

FSTA: Pokročilé statistické metody

Stochastické modelování s binárním endpointem

Jiří Jarkovský, Simona Littnerová

FSTA: Pokročilé statistické metody

Binární endpointy klinických dat

Mortalita a přežití: čím se liší?

- Základní endpointy v studiích na klinických datech
- Rozdílný význam i metodika analýzy

Pacient přijat
k hospitalizaci

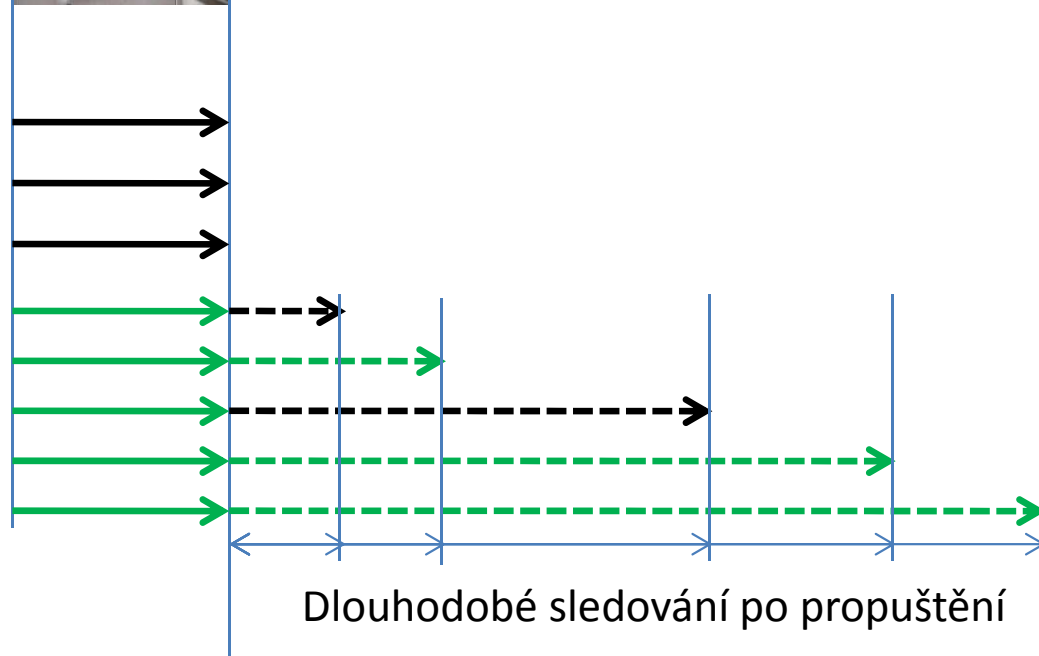


→ Úmrtí za hospitalizace

→ Žijící při propuštění

Analýza hospitalizační mortality

- bez vlivu času
- logistická regrese aj.



--> Úmrtí

- -> Žijící

Analýza přežití

- klíčový je vliv doby sledování
- cenzorování pacientů
- Kaplan-Meier, Coxova regrese

Pacient propuštěn nebo
zemřel za hospitalizace

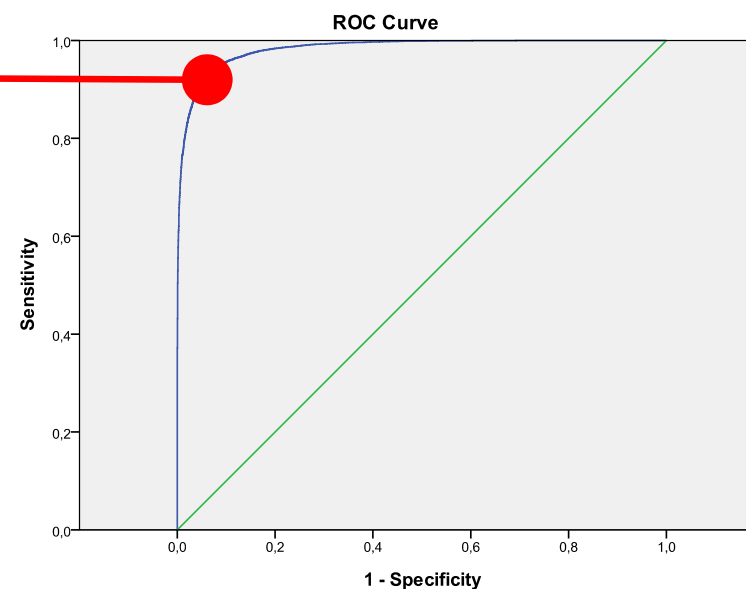
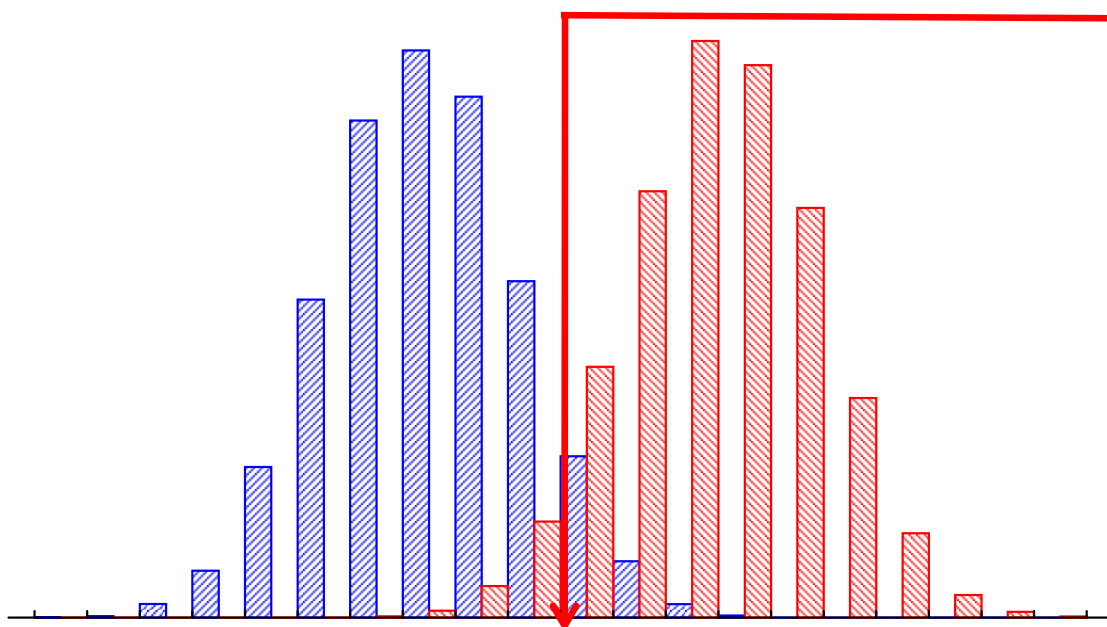
Analýza faktorů ovlivňujících mortalitu a přežití

- Identifikace faktorů zvyšujících riziko úmrtí je jedním z hlavních cílů analýzy dat v medicíně
- Podle typu endpointu (hospitalizační mortalita vs. dlouhodobé přežití) se odvíjí použité metody
- Dva obecné přístupy
 - Identifikace rizikových skupin pacientů a následná stratifikace do podsouborů (např. samostatná analýza kardiogenního šoku u ASS); modely jsou následně vytvářeny samostatně v rámci podsouborů
 - Vývoj modelů = rovnice, rozhodovací pravidla, neuronové sítě apod. kombinující prediktory (vysvětlující proměnné) za účelem vysvětlení endpointu (zde nejčastěji mortality)

ROC analýza

- Identifikace cutt offs pro kategorizaci spojitých proměnných aby při jejich užití v modelech byla maximalizována jejich sensitivita a specificita

Kde leží optimální hranice mezi skupinami?

