

## Model binární logistické regrese pro lékařská data

Sestavte model binární logistické regrese, který pro náhodně vybraného pacienta umožní predikovat pravděpodobnost, že se u něj vyskytne neklid po celkové anestézii. Závisle proměnnou veličinou je tedy neklid upravený (varianty 1 – vyskytl se neklid po celkové anestézii, 2 – nevyskytl se neklid po CA). Na výskyt neklidu mohou mít vliv tři kategoriální veličiny druh léku, ASA, Novalgin ano/ne a čtyři spojité veličiny věk, hmotnost, dávka, poměrová dávka. To posoudíme pomocí testů nezávislosti a pomocí dvouvýběrových testů. Veličiny, jejichž p-hodnota bude menší než 0,25, zařadíme do modelu binární logistické regrese. Přitom u veličiny druh léku bude referenční kategorií Rapifen, u veličiny ASA kategorie II a u veličiny Novalgin ano/ne varianta ne.

**Úkol 1.:** Na hladině významnosti 0,05 testujte hypotézy, že neklid upravený a kategoriální veličiny druh léku, ASA, Novalgin ano/ne jsou nezávislé. Nezapomeňte ověřit splnění podmínek dobré aproximace.

Výsledek pro neklid upravený x druh léku:

Kontingenční tabulka (Nalbuphin_Rapifen.sta)				
Tab. :				
	Neklid upraveny	Druh léku Nalbuphin	Druh léku Rapifen	Řádk. součty
Četnost	neklid byl	11	23	34
Sloupc. četn.		19,64%	39,66%	
Četnost	neklid nebyl	45	35	80
Sloupc. četn.		80,36%	60,34%	
Četnost	Vš.skup.	56	58	114

Souhrnná tab.: Očekávané četnosti (Nalbuphin_Rapifen.sta)			
Pearsonův chí-kv. : 5,45188, sv=1, p=,019547			
Neklid upraveny	Druh léku Nalbuphin	Druh léku Rapifen	Řádk. součty
neklid byl	16,70175	17,29825	34,0000
neklid nebyl	39,29825	40,70175	80,0000
Vš.skup.	56,00000	58,00000	114,0000

Hypotézu o nezávislosti zamítáme na asymptotické hladině významnosti 0,05.

### Výsledek pro neklid upravený x ASA:

Kontingenční tabulka (Nalbuphin_Rapifen.sta)				
Tab. :				
	Neklid upraveny	ASA I	ASA II	Řádk. součty
Četnost	neklid byl	31	3	34
Sloupc. četn.		31,00%	21,43%	
Četnost	neklid nebyl	69	11	80
Sloupc. četn.		69,00%	78,57%	
Četnost	Vš.skup.	100	14	114

Souhrnná tab.: Očekávané četnosti (Nalbuphin_Rapifen.sta)				
Pearsonův chí-kv. : ,537548, sv=1, p=,463451				
Neklid upraveny	ASA I	ASA II	Řádk. součty	
neklid byl	29,8246	4,17544	34,0000	
neklid nebyl	70,1754	9,82456	80,0000	
Vš.skup.	100,0000	14,00000	114,0000	

Hypotézu o nezávislosti nezamítáme na asymptotické hladině významnosti 0,05.

### Výsledek pro neklid upravený x Novalgin ano/ne:

Kontingenční tabulka (Nalbuphin_Rapifen.sta)				
Tab. :				
	Neklid upraveny	Novalgin ano/ne 0	Novalgin ano/ne 1	Řádk. součty
Četnost	neklid byl	9	25	34
Sloupc. četn.		28,13%	30,49%	
Četnost	neklid nebyl	23	57	80
Sloupc. četn.		71,88%	69,51%	
Četnost	Vš.skup.	32	82	114

Souhrnná tab.: Očekávané četnosti (Nalbuphin_Rapifen.sta)				
Pearsonův chí-kv. : ,061398, sv=1, p=,804300				
Neklid upraveny	Novalgin ano/ne 0	Novalgin ano/ne 1	Řádk. součty	
neklid byl	9,54386	24,45614	34,0000	
neklid nebyl	22,45614	57,54386	80,0000	
Vš.skup.	32,00000	82,00000	114,0000	

Hypotézu o nezávislosti nezamítáme na asymptotické hladině významnosti 0,05.

**Úkol 2.:** Před provedením dvouvýběrových testů ověřte normalitu proměnných věk, hmotnost, dávka, poměrová dávka ve skupinách pacientů, u nichž se vyskytl resp. nevyskytl neklid.

