

# Determinanty cen ojetých vozů Porsche 911

Adam Kelbl<sup>1</sup>

5. ledna 2025

**Abstrakt:** Cílem této seminární práce je pomocí regresní analýzy určit klíčové determinanty cen ojetých vozů Porsche 911. Podobné problematice se věnoval například Z. Sun (2024), který jako klíčové determinanty cen ojetých vozů v USA uvádí rok výroby vozu a jeho kilometrový nájezd, naopak barva exteriéru vozu na cenu dle jeho závěrů nemá vliv. Na datech z internetového autobazaru byla provedena vícenásobná lineární regrese pomocí metody nejmenších čtverců (OLS). Z důvodu přítomnosti heteroskedasticity v modelu byly pro výsledný model použity robustní standardní chyby. Analýza potvrzuje významný vliv kilometrového nájezdu a zanedbatelný vliv barvy vozu, přičemž výsledky rovněž naznačují, že série vozu a specifikace jako Turbo, GT, nebo RS cenu výrazně zvyšují. Práce poskytuje zajímavý vhled do cenotvorby a faktorů ovlivňujících hodnotu modelu Porsche 911 na trhu.

**Klíčová slova:** Cena ojetého vozu; Determinanty ceny; Porsche 911; Lineární regrese; OLS

**JEL klasifikace:** C21, C51, L62

## 1 Úvod

Automobilka Porsche je synonymem pro luxusní sportovní vozy, přičemž nejznámější model 911 je často označován za kultovní vůz a ztělesnění samotné značky. Díky své dlouhé historii, ikonické pověsti a široké škále specifikací představuje Porsche 911 zajímavý objekt zkoumání, zejména z hlediska faktorů ovlivňujících jeho cenu. Cílem této práce je identifikovat klíčové faktory určující ceny ojetých vozů tohoto modelu.

Pro analýzu byla využita vícenásobná lineární regrese pomocí metody nejmenších čtverců (OLS), která umožňuje modelovat vztah mezi cenou ojetých vozů a různými faktory, jako jsou kilometrový nájezd, série, specifikace a další proměnné. Podrobnější popis proměnných a jejich charakteristiky jsou uvedeny v sekci „Data“.

Problematika cen ojetých vozů a jejich predikce je poměrně dobře prozkoumané téma, kterému se věnovala řada odborných prací. Například studie Z. Suna (2024) identifikovala jako klíčové faktory ovlivňující cenu vozu rok výroby a jeho kilometrový nájezd, zatímco jiné studie zdůraznily typ převodovky (Pukeri & Safitri, 2020), nebo velikost motoru (Ceyhun, 2018).

Na úvod stanovme dvě hypotézy, které budou v rámci modelu testovány. První hypotéza zní: „Barva exteriéru vozu má statisticky významný vliv na jeho cenu, přičemž výrazné a neobvyklé barvy cenu zvyšují.“ Druhá hypotéza se zaměřuje na vliv typu převodovky na cenu vozidla. V případě modelů Porsche 911 je obecně přijímáno, že vozidla s manuální převodovkou mají

---

<sup>1</sup>Masarykova univerzita, Ekonomicko-správní fakulta, obor: Finance, 541344@mail.muni.cz

vyšší hodnotu. Formulujeme tedy druhou hypotézu následovně: „Vozidla vybavená automatickou převodovkou mají v průměru nižší cenu než vozidla s manuální převodovkou.“

Potřebná data byla získána z internetového autobazaru [Elferspot.com](#), který se specializuje výhradně na vozy značky Porsche. Pro sběr dat a jejich následnou ekonometrickou analýzu byl využit programovací jazyk Python, především knihovny Statsmodels, Pandas a BeautifulSoup. Data byla před samotnou analýzou upravena do vhodné podoby pomocí Excelu. Celý proces tvorby modelu je dostupný v GitHub repozitáři [zde](#).

## 2 Data

Veškerá data pocházejí z internetového autobazaru [Elferspot.com](#). Po procesu čištění dat je k dispozici 1473 pozorování. Dataset obsahuje 24 proměnných, které popisují cenu, technické specifikace a další charakteristiky jednotlivých vozů. Přehled všech proměnných a jejich popis je uveden v tabulce v příloze (Tabulka 2). Pro upřesnění následuje podrobnější vysvětlení některých z nich.