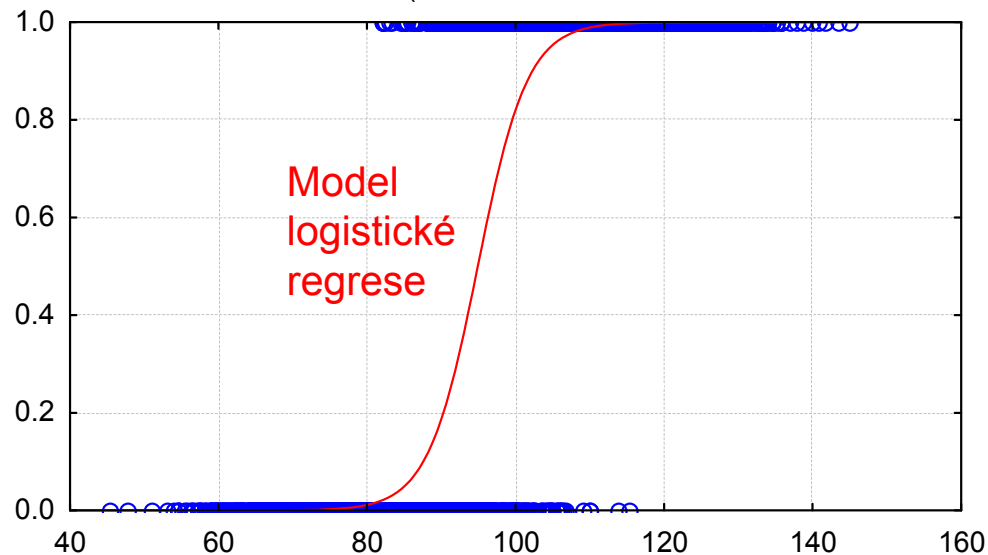


Identifikace hranice s nejvyšší sensitivitou a specificitou pro odlišení skupin

Logistická regrese

- Standardní metoda pro analýzu binárních charakteristik (zemřelý/žijící) bez vlivu času
- Modeluje závislost výskytu úmrtí na binárních, kategoriálních nebo spojitých proměnných
- Výsledkem rovnice je pravděpodobnost, že u daného pacienta nastane hodnocená endpoint
- Alternativou jsou např. rozhodovací stromy, neuronové sítě a další klasifikační metody

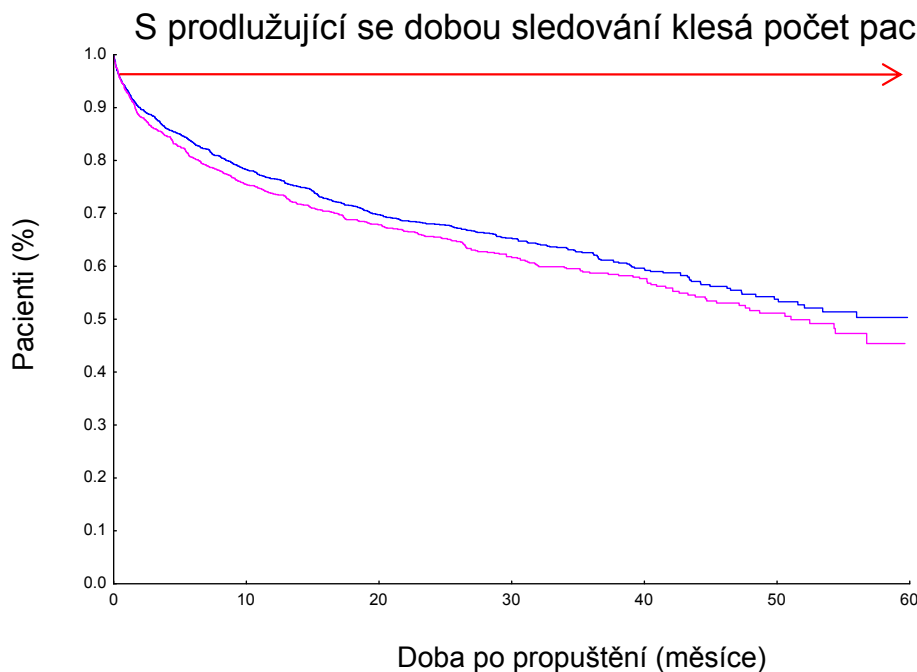
$$y = \frac{\exp(-28.41096581446 + (.29929760633475) * x)}{1 + \exp(-28.41096581446 + (.29929760633475) * x)}$$



Příklad logistické regrese: predikce binární charakteristiky (osa y) za pomoci spojitě proměnné (osa x)

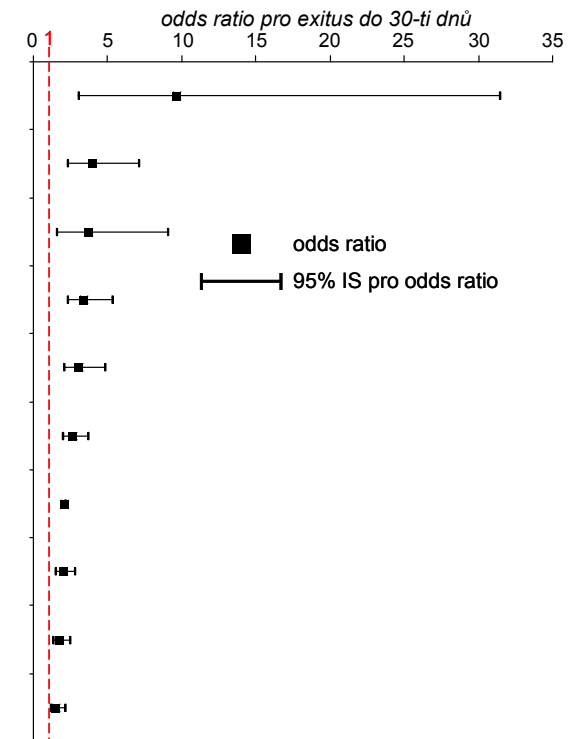
Coxův model proporcionálních rizik

- Standardní metoda pro analýzu dlouhodobého přežití (kromě dat o úmrtí musíme znát i dobu sledování pacientů)
- Modeluje závislost výskytu úmrtí na binárních, kategoriálních nebo spojitých proměnných při započítání doby sledování a cenzorování pacientů
- Doba sledování je klíčovou součástí výpočtu, sledovaný efekt musí mít „čas se projevit“



Odds ratio, relativní riziko a hazard ratio

- Různé způsoby vyjádření rizika pacientů
 - s danou vlastností (muži vs. ženy, kategorie NYHA, zestárnutí o 10 let)
 - pro výskyt události (mortalita, komplikace atd.)
- Většinou vyjádřeny jako OR/RR/HR (95% IS)
- **ODDS RATIO**
 - Spjato s logistickou regresí nebo kontingenčními tabulkami
 - Retrospektivní studie
- **RELATIVNÍ RIZIKO**
 - Odvozeno od kontingenčních tabulek
 - Prospektivní studie
- **HAZARD RATIO**
 - Spjato s Coxovým modelem proporcionálních rizik
 - Studie pracující s přežitím a cenzorovanými pacienty



Podobný, ale nikoliv identický význam.

Matematické vlastnosti výhodné pro různé účely.

Vazba na určité statistické metody a typy studií.

Relative risk vs. Odds ratio ?

Relative risk
(relativní riziko)



Odds ratio
(poměr šancí)

- Smysl RR a OR
- Výpočet
- Srovnatelnost
- Interpretace
- Výhody a nevýhody

- Aplikace v klinickém hodnocení

Smysl RR a OR

- Popis vlivu faktoru (léčba, klinický parametr) na výskyt události (úmrtí, progrese aj.)

Relative risk
(relativní riziko)



Odds ratio
(poměr šancí)

- ✓ Snadná přirozená interpretace rizik vyjádřených jako procento událostí

ALE

- ✓ Matematická omezení pro některé aplikace

- ✓ Pouze málo lidí má přirozenou schopnost interpretovat OR

ALE

- ✓ OR v řadě aplikací výhodnější matematické vlastnosti

Výpočet

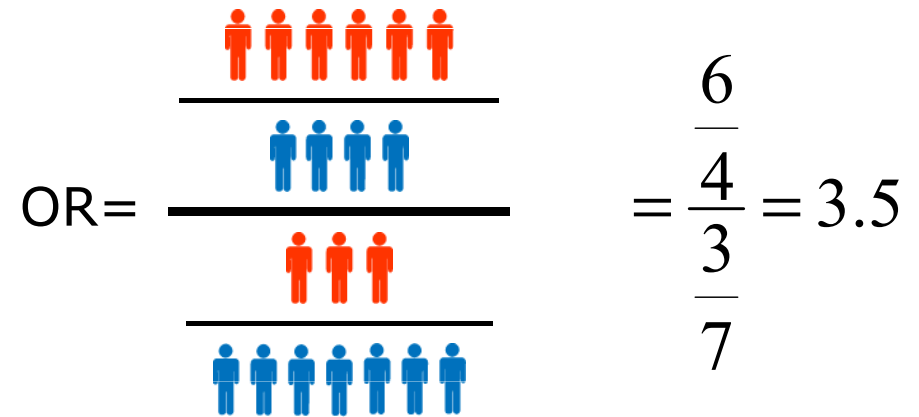
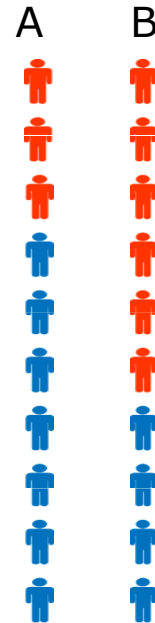
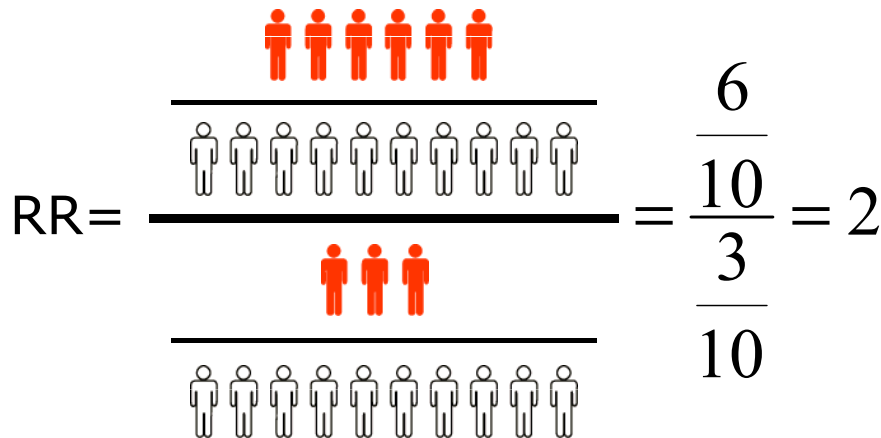
 event
  bez eventů