Výsledky pro pacienty, u nichž se vyskytl neklid:

Proměnná	Testy normality (Nalbuphin_Rapifen.sta)				
	Zhrnout podmínku: v15=1				
	N	max D	Lilliefors p	W	p
Věk	34	0,233684	p < ,01	0,832491	0,000116
Hmotnost	34	0,311151	p < ,01	0,715116	0,000001
Dávka	34	0,333163	p < ,01	0,724877	0,000001
Dávka mg/kg	34	0,372417	p < ,01	0,697266	0,000000

Výsledky pro pacienty, u nichž se nevyskytl neklid:

Proměnná	Testy normality (Nalbuphin_Rapifen.sta)				
	Zhrnout podmínku: v15=2				
	N	max D	Lilliefors p	W	p
Věk	80	0,202938	p < ,01	0,815354	0,000000
Hmotnost	80	0,173416	p < ,01	0,746045	0,000000
Dávka	80	0,224512	p < ,01	0,855445	0,000000
Dávka mg/kg	80	0,266870	p < ,01	0,763710	0,000000

Ve všech případech Lillieforsův i Shapirův – Wilkův test zamítá hypotézu o normalitě. Dále tedy použijeme neparametrické testy.

**Úkol 3.:** Na hladině významnosti 0,05 testujte dvouvýběrovým Wilcoxonovým testem, že rozložení proměnných věk, hmotnost, dávka, poměrová dávka ve skupinách pacientů, u nichž se vyskytl resp. nevyskytl neklid, je stejné.

Výsledky dvouvýběrového Wilcoxonova testu:

Proměnná	Mann-Whitneyův U Test (w oprava na spojitost) (Nalbuphin_Rapifen.sta)									
	Dle proměn. Neklid upraveny Označené testy jsou významné na hladině p < ,05000									
	Sčt poř. neklid byl	Sčt poř. neklid nebyl	U	Z	p-hodn.	Z upravené	p-hodn.	N platn. neklid byl	N platn. neklid nebyl	2*1str. přesné p
Věk	1710,000	4845,000	1115,000	-1,51438	0,129929	-1,53785	0,124087	34	80	0,130388
Hmotnost	1801,000	4754,000	1206,000	-0,95075	0,341733	-0,95292	0,340629	34	80	0,343429
Dávka	1558,000	4997,000	963,000	-2,45584	0,014056	-2,47110	0,013470	34	80	0,013560
Dávka mg/kg	1464,500	5090,500	869,500	-3,03496	0,002406	-3,03682	0,002391	34	80	0,002130

Na hladině významnosti 0,05 se prokázal rozdíl u proměnných dávka a poměrová dávka. U proměnné věk je p-hodnota menší než 0,25, proto ji do modelu zařadíme.

**Úkol 4.:** Porovnejte devianci nulového modelu s deviancí modelu se čtyřmi nezávisle proměnnými veličinami (druh léku, věk, dávka, poměrová dávka) a významnost poklesu testujte na hladině významnosti 0,05.

Devianci nulového modelu získáme tak, že ve vstupní tabulce pro logistickou regresi zadáme pouze závisle proměnnou veličinu Neklid upravený a žádné nezávisle proměnné veličiny.

Zvolíme Kvalita proložení a vybereme Statistiku kvality modelu.

Zjistíme, že deviance nulového modelu je 138,9365.

Deviance modelu se čtyřmi uvažovanými regresory je 130,0793, došlo tedy k nepříliš výraznému poklesu deviance. Test poměrem věrohodnosti poskytl p-hodnotu 0,0648, což je větší než hladina významnosti 0,05. Vidíme, že tato cesta nepovede k vybudování kvalitního modelu.

**Úkol 4.:** Vytvořte model se všemi sedmi nezávisle proměnnými, vypočtete jeho devianci a významnost poklesu oproti nulovému modelu testujte na hladině významnosti 0,05.

Deviance modelu se sedmi uvažovanými regresory je 118,2647, došlo tedy k významnému poklesu deviance, protože test poměrem věrohodnosti poskytl p-hodnotu 0,0043.

**Úkol 5.:** Odhadněte parametry modelu a podle výsledku Waldova testu ponechte v modelu ty proměnné, pro něž jsou p-hodnoty menší než 0,25. Interpretujte podíly šancí v tomto novém modelu

Tabulka odhadů parametrů:

