

# ZAVINĚNÍ DOPRAVNÍ NEHODY V ZÁVISLOSTI NA VĚKU, CUT-OFF SKÓRE

Provoz na pozemních komunikacích je každodenní součástí života většiny lidí a počet dopravních nehod neklesá (Policie České republiky, 2024). Řada odborníků se snaží zajistit co nejvyšší bezpečnost, buď zabezpečením automobilů nebo predikováním dopravních nehod a tím snížit jejich počet. Jedním z probíraných faktorů je věk řidiče. Je to proměnná, která není jednoznačná. Podle Šuchy a kol. nejvyšší počet vyhodovaných řidičů je mezi lidmi do 30 let (Šucha, Šrámková, Zámečník, Černochová, & Rehnová, 2016). Zároveň se s přibývajícím věkem zpomalují reakce a snižuje pozornost (Wood, a další, 2024).

Na Moodle máme k dispozici databázi řidičů, která obsahuje data od 2669 osob. Mezi sebraná data patří i údaje o věku řidičů a množství zaviněných nehod. Dalo by se předpokládat, že počet zaviněných nehod bude stoupat s věkem, ale mladší řidiči nemají tolik zkušeností, takže mohou častěji způsobit nehodu. Zajímalo mě, jestli se dá určit hranice, po jejímž překročení můžeme předpokládat zavinění dopravní nehody. Toto cut-off skóre se pokusím stanovit pomocí ROC analýzy.

Pro výpočet jsem použila data 2669 osob z databáze ridici.sta. Soubor je složen z 27 % žen a 63 % mužů, věkové rozmezí bylo 18–69 let a věkový průměr 26,17 roku.

Práce s využitím ROC analýzy vyžaduje několik proměnných, které jsem buď vzala z databáze, nebo jsem si je vypočítala.

## **Převzaté:**

- Věk zkoumané osoby
- Informaci, zda osoba zavinila dopravní nehodu

## **Vypočítané:**

- Prevalence zavinění dopravní nehody. Vzhledem k faktu, že počet osob, které zavinily nehodu se výrazně lišil od počtu těch, kteří nehodu nezavili, nemohla jsem předpokládat prevalenci 50 %. Proto jsem ji vypočítala jako průměr statistiky zavinění dopravní nehody. Průměr jsem si mohla dovolit použít díky reprezentativnosti souboru. Prevalence zavinění dopravní nehody se rovná 21,77 %.

Dále jsem si vypočítala další hodnoty, potřebné pro ROC analýzu:

- True positive – **TP** – tato hodnota představuje počet řidičů, kteří byli zařazeni do věkové skupiny s velkou pravděpodobností způsobení dopravní nehody a nehodu skutečně způsobili
- False positive – **FP** – tato hodnota představuje počet řidičů, kteří byli zařazeni do věkové skupiny s velkou pravděpodobností způsobení dopravní nehody. Kteří ale nehodu skutečně nezpůsobili
- True negative – **TN** – tato hodnota představuje počet řidičů, kteří byli zařazeni do věkové skupiny s malou pravděpodobností způsobení dopravní nehody a nehodu skutečně nezpůsobili
- False negative – **FN** – tato hodnota představuje počet řidičů, kteří byli zařazeni do věkové skupiny s malou pravděpodobností způsobení dopravní nehody, kteří ale nehodu skutečně způsobili
- **Senzitivita** – udává, v kolika procentech případů je náhodně vybraný jedinec při konkrétním cut-off správně zařazen do skupiny pozitivních případů. Tzn. je zařazen do pozitivní skupiny a nehodu skutečně způsobil. Vypočítá se jako  $TP/TP+FN$
- **Specificita** – udává, v kolika procentech případů je náhodně vybraný jedinec při konkrétním cut-off správně označen jako negativní (nezpůsobil nehodu) a opravdu ji nezpůsobil. Vypočítá se jako  $TN/TN+FP$
- **Joudenova statistika** – statistika, která určí výsledné cut-off skóre. Ta funguje dobře tam, kde je vyrovnaný počet skutečných případů (v našem případě zavinění dopravní nehody) a případů, kdy k nehodě nedojde. Spočítá se jako součet senzitivity a specificity, od kterého se odečte číslo 1. Čím vyšší hodnota statistiky, tím lepší.
- **Statistika I'** – V našem případě ale není rozložení zavinění nehody (Ano/Ne) rovnoměrné, pro výpočet cut-off skóre proto použiji statistiku, kterou jsem si pro své účely nazvala I'. Vypočítá se pro každý cut-off skór takto:

$Prevalence\ pozitivních\ případů * senzitivita + prevalence\ negativních\ případů * specificita$

Stejně jako u Joudenovy statistiky čím vyšší hodnota statistiky I', tím přesnější hodnota cut-off skóru.

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=226>.

Následující tabulka 1 uvádí přehled jednotlivých cut-off skóre.

Tabulka 1

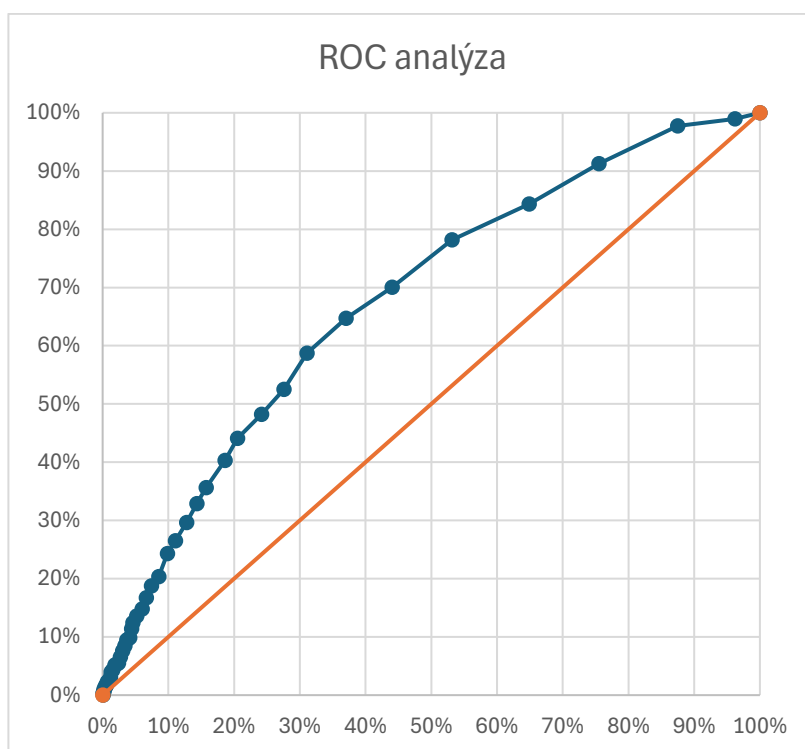
cut-off	TP	FP	TN	FN	senzitivita	specificita	1-spec	Statistika I'
18	581	2088	0	0	100%	0%	100%	21,77%
19	575	2008	80	6	99%	4%	96%	24,54%
20	568	1826	262	13	98%	13%	87%	31,10%
21	530	1576	512	51	91%	25%	75%	39,04%
22	490	1355	733	91	84%	35%	65%	45,82%
23	454	1110	978	127	78%	47%	53%	53,65%
24	407	919	1169	174	70%	56%	44%	59,05%
25	376	773	1315	205	65%	63%	37%	63,36%
26	341	649	1439	240	59%	69%	31%	66,69%
27	305	576	1512	276	52%	72%	28%	68,08%
28	280	504	1584	301	48%	76%	24%	69,84%
29	256	428	1660	325	44%	80%	20%	71,79%
30	234	389	1699	347	40%	81%	19%	72,42%
31	207	329	1759	374	36%	84%	16%	73,66%
32	191	299	1789	390	33%	86%	14%	74,19%
33	172	267	1821	409	30%	87%	13%	74,67%
34	154	231	1857	427	27%	89%	11%	75,35%
35	141	205	1883	440	24%	90%	10%	75,83%
36	118	178	1910	463	20%	91%	9%	75,98%
37	109	155	1933	472	19%	93%	7%	76,51%
38	97	138	1950	484	17%	93%	7%	76,70%
39	86	125	1963	495	15%	94%	6%	76,77%
40	79	108	1980	502	14%	95%	5%	77,14%
41	72	96	1992	509	12%	95%	5%	77,33%
42	66	92	1996	515	11%	96%	4%	77,26%
43	57	85	2003	524	10%	96%	4%	77,18%

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=226>.