- Srovnání výskytu události mezi dvěma rameny (A,B) studie

Relative risk
(relativní riziko)



Odds ratio
(poměr šancí)

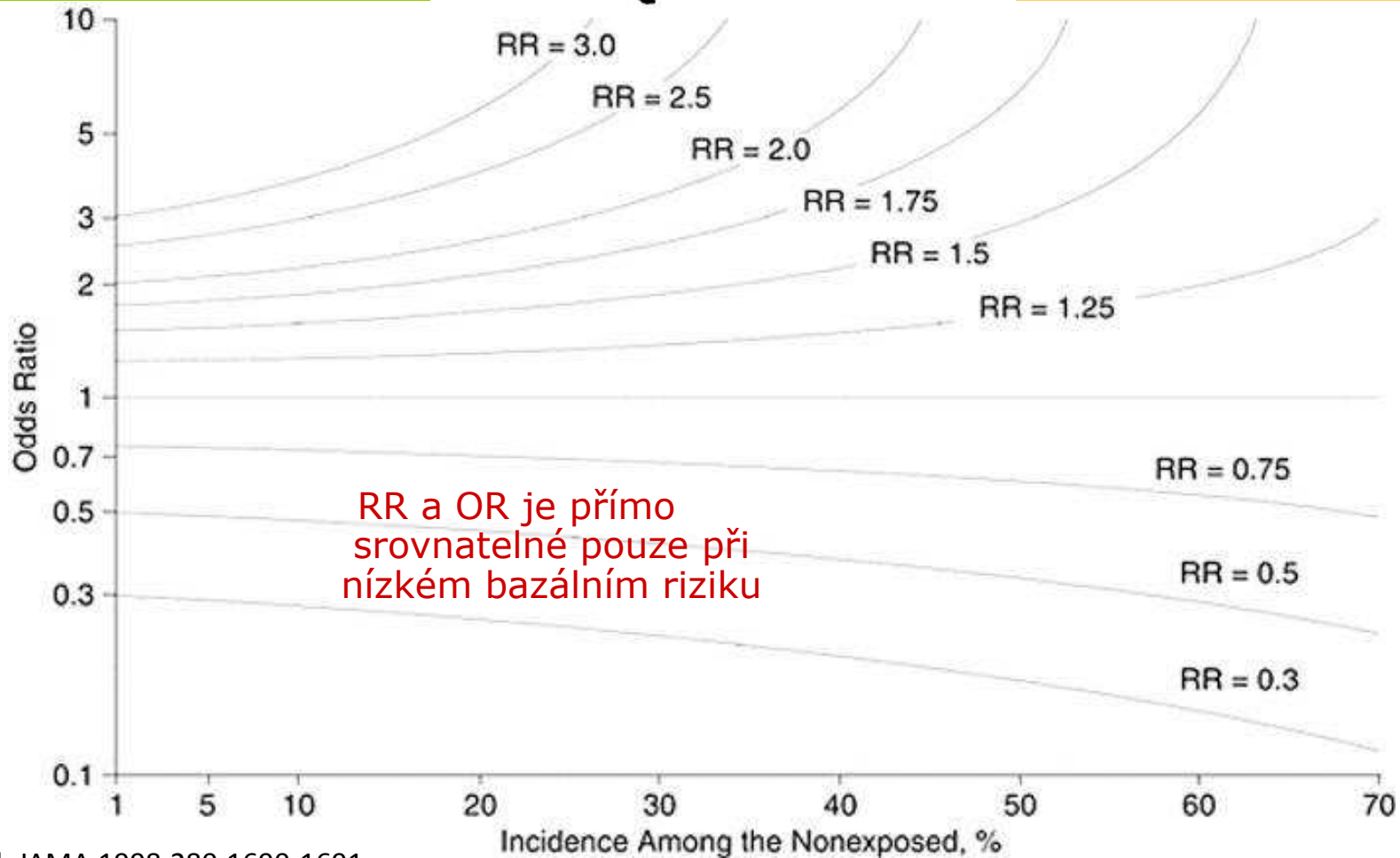


Vztah mezi RR a OR

Relative risk
(relativní riziko)



Odds ratio
(poměr šancí)



Zhang, J. et al. JAMA 1998;280:1690-1691.

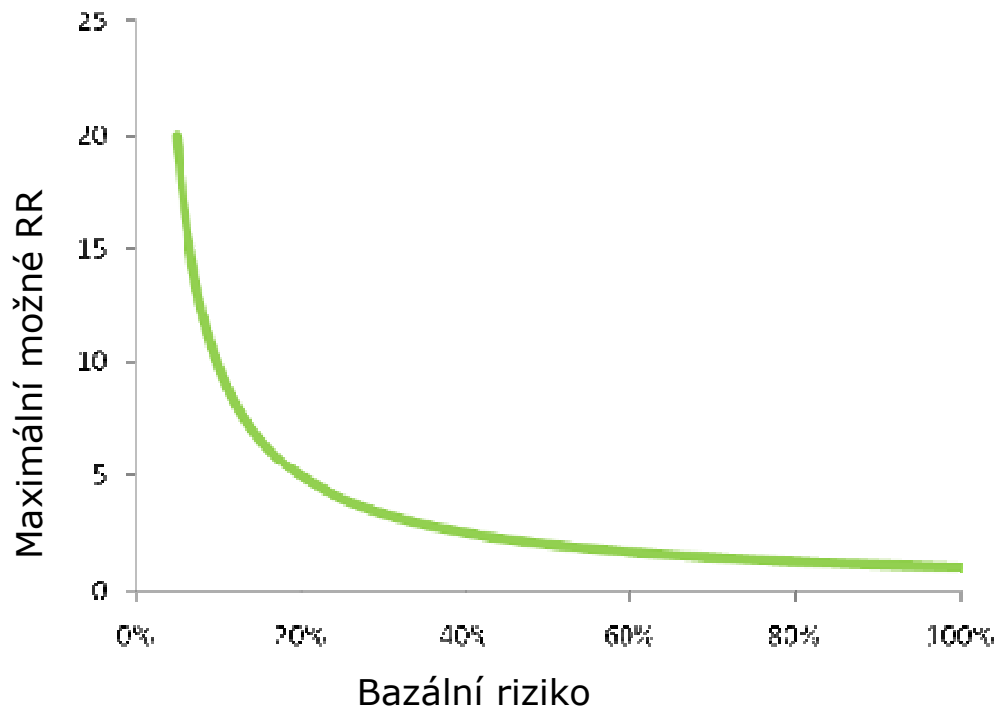
Srovnatelnost RR a OR I: maximum

Relative risk
(relativní riziko)



Odds ratio
(poměr šancí)

- ✓ RR mění své maximum podle bazálního rizika



- ✓ RR ve studiích s různým bazálním rizikem jsou nesrovnatelná !!!!

- ✓ Odds ratio má vždy rozsah od 0 do nekonečna
- ✓ Velikost OR není závislá na velikosti bazálního rizika