Neklid upraveny - Odhady parametrů (Nalbuphin_Rapifen.sta)								
Rozdělení : BINOMICKÉ, Linkující funkce: LOGIT								
Modelovaná pravděpodobnost, že Neklid upraveny = neklid byl								
Efekt	Úroveň Efekt	Sloupec	Odhad	Standard chyba	Wald. Stat.	Dolní LS 95,0%	Horní LS 95,0%	p
Abs.člen		1	-1,21016	1,00336	1,454692	-3,1767	0,75639	0,227777
Věk		2	-0,60297	0,24680	5,969086	-1,0867	-0,11926	0,014559
Hmotnost		3	0,17389	0,06729	6,678929	0,0420	0,30578	0,009756
Dávka		4	-0,86261	0,55172	2,444501	-1,9440	0,21874	0,117937
Dávka mg/kg		5	8,53431	15,09212	0,319769	-21,0457	38,11432	0,517147
Druh léku	Nalbuphin	6	-0,53602	1,76557	0,092169	-3,9965	2,92444	0,761437
ASA	I	7	0,24545	0,79510	0,095297	-1,3129	1,80381	0,757548
Novalgin ano/ne	0	8	1,36682	0,90154	2,298540	-0,4002	3,13381	0,129496
Měřítko			1,00000	0,00000		1,0000	1,00000	

V modelu ponecháme proměnné Věk, Hmotnost, Dávka, Novalgin.

Dostaneme novou tabulku odhadů parametrů:

Neklid upraveny - Odhady parametrů (Nalbuphin_Rapifen.sta)								
Rozdělení : BINOMICKÉ, Linkující funkce: LOGIT								
Modelovaná pravděpodobnost, že Neklid upraveny = neklid byl								
Efekt	Úroveň Efekt	Sloupec	Odhad	Standard chyba	Wald. Stat.	Dolní LS 95,0%	Horní LS 95,0%	p
Abs.člen		1	-0,790356	0,586483	1,816080	-1,93984	0,359129	0,177781
Věk		2	-0,607933	0,243202	6,248550	-1,08460	-0,131267	0,012429
Hmotnost		3	0,168010	0,064711	6,740808	0,04118	0,294842	0,009423
Dávka		4	-0,616158	0,202831	9,228171	-1,01370	-0,218616	0,002383
Novalgin ano/ne	0	5	1,357499	0,730567	3,452699	-0,07439	2,789385	0,063149
Měřítko			1,000000	0,000000		1,00000	1,000000	

### Tabulka podílů šancí:

Neklid upraveny - Poměry šancí (Nalbuphin_Rapifen.sta) Rozdělení : BINOMICKÉ, Linkující funkce: LOGIT Modelovaná pravděpodobnost, že Neklid upraveny = neklid byl						
Efekt	Úroveň Efekt	Sloupec	Šance Poměr	Dolní LS 95,0%	Horní LS 95,0%	p
Abs.člen		1				
Věk		2	0,544475	0,338037	0,87698	0,012429
Hmotnost		3	1,182949	1,042038	1,34291	0,009423
Dávka		4	0,540015	0,362874	0,80363	0,002383
Novalgin ano/ne	0	5	3,886462	0,928313	16,27101	0,063149
Měřítko			1,000000			

Pokud se věk pacienta zvýší o rok, poklesne šance na výskyt neklidu 0,54x. Podobně pro ostatní parametry.

### Úkol 6.: Proved'te H-S test a Pearsonův test dobré shody.

Neklid upraveny - Kvalita proložení: Hosmer-Lemeshow Test (Nalbuphin_Rapifen.sta) Rozdělení : BINOMICKÉ, Linkující funkce: LOGIT Hosmer Lemeshow = 6,3157, p hodn. = 0,611911											
Odezva	Skupi1a	Skupi2a	Skupi3a	Skupi4a	Skupi5a	Skupi6a	Skupi7a	Skupi8a	Skupi9a	Skupi10	Row Tot.
0: Pozorov.	11,0	10,0	9,0	8,0	10,0	8,0	6,0	8,0	3,0	7,0	80
Očekáv.	10,6	10,2	9,5	8,6	9,5	7,6	7,0	6,5	5,6	4,8	
1: Pozorov.	0,0	1,0	2,0	3,0	3,0	3,0	5,0	3,0	8,0	6,0	34
Očekáv.	0,4	0,8	1,5	2,4	3,5	3,4	4,0	4,5	5,4	8,2	
Vš. skup.	11,0	11,0	11,0	11,0	13,0	11,0	11,0	11,0	11,0	13,0	114

H-S test poskytl p-hodnotu 0,6119, tedy na hladině významnosti 0,05 nezamítáme hypotézu, že model souhlasí s daty.

Testová statistika Pearsonova testu nabyla hodnoty 104,6207.

Kritický obor  $W = \langle \chi^2_{0,95}(109), \infty \rangle = \langle 134,3688, \infty \rangle$ , tedy nulovou hypotézu nezamítáme na hladině významnosti 0,05.

### Úkol 7.: Vypoč'te Nagelkerkův koeficient.

Nagelkerkův koeficient nabývá hodnoty 0,2306, což svědčí o tom, že náš model není příliš vzdálen od nulového modelu.