Umělé proměnné *IS964*, *IS993* atd., nabývají pro jednotlivé série modelu 911 hodnoty 1. Dále jsou v datasetu umělé proměnné, které nesou informace o specifikaci vozu, například *S*, pro sportovnější verze modelu, nebo *Turbo*, pro auta s přeplňovaným motorem. U aut se specifikací „GT“ (GT3/GT2) nabývá stejnojmenná umělá proměnná hodnoty 1 a značí vysoký výkon a sportovnější specifikaci. Pokud se jedná o vůz s přídomkem „RS“, tedy auta speciálně určená na závodní okruh (německy „Rennsport“), umělá proměnná *RS* nabývá hodnoty 1. Umělá proměnná *Special* nabývá hodnoty 1, pokud se jedná o speciální, nejčastěji jubilejní, edici vozu. Proměnné *Cabriolet* a *Targa* nabývají hodnoty 1, pokud se jedná o daný typ karoserie. Proměnná *PDK* nabývá hodnoty 1, pokud je vůz vybaven technologií dvouspojkové automatické převodovky („Porsche Doppelkupplung“). Proměnné *ExtCol* a *IntColor* vyjadřují barvu exteriéru a interiéru vozu, konkrétně jestli se jedná o výraznou a neobyčejnou barvu (červená, žlutá, zelená, ...), či barvu obyčejnou (černá, šedá, bílá, ...).

Na závěr je vhodné zmínit, že z důvodu dokonalé multikolinearity některé specifikace nemají vlastní proměnnou, ale jsou implicitně zahrnuty, pokud se jiné proměnné rovnají 0. Například, pokud jsou proměnné *Turbo* a *GT* rovny 0, znamená to, že vůz má základní specifikaci Carrera. Pokud jsou proměnné *Cabriolet* a *Targa* rovny 0, znamená to, že vůz má karoserii Coupé. Pokud jsou všechny umělé proměnné vyjadřující sérii vozu rovny 0, znamená to, že se jedná o vůz nejnovější série 992.

## 3 Ekonometrický model

Tato sekce se zaměřuje na klíčovou část empirického výzkumu, a to na samotný ekonometrický model. Představená data byla analyzována pomocí vícenásobné lineární regrese provedené metodou nejmenších čtverců (OLS). Tento přístup umožňuje zkoumat vliv jednotlivých proměnných na výslednou cenu, ověřovat stanovené hypotézy a formulovat relevantní závěry. Součástí je také ověření klasických předpokladů a testování na přítomnost multikolinearity pro zajištění přesnosti odhadnutého modelu.

### 3.1 Odhadnutý model

V prvním modelu byly použity všechny dostupné proměnné. Pro model byla zvolena funkční forma lin-lin, která předpokládá čistě lineární vztah mezi nezávislými proměnnými a závislou proměnnou. Shrnutí klíčových statistických ukazatelů tohoto modelu je uvedeno v tabulce v příloze (Tabulka 3). Na základě p-hodnot jednotlivých proměnných je posouzena jejich statistická významnost a proměnné, jejichž p-hodnota přesahuje hladinu významnosti 10 %, budou z modelu odstraněny, aby se předešlo zahrnutí irelevantních proměnných.

Jedním z klíčových testů o vhodné specifikaci modelu je tzv. Ramseyův RESET test, který pomocí přidání nelineárních vztahů do modelu testuje jeho adekvátní specifikaci. Pro první model, a to jak včetně, tak bez statisticky nevýznamných proměnných, vykazuje p-hodnota RESET testu velmi nízké hodnoty blízké 0. To naznačuje neadekvátní specifikaci modelu, pravděpodobně způsobenou opomenutím nelineárních vztahů nebo interakčních členů. Ani po přidání druhých či třetích mocnin spojených vysvětlujících proměnných, nebo jejich zlogaritmování, RESET test modelu nedosahuje uspokojivé p-hodnoty, tedy alespoň hodnoty vyšší než 1 % hladina významnosti. Zároveň byl do modelu přidán interakční člen *MileagexYear*, který udává, jak se vliv nájezdu na cenu mění v závislosti na roce výroby vozu. Ani ten ale nezvyšuje p-hodnotu RESET testu.