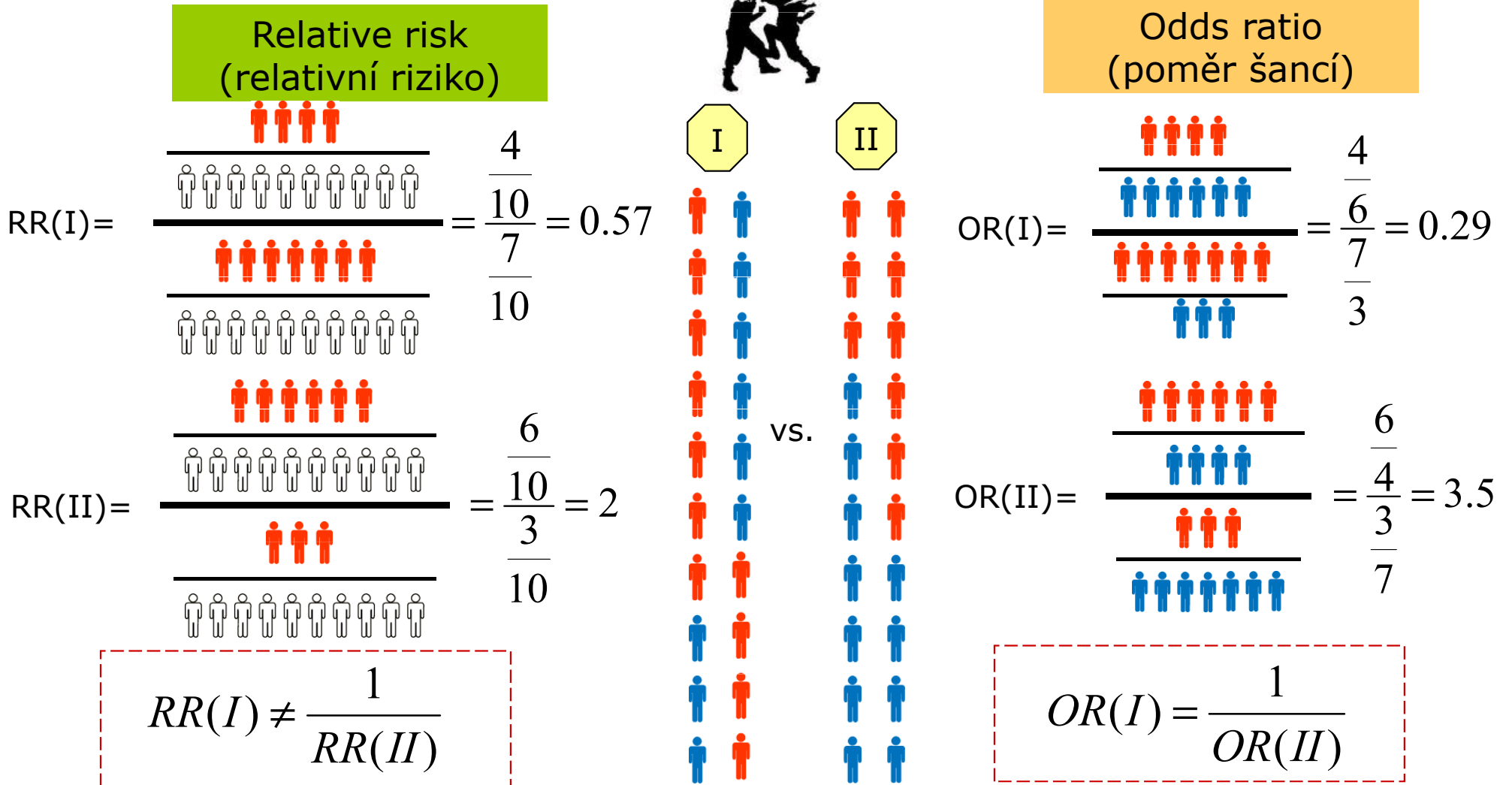
- ✓ OR lze použít pro srovnání studií s různým bazálním rizikem !!!!

- ✓ Výhodné pro metaanalýzu

Srovnatelnost RR a OR II: symetrie

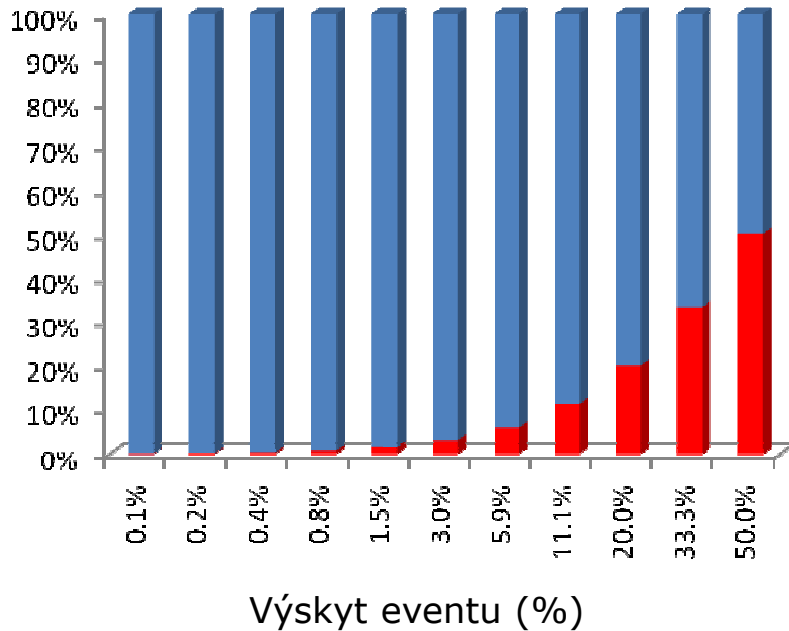
- Existuje mezi RR a O rozdíl v případě

výměny definice eventu a non-eventu?

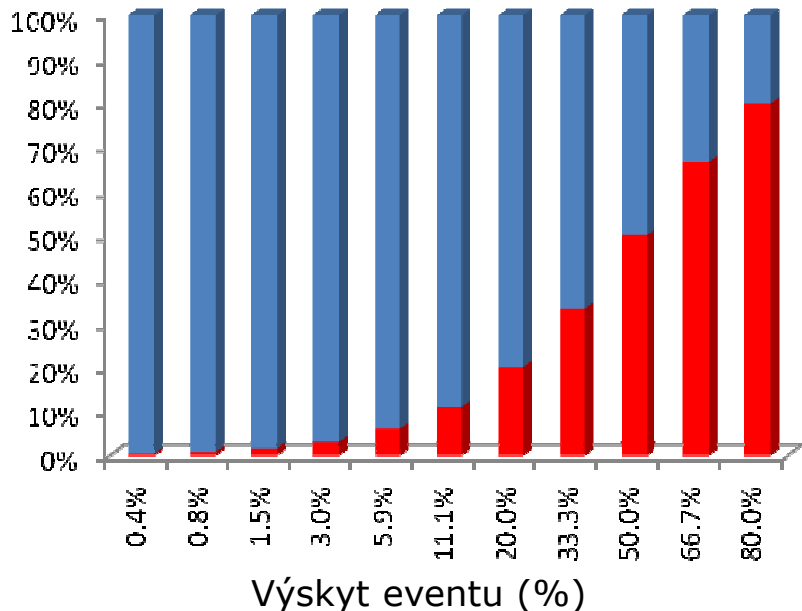


RR a OR ve studiích s různou mírou bazálního rizika

Control

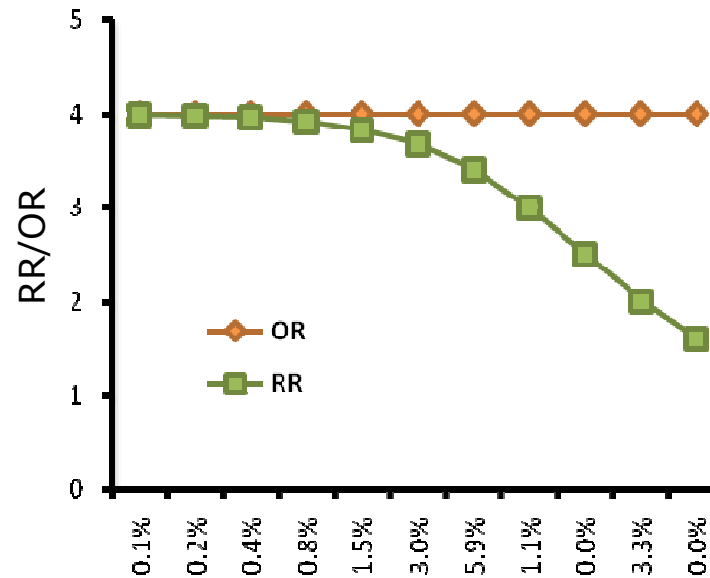


Case



Odds ratio

Ve skupině „Case“ připadá na jednoho pacienta bez eventu 4x tolik pacientů s eventem než ve skupině „Control“



Bazální (control) výskyt eventu (%)

Relative risk

Pacient ve skupině „Case“ má x-krát zvýšenou pravděpodobnost výskytu eventu než pacient ve skupině „Control“. X-krát závisí na bazálním výskytu eventu.

