaci geometrického průměru sens a spec

[1] Youdenova J statistika je definována jako: $J = se + sp - 1$; při hledání maxima lze ale člen (-1) zanedbat

Hledání cut-off – vážená kritéria (dle R balíku pROC)

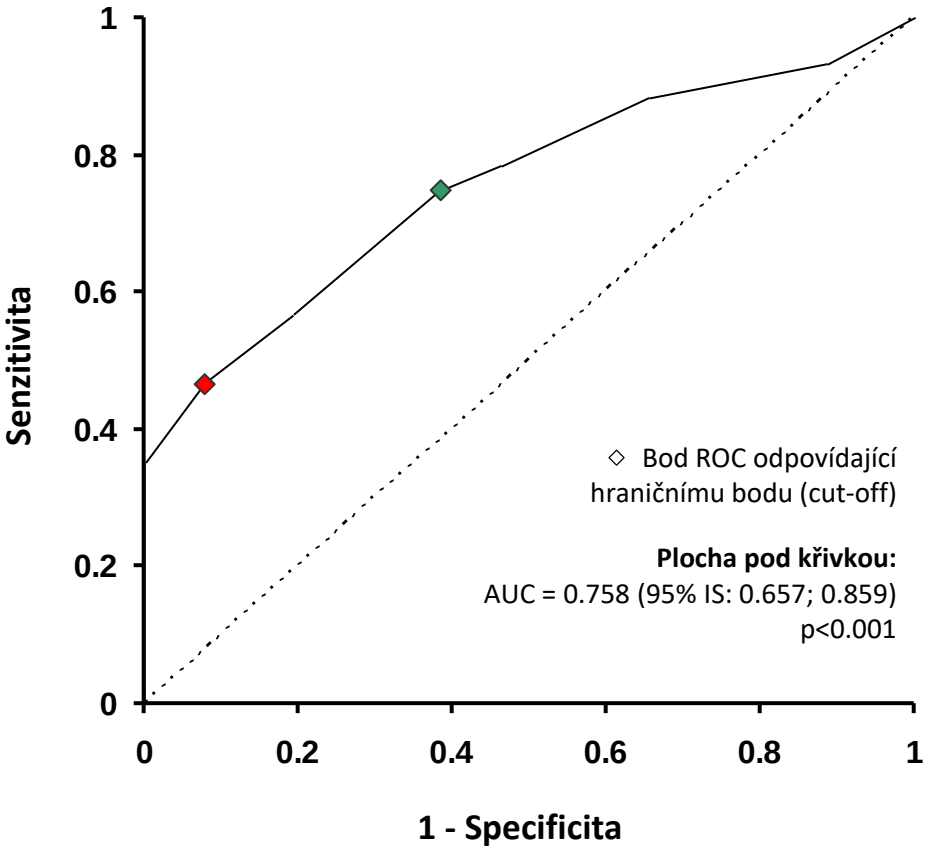
Kritérium	Vzoreček
Youdenova J statistika [1] – maximalizace vzdálenosti od diagonály	$\max(se + r * sp)$
Nejbližší bod levému hornímu rohu grafu	$\min((1-se)^2 + r * (1-sp)^2)$

kde: $r = \frac{1 - prevalence}{cost * prevalence}$ $prevalence = \frac{n_{cases}}{n_{cases} + n_{controls}}$

$cost$ – penalizace falešně negativních výsledků

defaultně: $prevalence = 0{,}5$ a $cost = 1$

Hledání cut-off – doplnění II



Sens	Spec	Sens+ Spec	closest. topleft	Sens* Spec
1.000	0.000	1.000	1.000	0.000
0.933	0.115	1.049	0.787	0.108
0.883	0.346	1.229	0.441	0.306
0.783	0.538	1.322	0.260	0.422
0.750	0.615	1.365	0.210	0.462
0.567	0.808	1.374	0.225	0.458
0.467	0.923	1.390	0.290	0.431
0.350	1.000	1.350	0.423	0.350
0.217	1.000	1.217	0.614	0.217
0.150	1.000	1.150	0.723	0.150
0.050	1.000	1.050	0.903	0.050
0.033	1.000	1.033	0.934	0.033
0.000	1.000	1.000	1.000	0.000