Jako nejlepší se jeví model ve funkční podobě log-lin, bez zmíněného interakčního členu či druhých mocnin vysvětlujících proměnných. Shrnutí tohoto modelu je uvedeno v tabulce v příloze (Tabulka 4). P-hodnota RESET testu modelu je rovna 0.0002 a koeficient determinace je roven 0.856. Stále tak zamítáme hypotézu o adekvátní specifikaci modelu, avšak koeficient determinace nám říká, že model vysvětluje přibližně 85,6 % variability závislé proměnné na základě zahrnutých vysvětlujících proměnných, což je ze všech předchozích modelů nejvyšší hodnota.

### 3.2 Klasické předpoklady

Abychom si mohli být jistí, že odhady získané naším modelem jsou podle Gauss-Markovovy věty nestranné, konzistentní a efektivní, je potřeba tento model testovat na splnění tzv. klasických předpokladů, které zní následovně:

1. Nulová střední hodnota reziduí
2. Konstantní rozptyl reziduí (Homoskedasticita)
3. Vzájemná nekorelovanost jednotlivých reziduí (Absence autokorelace)
4. Normalita reziduí
5. Nezávislost vysvětlujících proměnných a reziduí.

Jelikož předpokládáme, že jsme zvolili správné vysvětlující proměnné a zároveň máme v modelu zahrnutou konstantu, náš model splňuje první předpoklad o nulové střední hodnotě reziduí.

Pro testování, zda mají rezidua modelu konstantní rozptyl, byl zvolen Breusch-Paganův test heteroskedasticity. Výsledek testu, resp. p-hodnota testu (viz Tabulka 1 níže), značí přítomnost heteroskedasticity, tedy nekonstantního rozptylu reziduí. To potvrzuje i vizualizace vztahu reziduí a predikovaných hodnot, která má mírně kuželovitý tvar (Obrázek 1 v příloze). Z tohoto důvodu pro model použijeme robustní odhad směrodatných odchylek, který umožňuje získat správné směrodatné odchylky a testovací statistiky i v přítomnosti heteroskedasticity. Tím

zajistíme, že interpretace výsledků bude méně ovlivněna porušením tohoto předpokladu.

Vzhledem k tomu, že zkoumaný model je založen na průřezových datech, která představují nezávislé pozorování v jednom časovém okamžiku, není nutné provádět test na přítomnost autokorelace. Autokorelace je obvykle relevantní pouze při práci s časovými řadami. U našeho modelu tak lze předpokládat, že problém autokorelace nevzniká a předpoklad o nezávislosti jednotlivých pozorování je splněn.

K ověření předpokladu normálního rozdělení reziduí byl použit Jarque-Berův test. Výsledky testu indikují, že rezidua modelu nemají normální rozdělení (viz Tabulka 1 níže). Nicméně při dostatečně velkém počtu pozorování lze využít asymptotickou teorii, podle níž se rozdělení reziduí postupně přibližuje normálnímu rozdělení, a to i v případech, kdy rezidua původně normální rozložení nevykazují. V případě našeho modelu s 1473 pozorováními tak můžeme předpoklad normality považovat za splněný.

Posledním klasickým předpokladem je nezávislost (nekorelovanost) vysvětlujících proměnných a reziduí. Tento předpoklad byl ověřen pomocí korelačních koeficientů mezi vysvětlujícími proměnnými a rezidui modelu. Jelikož všechny vypočtené korelační koeficienty dosahují hodnot velmi blízkých nule, lze předpokládat, že model splňuje podmínku exogenity proměnných. Výsledky korelačních koeficientů jsou uvedeny v příloze (Tabulka 5).

Tabulka 1: Testy klasických předpokladů

Test	Výsledek	Důsledek
Breusch-Pagan (heteroskedasticita)	$5.4 \times 10^{-40}$	Heteroskedasticita
Jarque-Bera (normalita reziduí)	$8.2 \times 10^{-87}$	Nenormální rozdělení
Korelace mezi regresory a rezidui	Nízké korelační koeficienty	Nezávislost $X$ a $\varepsilon$

### 3.3 Multikolinearita

Před odhadem finálního modelu je potřeba prozkoumat vzájemné korelace mezi vysvětlujícími proměnnými. Vysoká korelace mezi některými proměnnými by naznačovala, že tyto proměnné nesou stejnou či podobnou informaci, což by mohlo vést k problému multikolinearity v modelu. Na grafické korelační matici uvedené v příloze (Obrázek 2) lze vidět, že nejvyšší korelační koeficienty se pohybují kolem hodnoty 0.5, což naznačuje, že se multikolinearita v modelu nevyskytuje.