RR a OR v prospektivních a retrospektivních studiích

Prospektivní studie

- ✓ Sledování výskytu eventů a následná analýza jeho příčin
- ✓ Převážně kohortní studie



- ✓ Bazální výskyt eventů je dán vlastnostmi kohorty pacientů
- ✓ Bezproblémové využití RR



Relative risk
(relativní riziko)

Retrospektivní studie

- ✓ **Zpětné sledování příčin eventů**
- ✓ **Převážně case-control studie**
- ✓ **Výběrem pacientů ovlivňujeme bazální výskyt eventů**



- ✓ RR nelze použít – ovlivněno bazálním výskytem eventů
- ✓ Využití OR – není ovlivněno designem studie



Odds ratio
(poměr šancí)

Relative risk vs. Odds ratio: shrnutí

Relative risk (relativní riziko)



Odds ratio (poměr šancí)

- ☑ Intuitivně snadno interpretovatelné
- ☑ Pro prospektivní studie
- ☑ Maximum se liší podle bazální hodnoty výskytu eventů

- ☑ Retrospektivní studie
- ☑ Aplikace v metaanalýze
- ☑ Standardní výstup logistické regrese
- ☑ Rozsah vždy 0 až nekonečno, není ovlivněno bazálním výskytem eventů
- ☑ Obtížnější interpretace

Hazard ratio

- ☑ Významově leží mezi RR a OR
- ☑ Standardní výstup Coxova modelu proporcionálních rizik

FSTA: Pokročilé statistické metody

ROC analýza

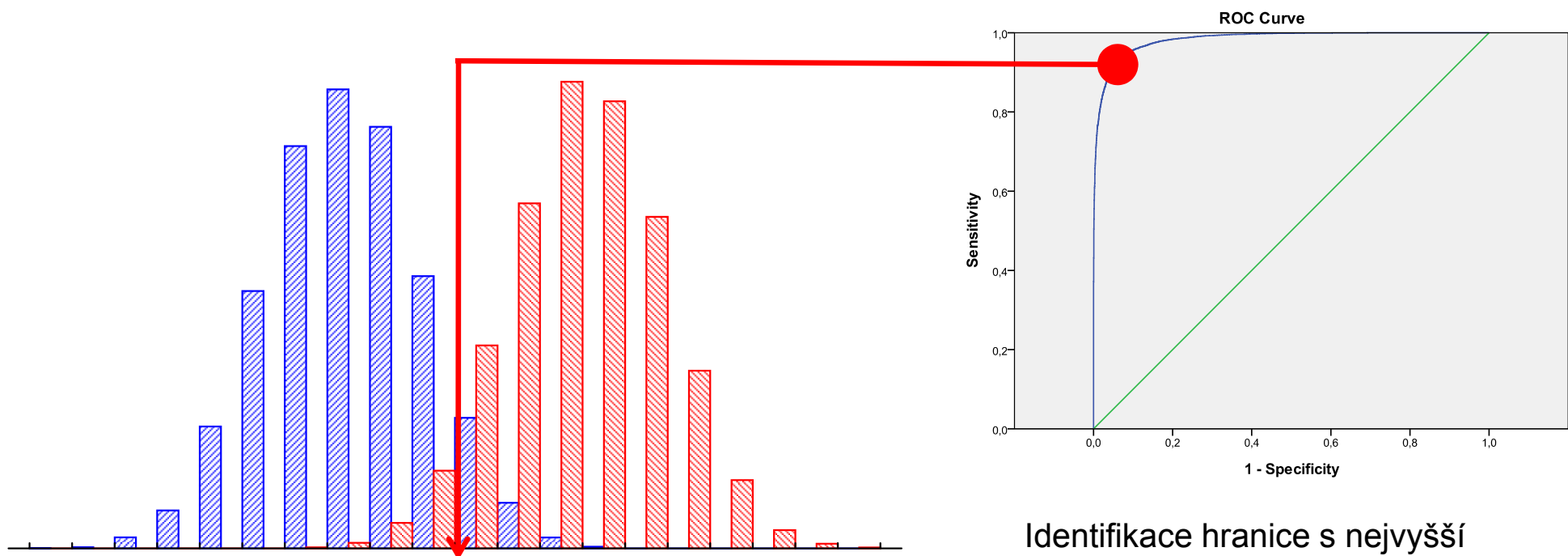
ROC analýza

- Nástroj pro identifikaci cut-off (hranice rozdělení spojitých dat) ve spojitých datech vzhledem k co nejlepšímu odlišení binárního endpointu
- Výsledkem je binarizace spojitě proměnné, která je často lépe interpretovatelná než výsledky na spojitých datech
- Identifikace konkrétního cut-off souvisí s preferencí buď sensitivity nebo specificity pro identifikaci endpointu
- Upřednostnění sensitivity nebo specificity je do určité míry subjektivní dle reálného cíle analýzy
 - Vysoká sensitivity – screeningový test, kdy je třeba zachytit všechny možné nemocné (např. závažné onemocnění, které je třeba zachytit v počátečním stadiu)
 - Vysoká specificita – pokud je nezbytné odchytit pouze skutečně nemocné pacienty (např. nechceme vystavovat pacienty zbytečné léčbě málo závažného onemocnění)

ROC analýza

- Identifikace cutt offs pro kategorizaci spojitých proměnných aby při jejich užití v modelech byla maximalizována jejich sensitivita a specificita

Kde leží optimální hranice mezi skupinami?



Identifikace hranice s nejvyšší sensitivitou a specificitou pro odlišení skupin

Sensitivita a specificita

- Klíčové pojmy v popisu vztahu dvou binárních proměnných = situace kdy predikujeme binární endpoint binárním prediktorem

| | 1 – nemocný | 0 - zdravý |
|------------------------|--------------------|--------------------|
| 1 – riziková skupina | Skutečně pozitivní | Falešně pozitivní |
| 0 – neriziková skupina | Falešně negativní | Skutečně negativní |

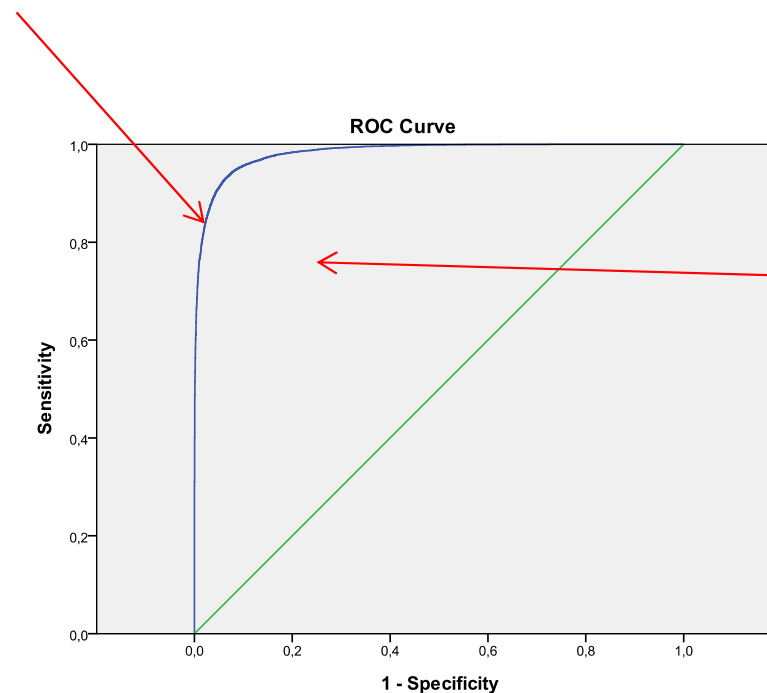
$$\textit{sensitivita} = \frac{\textit{skutečně pozitivní}}{\textit{skutečně pozitivní} + \textit{falešně negativní}}$$

$$\textit{specificita} = \frac{\textit{skutečně negativní}}{\textit{skutečně negativní} + \textit{falešně pozitivní}}$$

