

Šance na přežití u cestujících Titanicu z hlediska věku a palubní třídy

Tyto data poskytují detailní informace o osudu pasažérů lodi Titanic, které byly rozčleněny dle ekonomického statusu, pohlaví, věku a šanci na přežití. Data činí životní osudy 1313 pasažérů, kteří cestovali na Titanicu. Data byla získány na stránce: <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/Titanic.html>. Čtyřdimenzionální pole obsahuje následující proměnné a jejich úrovně:

Třída:	první, druhá, třetí, posádka
Pohlaví:	muž, žena
Věk:	dítě, dospělý
Přežití:	ano, ne (1,0)

Pro následnou práci s daty byla využita metoda logistické regrese, jenž se zabývá problematikou odhadu pravděpodobnosti závislé proměnné na základě určitých nám známých skutečností, jenž reprezentují proměnné nezávislé, které mohou ovlivnit výskyt jevu (Hosmer & Lemeshow, 2000). Data byla zpracována v programu JASP 0.17.1.

Když interpretujeme poměry šancí na přežití katastrofy, musíme mít na paměti, že hodnota 0 spojitých proměnných a úrovně prvního faktoru kategorických proměnných slouží jako referenční třída pro model – a pravděpodobnosti pro tuto referenční třídu jsou popsány ve výsledcích. V našem případě máme věk jako spojitou proměnnou a palubní třídu cestujících jako kategorickou proměnnou. Naši referenční třídou jsou tedy novorozenci nebo cestující ve věku 0 let, kteří cestovali první třídou. Jejich šance na přežití byly asi 7,6 ku 1, což znamená, že jejich pravděpodobnost přežití byla 7,6krát vyšší než pravděpodobnost, že se utopí. Je rovněž patrné, že byla šance na přežití daleko menší, když člověk cestoval třídou druhou či třetí. U pasažérů třetí třídy byla šance na přežití 10krát menší, než u pasažérů třídy první.

Logistická regrese

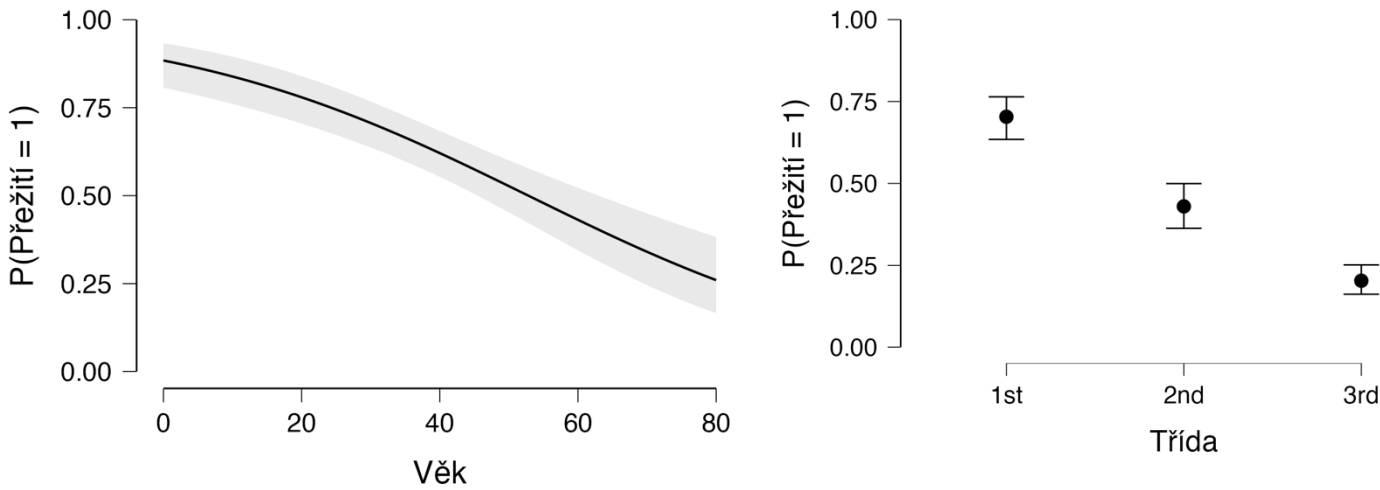
Model Summary – Přežití

Model	Deviance	AIC	BIC	df	X ²	p	McFadden R ²	Nagelkerke R ²	Tjur R ²	Cox & Snell R ²
H ₀	1025.573	1027.573	1032.201	755						
H ₁	909.917	917.917	936.429	752	115.656	< .001	0.113	0.191	0.148	0.142

Coefficients

	Estimate	Standard Error	Odds Ratio	z	Wald Test		
					Wald Statistic	df	p
(Intercept)	2.034	0.307	7.642	6.627	43.918	1	< .001
Věk	−0.039	0.007	0.962	−5.882	34.602	1	< .001
Třída (2nd)	−1.146	0.220	0.318	−5.202	27.062	1	< .001
Třída (3rd)	−2.232	0.229	0.107	−9.745	94.962	1	< .001

Note. Přežití level '1' coded as class 1.



Následující tabulka ukazuje, jak dobře náš model předpovídá výsledky, v našem případě, zda někdo přežil na základě svého věku a třídy.

Performance Diagnostics

Confusion matrix

Observed	Predicted		% Correct
	0	1	
0	350	93	79.007
1	155	158	50.479
Overall % Correct			67.196

Note. The cut-off value is set to 0.5

Performance metrics

	Value
Sensitivity	0.505
Specificity	0.790

Literatura

1. British Board of Trade (1990), *Report on the Loss of the 'Titanic' (S.S.)*. British Board of Trade Inquiry Report (reprint). Gloucester, UK: Allan Sutton Publishing.
2. Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed). John Wiley.