### 3.4 Finální model

Pro finální model byl z důvodu přítomné heteroskedasticity zvolen robustní odhad směrodatných odchylek typu HC1. Ten sice neeliminuje heteroskedasticitu v modelu, ale zpřesní směrodatné odchylky koeficientů, testovací statistiky a intervaly spolehlivosti. Detailní shrnutí tohoto modelu se nachází v příloze (Tabulka 6). Tento model dosahuje koeficientu determinace 0.857, což znamená, že vysvětluje přibližně 85,6 % variability závislé proměnné.

Jedná se o model v podobě log-lin, kdy vysvětlovaná proměnná je v logaritmu a vysvětlující proměnné v lineární podobě. Význam koeficientů lze tedy vysvětlit následovně: „Pokud se hodnota některé vysvětlující proměnné zvýší o jednotku (například pokud nájezd vozu vzroste

o 1 km), změni se hodnota závislé proměnné (v tomto případě cena vozu) o procentní hodnotu odpovídající danému koeficientu.“

## 4 Interpretace výsledků

Nyní je vhodné interpretovat výsledky modelu, především jeho koeficienty, a ověřit dříve stanovené hypotézy. Shrnutí modelu je dostupné v příloze (Tabulka 6). Konstanta modelu nám obecně říká, jaká je hodnota vysvětlované proměnné, pokud jsou všechny vysvětlující proměnné rovny 0. V případě našeho modelu se tedy jedná o cenu Porsche 992 Carrera s manuální převodovkou s nulovým nájezdem a teoreticky „s žádným výkonem a objemem motoru“. V případě tohoto modelu je tedy konstanta spíše teoretická a vůz by měl hodnotu  $e^{11.4855}$ , tedy zhruba 97295 EUR. Zajímavější jsou ale koeficienty jednotlivých proměnných. Například série 964 zvýší cenu oproti sérii 992 o 48,63 %. Naopak série 996 je v průměru o 42,2 % levnější než série 992. Specifikace „RS“ v průměru zvyšuje cenu vozu o 52 %.

Nyní otestujeme hypotézy stanovené v úvodu. První hypotéza zněla: „Barva exteriéru vozu má statisticky významný vliv na jeho cenu, přičemž výrazné a neobvyklé barvy cenu zvyšují.“ Proměnná *ExtCol* byla v procesu tvorby modelu odstraněna z důvodu její statistické nevýznamnosti. Proto můžeme zamítnout hypotézu, že barva vozu má statisticky významný vliv na jeho cenu.

Druhá hypotéza zněla: „Vozidla vybavená automatickou převodovkou mají v průměru nižší cenu než vozidla s manuální převodovkou.“ Jelikož je ve výsledném modelu koeficient pro proměnnou *Automat* negativní a p-hodnota velmi nízká, znamená to, že vozidla s automatickou převodovkou mají v průměru nižší cenu než vozidla s manuální převodovkou, což potvrzuje stanovenou hypotézu.

## 5 Závěr

Hlavním cílem této práce bylo pomocí regresní analýzy určit klíčové determinanty cen ojetých vozů Porsche 911. Domnívám se, že tento cíl se mi v rámci mých znalostí ekonometrie a dostupných dat podařilo splnit. Na datech z internetového autobazaru byla provedena vícenásobná lineární regrese metodou OLS. Výsledný model ve funkční podobě log-lin je schopný vysvětlit necelých 86 % variability vysvětlované proměnné. To považuji za poměrně dobrý výsledek. Jedním z úskalí tohoto modelu je nevyhovující výsledek Ramseyho RESET testu, který naznačuje neadekvátní specifikaci modelu.