Výstupy ROC

Sensitivita a specificita v každém bodě křivky – mohou být doplněny o IS

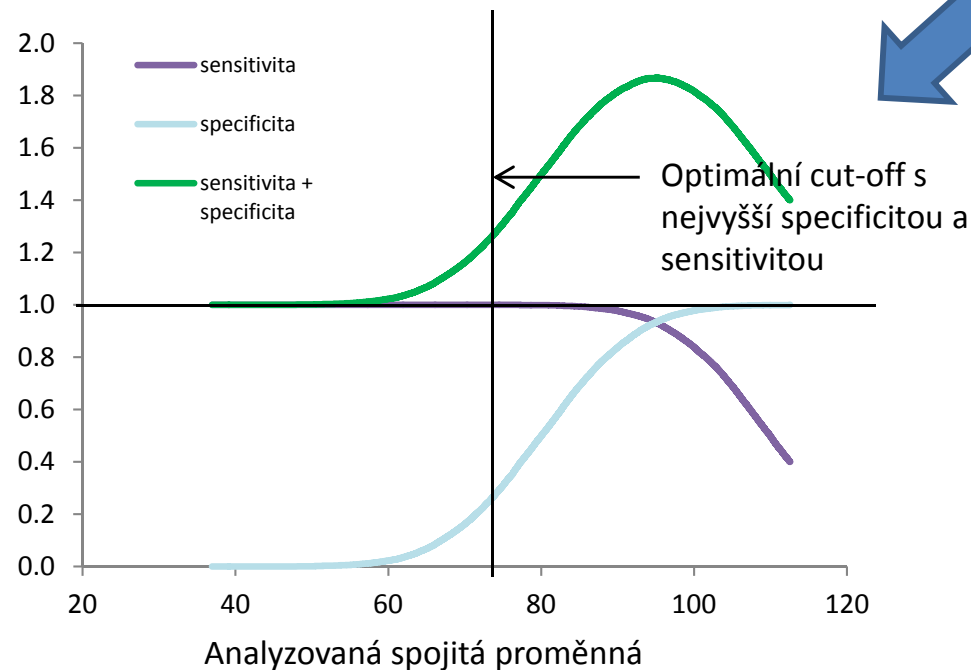
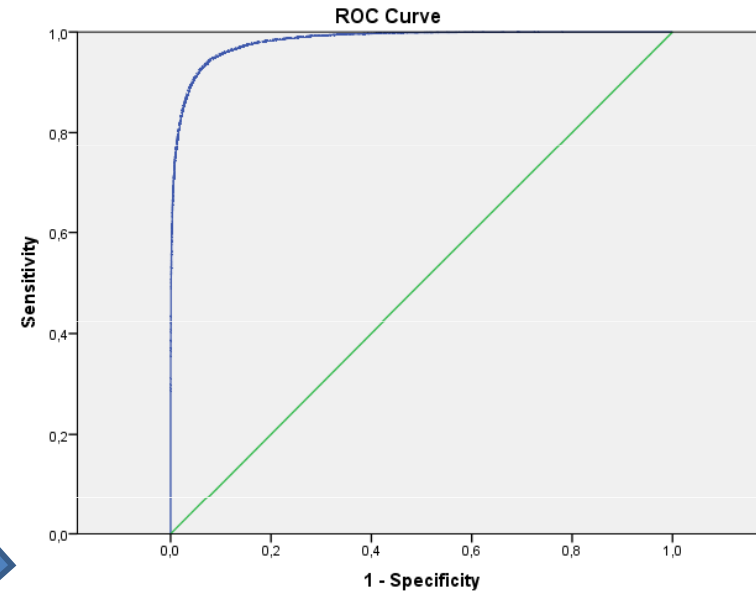
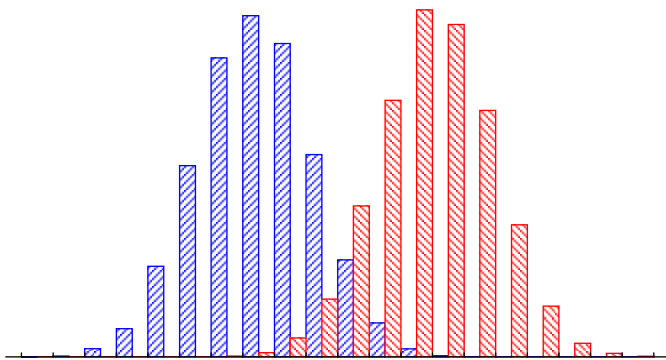
Nejlepší kombinace sensitivity a specificity určuje příslušný dělicí bod spojité proměnné
Při identifikaci cut-off je třeba také kontrolovat, aby výsledná riziková skupina neobsahovala pouze minimum hodnot (cut-off oddělující jednoho pacinta nemá téměř smysl)



AUC (plocha pod křivkou) + IS
Čím odlišnější od 0.5, tím lepší identifikace endpointu
Testování významnosti AUC

ROC – příklad I

Odlišení dvou skupin pacientů
(modří=zdraví; červení=nemocní)



Area Under the Curve

Test Result Variable(s): Var1

| Area | Std. Error ^a | Asymptotic Sig. ^b | Asymptotic 95% Confidence Interval | |
|------|-------------------------|------------------------------|------------------------------------|-------------|
| | | | Lower Bound | Upper Bound |
| ,983 | ,000 | ,000 | ,982 | ,984 |

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

FSTA: Pokročilé statistické metody

Logistická regrese

Logistická regrese

- Logistická regrese je základním nástrojem pro analýzu závislosti binárního endpointu (úmrtí, komplikace atd.) na spojitých nebo binárních prediktorech
- Cílem analýzy je:
 - Identifikace vztahů mezi prediktory a endpointem a jejich popis (odds ratio)
 - Vytvoření predikčního modelu umožňujícího zařazení pacientů do hodnocených skupin
- Logistická regrese patří do skupiny zobecněných lineárních modelů (lineární statistické modely s linkovací funkcí)

Princip logistické regrese I

- V logistické regresi modelujeme vliv spojitych nebo binárních prediktorů na endpoint s binomickým rozdělením - > není možné použít klasickou lineární regresi
- Predikujeme pravděpodobnost výskytu jevu pomocí rovnice:

$$P(x) = \frac{\exp(a + b * x)}{1 + \exp(a + b * x)}$$

exp(rovnice)

- Kde $\frac{\exp(rovnice)}{1 + \exp(rovnice)}$ je tzv. logit, linkovací funkce pro logistickou regresi a rovnice $a+b*x$ je použitý lineární model
- Pojem linkovací funkce je spjat se zobecněnými lineárními modely, kdy linkovací funkce převádí problém nelineární závislosti y na x na lineární model
- Zjednodušeně řečeno „nelineární vztah=linkovací funkce(lineární model)“
- Zobecněný lineární model s linkovací funkcí „identita“ = lineární model

Princip logistické regrese II

- Model logistické regrese lze popsat také jako:
tedy, že logaritmus šance na výskyt jevu
lze popsat jeho lineární závislostí na x

$$\log \left[\frac{P(x)}{1 - P(x)} \right] = a + b * x$$

- S využitím této znalosti můžeme popsat jak se mění šance na výskyt jevu (např. úmrtí) při změně x , konkrétně poměr šancí pro změnu x o Δ

$$\log \left[\frac{\frac{P(x + \Delta)}{1 - P(x + \Delta)}}{\frac{P(x)}{1 - P(x)}} \right] = b * \Delta$$
$$\frac{\frac{P(x + \Delta)}{1 - P(x + \Delta)}}{\frac{P(x)}{1 - P(x)}} = \exp(b * \Delta)$$

- Pro $\Delta=1$, tak získáme hodnotu odds ratioa pro změnu x o 1 jako $\exp(\text{regresní koeficient } x)$

$$\frac{\frac{P(x + 1)}{1 - P(x + 1)}}{\frac{P(x)}{1 - P(x)}} = \exp(b)$$

Kategoriální data jako prediktory v logistické regresi

- Kategoriální a ordinální data mohou do analýzy vstupovat jako binární proměnné
- Kategoriální data (nelze seřadit) -> dummies
- Ordinální data (lze seřadit)
 - Dummies
 - Definice referenční kategorie (obvykle kategorie s nejnižším rizikem pro hodnocený endpoint)
- Příklad: The New York Heart Association (NYHA) Functional Classification

| Původní | Dummies | | | | Vzhledem k referenci | | | |
|---------|---------|--------|---------|----------|----------------------|-------------|--------------|-------------|
| | NYHA | NYHA I | NYHA II | NYHA III | NYHA IV | NYHA II ref | NYHA III ref | NYHA IV ref |
| I | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| I | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| I | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| II | 0 | 1 | 0 | 0 | 0 | 1 | | |
| II | 0 | 1 | 0 | 0 | 0 | 1 | | |
| III | 0 | 0 | 0 | 0 | 0 | | 1 | |
| III | 0 | 0 | 0 | 0 | 0 | | 1 | |
| IV | 0 | 0 | 0 | 1 | 1 | | | 1 |
| IV | 0 | 0 | 0 | 1 | 1 | | | 1 |

Výstupy logistické regrese

- Koeficienty rovnice logistické regrese
- Odds ratio jako popis rizikovosti prediktorů pro výskyt endpointu

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|---------|-------|--------|----|------|---------|---------------------|----------|
| | | | | | | | Lower | Upper |
| Step 1 ^a SEPALLEN | 5,140 | 1,007 | 26,080 | 1 | ,000 | 170,773 | 23,748 | 1228,028 |
| Constant | -27,831 | 5,434 | 26,236 | 1 | ,000 | ,000 | | |

a. Variable(s) entered on step 1: SEPALLEN.

- Popis celkových výsledků a kvality modelu

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|---------------------|----------------------|---------------------|
| 1 | 64,211 ^a | ,525 | ,700 |

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

| Step | Chi-square | df | Sig. |
|------|------------|----|------|
| 1 | 6,832 | 8 | ,555 |

Classification Table^a

| | | Predicted | | |
|----------|--------------------|------------|----|--------------------|
| | | VERSICOL | | Percentage Correct |
| Observed | ,00000000 | 1,00000000 | | |
| Step 1 | VERSICOL ,00000000 | 45 | 5 | 90,0 |
| | 1,00000000 | 6 | 44 | 88,0 |
| | Overall Percentage | | | 89,0 |

a. The cut value is .500

Odds ratio

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|---------|-------|--------|----|------|---------|---------------------|----------|
| | | | | | | | Lower | Upper |
| Step 1 ^a SEPALLEN | 5,140 | 1,007 | 26,080 | 1 | ,000 | 170,773 | 23,748 | 1228,028 |
| Constant | -27,831 | 5,434 | 26,236 | 1 | ,000 | ,000 | | |

a. Variable(s) entered on step 1: SEPALLEN.