44	55	76	2012	526	9%	96%	4%	77,44%
45	49	70	2018	532	8%	97%	3%	77,44%
46	44	63	2025	537	8%	97%	3%	77,52%
47	38	56	2032	543	7%	97%	3%	77,56%
48	32	50	2038	549	6%	98%	2%	77,56%
49	31	45	2043	550	5%	98%	2%	77,71%
50	30	39	2049	551	5%	98%	2%	77,89%
51	25	32	2056	556	4%	98%	2%	77,97%
52	23	27	2061	558	4%	99%	1%	78,08%
53	18	25	2063	563	3%	99%	1%	77,97%
54	15	23	2065	566	3%	99%	1%	77,93%
55	14	18	2070	567	2%	99%	1%	78,08%
56	13	14	2074	568	2%	99%	1%	78,19%
57	10	11	2077	571	2%	99%	1%	78,19%
58	9	10	2078	572	2%	100%	0%	78,19%
59	9	8	2080	572	2%	100%	0%	78,27%
<b>60</b>	<b>6</b>	<b>4</b>	<b>2084</b>	<b>575</b>	<b>1%</b>	<b>100%</b>	<b>0%</b>	<b>78,31%</b>
61	4	4	2084	577	1%	100%	0%	78,23%
62	3	2	2086	578	1%	100%	0%	78,27%
64	2	2	2086	579	0%	100%	0%	78,23%
65	2	1	2087	579	0%	100%	0%	78,27%
69	0	1	2087	581	0%	100%	0%	78,19%

Z tabulky vyplývá, že použitím statistiky I' vyšel jako nejlepší cut-off skór věk 60 let. To znamená, že u šedesátiletého člověka je větší pravděpodobnost zavinění nehody.

Dále jsem pomocí ROC analýzy zobrazila data v grafu č. 1



graf č. 1

V grafu je důležitá plocha pod křivkou, která se označuje AUC (area under the curve). Čím je větší, tzn., křivka je vykreslena co nejvíce vlevo nahoru, tím daná proměnná přesněji predikuje výskyt dané události. Plocha pod křivkou je 0,69, což ukazuje, že by se věk při predikci měl brát v potaz, nicméně není dostačující. Oranžová přímka je pouze ilustrativní.

Pro srovnání, jak je důležité zohlednit prevalenci výskytu události, uvádím výsledek Joudenovy statistiky, která jako cut-off skór určila věk 25 let, což je opravdu velký rozdíl oproti statistice I'.

## Bibliografie

- Policie České republiky. (2024). *Statistika nehodovosti*. Získáno 2024-03-12, z Policie České republiky: <https://www.policie.cz/clanek/statistika-nehodovosti-900835.aspx>
- Šucha, M., Šrámková, L., Zámečník, P., Černochová, D., & Rehnová, V. (2016). ŘIDIČI PODSTUPUJÍCÍ DOPRAVNĚPSYCHOLOGICKÉ VYŠETŘENÍ V ČR. *Psychologie Pro Praxi*(1/2), 47-62. Získáno 2024-03-12, z <https://eds.p.ebscohost.com/eds/pdfviewer/pdfviewer?vid=1&sid=0f1c0605-6bff-4f67-a222-25cff883b964%40redis>
- Wood, J., Henry, E., Black, A., Kaye, S.-A., Glaser, S., Rakotonirainy, A., & Anstey, K. (2024). Exploring perceptions of Advanced Driver Assistance Systems (ADAS) in older drivers with age-related declines. *Transportation Research Part F: Traffic Psychology and Behaviour*, 100, 419 - 430. doi:10.1016/j.trf.2023.12.006