Práce poskytuje zajímavý vhled do faktorů ovlivňujících hodnotu modelu Porsche 911 na trhu. Model nám říká, jak určité specifikace ovlivňují cenu tohoto modelu. Například série 964 zvýší cenu auta oproti sérii 992 o 48,63 %. Auta s karoserií typu Targa (tzn. auta s částečně odnímatelnou střechou), mají v průměru o 7 % vyšší hodnotu, než auta s karoserií Coupé. Jako statisticky nevýznamné se ukázaly faktory jako pohon všech kol, technologie PDK (dvouspojková automatická převodovka), nebo barva exteriéru a interiéru.

Pro přesnější výsledky výzkumu faktorů ovlivňujících cenu Porsche 911, by bylo vhodné zahrnout více proměnných, jako je například počet předchozích majitelů nebo historie nehod. Dále je třeba brát v potaz, že data o cenách nejsou skutečné uzavřené prodejní ceny, ale pouze nabízené ceny vozů ze strany prodávajících, což může způsobovat zkreslení cen směrem nahoru.

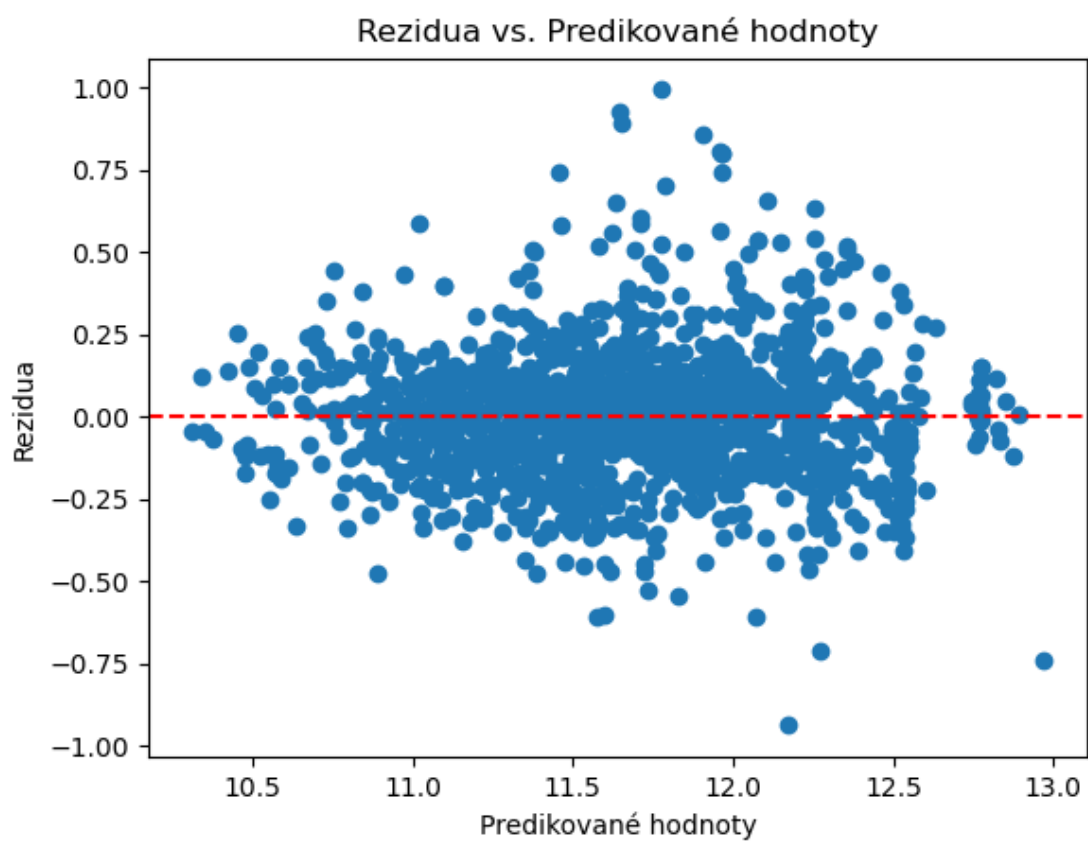
## Reference

- [1] Sun, Z. (2024). Research on factors affecting second-hand car market prices. *Theoretical and Natural Science*, 36, 128-135.
- [2] Puteri, C. K. & Safitri, L. N. (2020). Analysis of linear regression on used car sales in Indonesia. *Journal of Physics: Conference Series*, 1469.
- [3] Ceyhun, O. (2018). Multiple Linear Regression Applications Automobile Pricing. *Real Estate Management and Valuation*, 15.