- Popisuje míru rizika spjatou:
 - **U spojitých proměnných** se změnou hodnoty o 1 (z tohoto důvodu se spojité proměnné často převádí na interpretovatelné jednotky – např. věk po desetiletích, koncentrace po stovkách jednotek)
 - **U binárních proměnných** spjatých s výskytem vlastnosti (kódováno jako 1)
 - U klasických dummies jde o riziko vůči všem ostatním pacientům bez dané vlastnosti
 - U binárních proměnných kódovaných vůči referenční kategorii jde o nárůst oproti pacientům v referenční kategorii
- Odds ratio je exponenciální hodnota koeficientu regresní rovnice

Hodnocení kvality modelu I

- Klasifikace objektů na základě vytvořené rovnice logistické regrese, analyzován je úspěch klasifikace (zcela korektní by pak bylo toto testování na nezávislém souboru)
- Standardně je jako hranice pro klasifikaci využíváno $p=0.5$ a je možné ji popsat standardními ukazateli vztahu dvou binárních proměnných jako je sensitivita, specificita, pozitivní a negativní prediktivní hodnota apod.

Classification Table^a

| | | Predicted | | |
|----------|--------------------|------------|----|--------------------|
| | | VERSICOL | | Percentage Correct |
| Observed | ,00000000 | 1,00000000 | | |
| Step 1 | VERSICOL ,00000000 | 45 | 5 | 90,0 |
| | 1,00000000 | 6 | 44 | 88,0 |
| | Overall Percentage | | | 89,0 |

a. The cut value is .500

Hodnocení kvality modelu II

- **Hosmer & Lemeshow test**
- Jde o aplikaci chi-square testu na výsledky predikce logistické regrese
- Jsou srovnávány pozorované počty pacientů v rizikových skupinách vs. počty očekávané dle výsledků logistického modelu (obvykle na bázi decilů rizikové funkce)
- V případě shody pozorovaných a očekávaných četností je model považován za dobře kalibrovaný

Hosmer and Lemeshow Test

| Step | Chi-square | df | Sig. |
|------|------------|----|------|
| 1 | 6,832 | 8 | ,555 |

Contingency Table for Hosmer and Lemeshow Test

| | | VERSICOL = .00000000 | | VERSICOL = 1.00000000 | | Total |
|--------|----|----------------------|----------|-----------------------|----------|-------|
| | | Observed | Expected | Observed | Expected | |
| Step 1 | 1 | 11 | 10,862 | 0 | ,138 | 11 |
| | 2 | 9 | 9,464 | 1 | ,536 | 10 |
| | 3 | 8 | 8,938 | 2 | 1,062 | 10 |
| | 4 | 8 | 7,508 | 1 | 1,492 | 9 |
| | 5 | 9 | 6,756 | 2 | 4,244 | 11 |
| | 6 | 2 | 4,131 | 10 | 7,869 | 12 |
| | 7 | 3 | 1,794 | 8 | 9,206 | 11 |
| | 8 | 0 | ,454 | 10 | 9,546 | 10 |
| | 9 | 0 | ,087 | 10 | 9,913 | 10 |
| | 10 | 0 | ,006 | 6 | 5,994 | 6 |

Hodnocení kvality modelu III

- -Pseudo R²
 - Nejde o přímou obdobu R² u lineární regrese
 - Popisuje zlepšení výsledků modelu oproti modelu pouze s interceptem (tedy bez vlivu prediktoru)
 - Existuje řada algoritmů pro výpočet pseudo R²
 - Cox & Snell – analyzuje zlepšení výsledků modelu oproti samotnému interceptu

$$R^2 = 1 - \left[\frac{L(M_{intercept})}{L(M_{kompletni})} \right]^{2/N}$$

- Nagelkerke – adjustuje Cox & Snell na maximální možnou hodnotu

$$R^2 = \frac{1 - \left[\frac{L(M_{intercept})}{L(M_{kompletni})} \right]^{2/N}}{1 - L(M_{intercept})^{2/N}}$$

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|---------------------|----------------------|---------------------|
| 1 | 64,211 ^a | ,525 | ,700 |

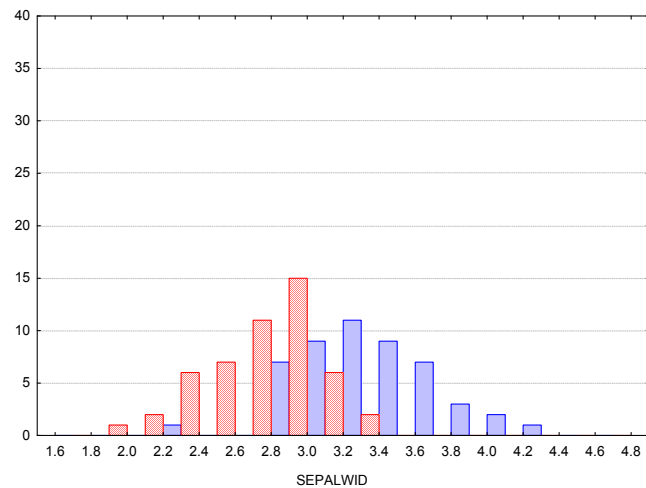
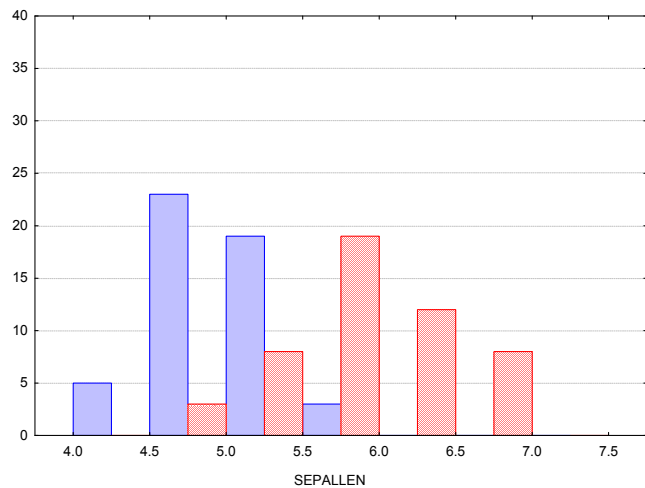
a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Vícerozměrná logistická regrese

- Požadavky na kvalitní predikční model
 - Maximální predikční síla
 - Maximální interpretovatelnost
 - Minimální složitost
- Tvorba modelů
 - Neobsahuje redundantní proměnné
 - Je otestován na nezávislých datech
- Výběr proměnných
 - Algoritmy typu dopředné a zpětné eliminace jsou pouze pomocným ukazatelem při výběru proměnných finálního modelu
 - Při výběru proměnných se uplatní jak klasické statistické metody (ANOVA), tak expertní znalost významu proměnných a jejich zastupitelnosti

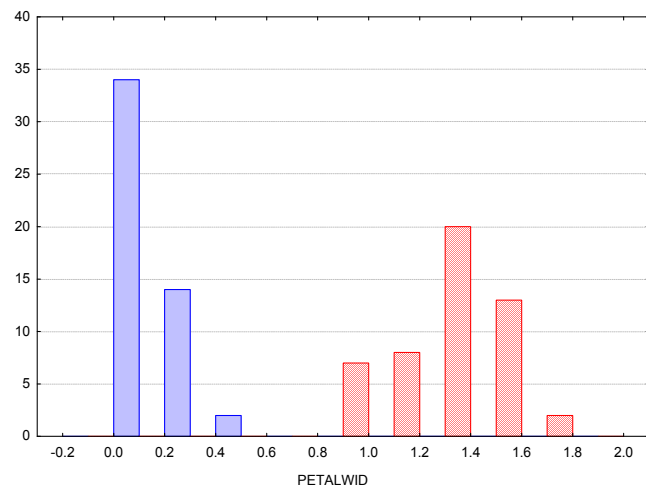
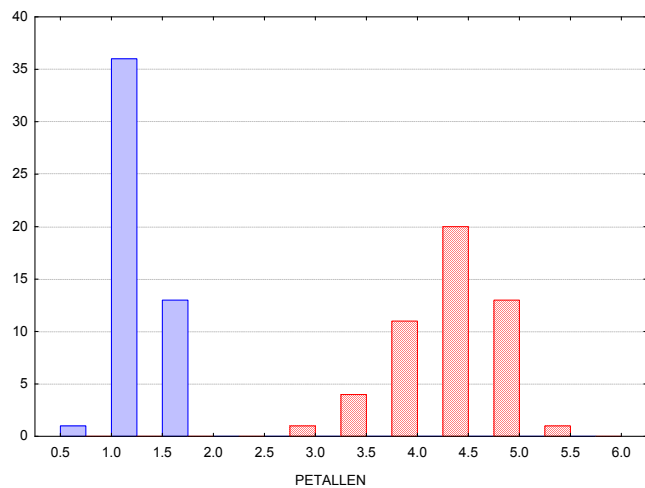
Vícerozměrná logistická regrese: příklad I