### Úkol 5.: Sestavte klasifikační tabulku.

#### Klasifikační tabulka:

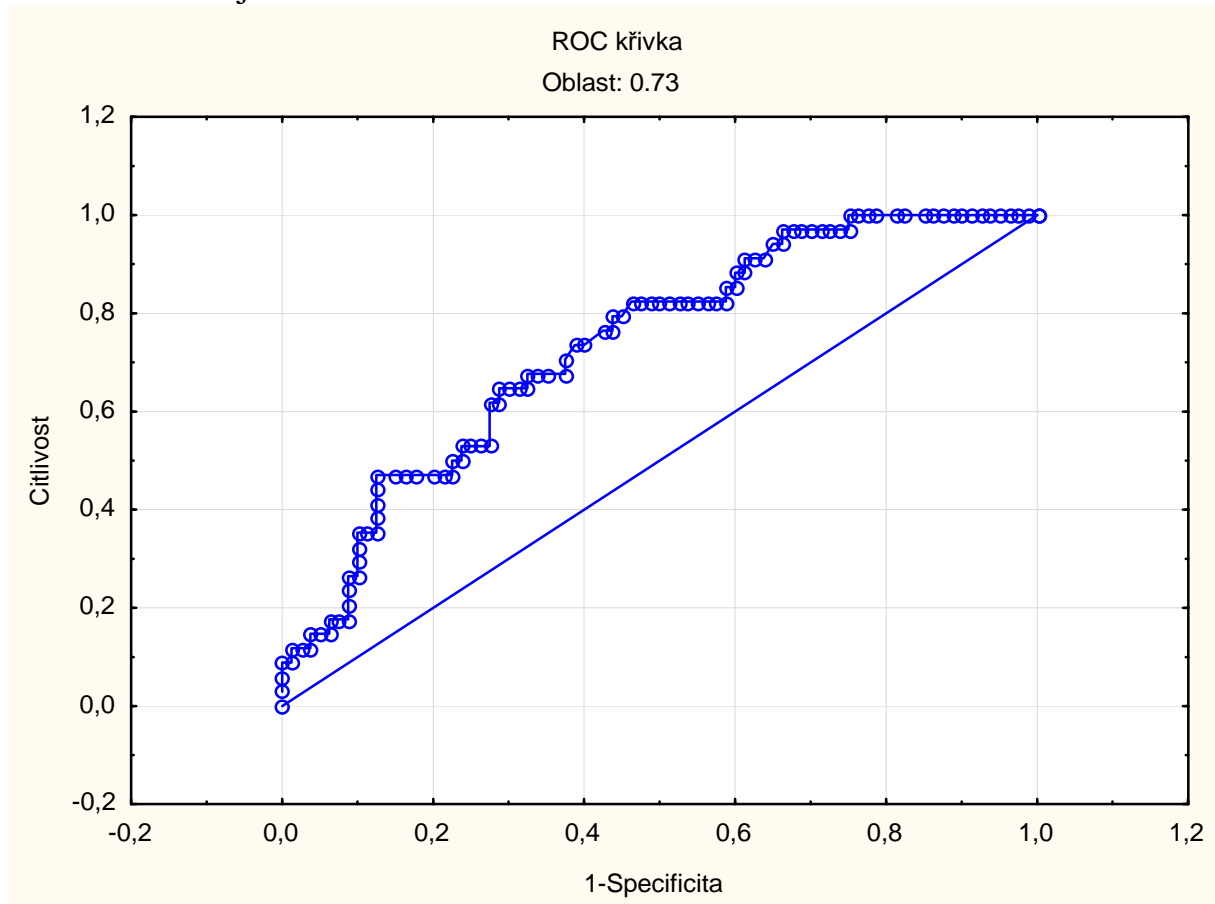
Klasifikace případů (Nalbuphin_Rapifen.sta) Odds ratio: 4,304348 Log odds ratio: 1,459626			
	Předpovězená: neklid byl	Předpovězená: neklid nebyl	Procento správných
Pozorované: neklid byl	11	23	32,3529412
Pozorované: neklid nebyl	8	72	90

Z 34 pacientů, u nichž se vyskytl neklid po CA, model správně zařadil 11 pacientů, tj. odhad senzitivity = 32,4 %.

Z 80 pacientů, u nichž se neklid po CA nevyskytl, model správně zařadil 72, tj. odhad specificity = 90 %.

Celkové procento úspěšné klasifikace je tedy  $83/114 = 72,8 \%$ .

**Úkol 8.:** Sestrojte ROC křivku.



**Úkol 9.:** Vyzkoušejte tvorbu modelu všemi dostupnými metodami, které jsou implementovány v systému STATISTICA a posuďte kvalitu dosažených výsledků.