## A Přílohy

Tabulka 2: Seznam použitých proměnných a jejich popis

Proměnná	Popis
Price	Cena vozu (v EUR)
Year	Rok výroby vozu
IS911	Umělá proměnná: 1 pro sérii 911 (1964–1989), jinak 0
IS964	Umělá proměnná: 1 pro sérii 964 (1989–1994), jinak 0
IS993	Umělá proměnná: 1 pro sérii 993 (1995–1998), jinak 0
IS996	Umělá proměnná: 1 pro sérii 996 (1999–2005), jinak 0
IS997	Umělá proměnná: 1 pro sérii 997 (2005–2012), jinak 0
IS991	Umělá proměnná: 1 pro sérii 991 (2012–2019), jinak 0
IS992	Umělá proměnná: 1 pro sérii 992 (2019–současnost), jinak 0
Special	Umělá proměnná: 1 pro speciální edice (Exkluzivní edice, Clubsport, ...), jinak 0
S	Umělá proměnná: 1 pro verzi S (Super/Sport), jinak 0
RS	Umělá proměnná: 1 pro verzi RS (Rennsport), jinak 0
Turbo	Umělá proměnná: 1 pro verzi Turbo, jinak 0
GT	Umělá proměnná: 1 pro verzi GT (Gran Turismo), jinak 0
Targa	Umělá proměnná: 1 pro verzi Targa (částečně odnímatelná střecha), jinak 0
Cabriolet	Umělá proměnná: 1 pro kabriolet, jinak 0
Automat	Umělá proměnná: 1 pro automatickou převodovku, jinak 0
PDK	Umělá proměnná: 1 pro převodovku PDK, jinak 0
AWD	Umělá proměnná: 1 pro pohon všech kol, jinak 0
ExtCol	Umělá proměnná: 1 pokud má auto výraznou barvu exteriéru, jinak 0
IntColor	Umělá proměnná: 1 pokud má auto výraznou barvu interiéru, jinak 0
Mileage	Nájezd vozu (v kilometrech)
Power	Výkon vozu (v koních)
CylCap	Velikost motoru (v litrech)