- Identifikace dvou druhů kosatců: VERSICOL vs. ostatní

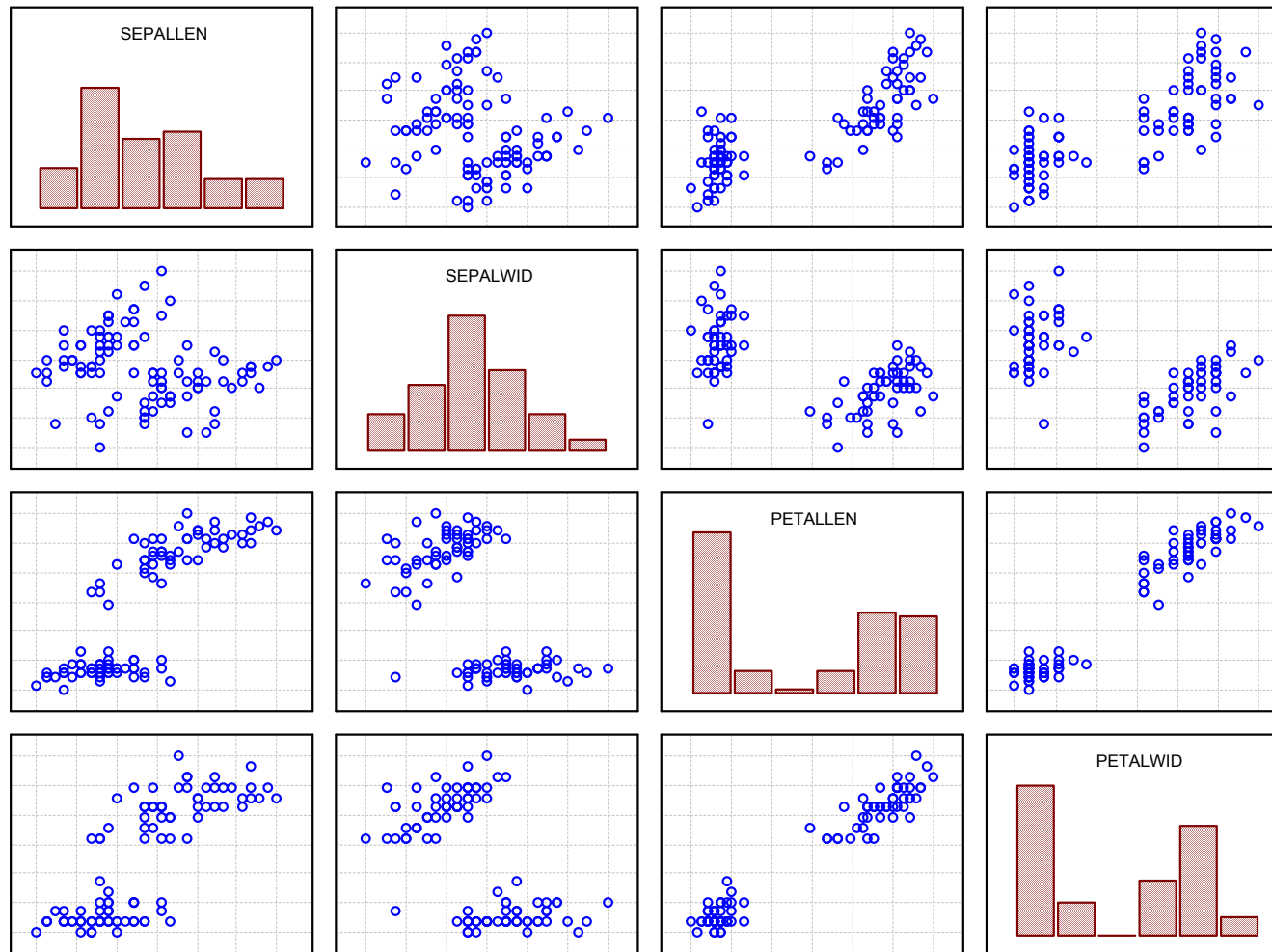


VERSICOL

Lze jednoznačně pozorovat různou diskriminační schopnost různých proměnných v jednorozměrné analýze



Vícerozměrná logistická regrese: příklad II



Lze jednoznačně pozorovat korelace prediktorů

Vícerozměrná logistická regrese: příklad III

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|---------|-------|--------|----|------|---------|---------------------|----------|
| | | | | | | | Lower | Upper |
| Step 1 ^a SEPALLEN | 5,140 | 1,007 | 26,080 | 1 | ,000 | 170,773 | 23,748 | 1228,028 |
| Constant | -27,831 | 5,434 | 26,236 | 1 | ,000 | ,000 | | |

a. Variable(s) entered on step 1: SEPALLEN.

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|--------|-------|--------|----|------|---------|---------------------|-------|
| | | | | | | | Lower | Upper |
| Step 1 ^a SEPALWID | -6,552 | 1,350 | 23,552 | 1 | ,000 | ,001 | ,000 | ,020 |
| Constant | 20,230 | 4,165 | 23,594 | 1 | ,000 | 6,109E8 | | |

a. Variable(s) entered on step 1: SEPALWID.

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|---------|----------|------|----|------|----------|---------------------|-------|
| | | | | | | | Lower | Upper |
| Step 1 ^a PETALLEN | 29,384 | 3484,594 | ,000 | 1 | ,993 | 5,771E12 | ,000 | . |
| Constant | -72,391 | 8843,695 | ,000 | 1 | ,993 | ,000 | | |

a. Variable(s) entered on step 1: PETALLEN.

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|---------|----------|------|----|------|----------|---------------------|-------|
| | | | | | | | Lower | Upper |
| Step 1 ^a PETALWID | 83,548 | 9248,229 | ,000 | 1 | ,993 | 1,924E36 | ,000 | . |
| Constant | -65,795 | 7499,174 | ,000 | 1 | ,993 | ,000 | | |

a. Variable(s) entered on step 1: PETALWID.

Vícerozměrná logistická regrese: příklad IV

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|--------|-----------|------|----|-------|-------------|---------------------|-------|
| | | | | | | | Lower | Upper |
| Step 1 ^a SEPALLEN | -6,991 | 21075,525 | ,000 | 1 | 1,000 | ,001 | ,000 | . |
| SEPALWID | -6,107 | 10266,630 | ,000 | 1 | 1,000 | ,002 | ,000 | . |
| PETALLEN | 14,168 | 15685,907 | ,000 | 1 | ,999 | 1422296,317 | ,000 | . |
| PETALWID | 19,646 | 23995,058 | ,000 | 1 | ,999 | 3,407E8 | ,000 | . |
| Constant | 3,838 | 65766,953 | ,000 | 1 | 1,000 | 46,418 | | |

a. Variable(s) entered on step 1: SEPALLEN, SEPALWID, PETALLEN, PETALWID.

Variables in the Equation

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|------------------------------|---------|-----------|------|----|-------|-------------|---------------------|-------|
| | | | | | | | Lower | Upper |
| Step 1 ^a SEPALLEN | -6,991 | 21075,525 | ,000 | 1 | 1,000 | ,001 | ,000 | . |
| SEPALWID | -6,107 | 10266,630 | ,000 | 1 | 1,000 | ,002 | ,000 | . |
| PETALLEN | 14,168 | 15685,907 | ,000 | 1 | ,999 | 1422296,317 | ,000 | . |
| PETALWID | 19,646 | 23995,058 | ,000 | 1 | ,999 | 3,407E8 | ,000 | . |
| Constant | 3,838 | 65766,953 | ,000 | 1 | 1,000 | 46,418 | | |
| Step 2 ^a SEPALWID | -9,304 | 6916,884 | ,000 | 1 | ,999 | ,000 | ,000 | . |
| PETALLEN | 9,530 | 11139,407 | ,000 | 1 | ,999 | 13771,081 | ,000 | . |
| PETALWID | 24,744 | 27802,702 | ,000 | 1 | ,999 | 5,573E10 | ,000 | . |
| Constant | -15,605 | 23339,954 | ,000 | 1 | ,999 | ,000 | | |
| Step 3 ^a SEPALWID | -11,772 | 6256,715 | ,000 | 1 | ,998 | ,000 | ,000 | . |
| PETALLEN | 20,869 | 3186,452 | ,000 | 1 | ,995 | 1,157E9 | ,000 | . |
| Constant | -17,010 | 19097,759 | ,000 | 1 | ,999 | ,000 | | |
| Step 4 ^a PETALLEN | 29,384 | 3484,594 | ,000 | 1 | ,993 | 5,771E12 | ,000 | . |
| Constant | -72,391 | 8843,695 | ,000 | 1 | ,993 | ,000 | | |

a. Variable(s) entered on step 1: SEPALLEN, SEPALWID, PETALLEN, PETALWID.

Logistická regrese: shrnutí

- Základní nástroj pro identifikaci faktorů ovlivňujících výskyt binárních endpointů a predikci individuální pravděpodobnosti výskytu endpointů
- Použitelná jako obdoba diskriminační analýzy pro 2 skupiny
- Popisuje míru rizikovosti prediktorů pro binární endpoint ve formě odds ratio
- Pro vícerozměrné modely je důležité analyzovat redundanci parametrů a stabilitu vícerozměrných modelů
- Pro praktické nasazení modelů je nezbytná jejich krosvalidace, popřípadě jiné metody testování nasazení modelů na nezávislých datech
- Neumí pracovat s cenzorovanými daty
- **Standardní metodika analýzy rizikových faktorů pro binární endpointy (hospitalizační mortalita apod.)**