Obrázek 1: Graf vztahu reziduů a predikovaných hodnot



Tabulka 3: Model 1

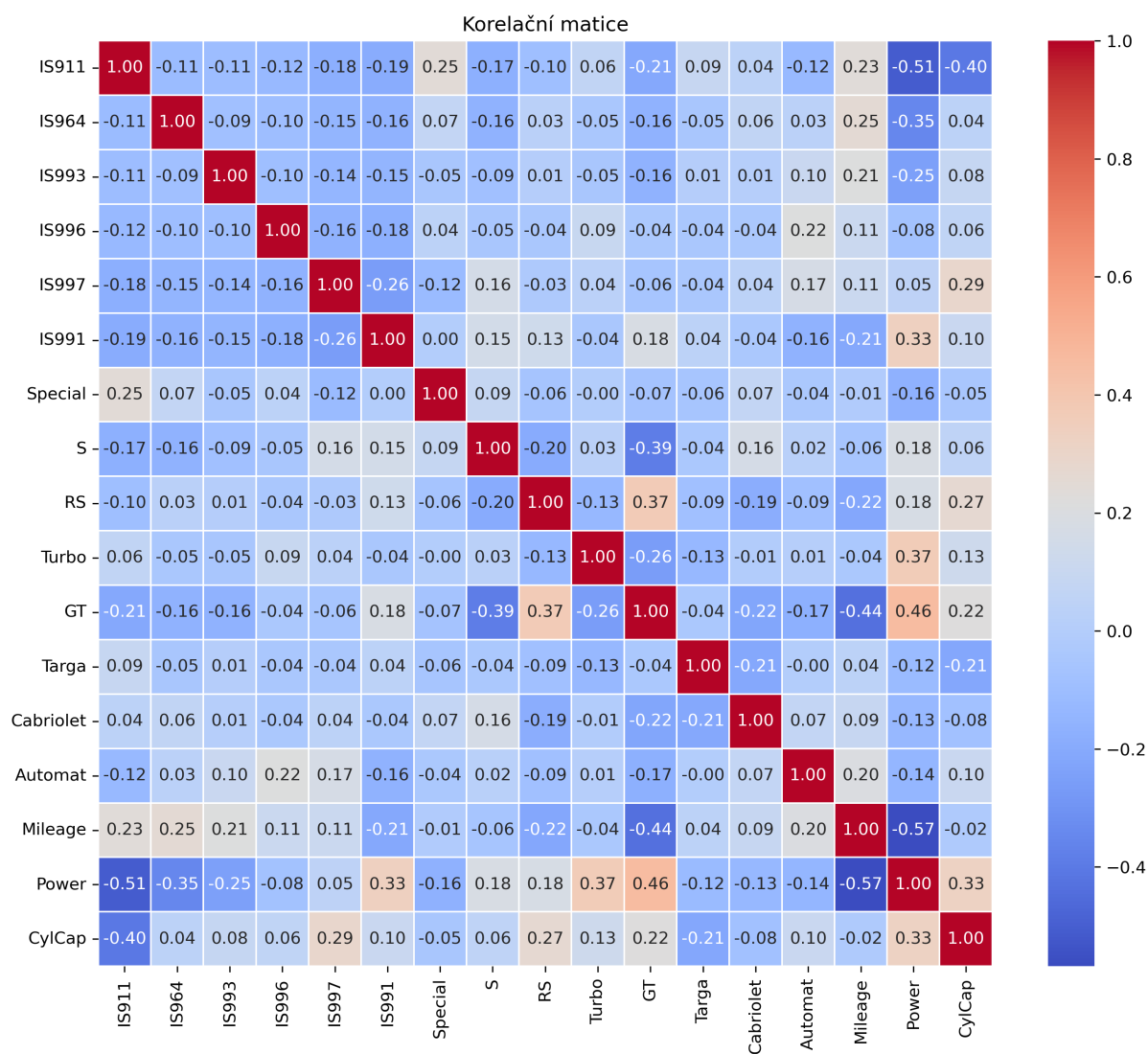
<b>Proměnná</b>	<b>Koeficient</b>	<b>P-hodnota</b>
const	20120	0.193
IS911	68060	0.000
IS964	64020	0.000
IS993	52710	0.000
IS996	-40100	0.000
IS997	-41050	0.000
IS991	-37220	0.000
Special	50440	0.000
S	8339.2	0.003
RS	103900	0.000
Turbo	4059.7	0.428
GT	17760	0.000
Targa	11610	0.001
Cabriolet	-9568	0.000
Automat	-14680	0.000
PDK	-6741.2	0.043
AWD	-3806.2	0.116
IntColor	-3506.3	0.236
ExtCol	-94.9	0.963
Mileage	-0.389	0.000
Power	356.98	0.000
CylCap	-931.05	0.798
Koeficient determinace:		0.782
Adj. koeficient determinace:		0.779
RESET test (p-hodnota):		0.000

Tabulka 4: Model 2

Proměnná	Koeficient	Směrodatná odchylka	P-hodnota
const	11.4855	0.087	0.000
IS911	0.4178	0.046	0.000
IS964	0.4863	0.043	0.000
IS993	0.4223	0.039	0.000
IS996	-0.4288	0.032	0.000
IS997	-0.3016	0.025	0.000
IS991	-0.1936	0.018	0.000
Special	0.3401	0.024	0.000
S	0.1192	0.016	0.000
RS	0.5208	0.024	0.000
Turbo	0.2188	0.029	0.000
GT	0.3128	0.024	0.000
Targa	0.0727	0.020	0.000
Cabriolet	-0.0703	0.012	0.000
Automat	-0.1830	0.019	0.000
Power	0.0017	0.000	0.000
Mileage	$-3.379 \times 10^{-6}$	$1.24 \times 10^{-7}$	0.000
CylCap	-0.1076	0.021	0.000
<b>Koeficient determinace:</b>			0.857
<b>Adj. koeficient determinace:</b>			0.856
<b>RESET test (p-hodnota):</b>			0.0002

Tabulka 5: Korelace vysvětlujících proměnných a reziduí

Proměnná	Korelace
IS911	$1.46 \times 10^{-11}$
IS964	$1.96 \times 10^{-11}$
IS993	$1.63 \times 10^{-11}$
IS996	$1.03 \times 10^{-11}$
IS997	$5.98 \times 10^{-12}$
IS991	$-1.55 \times 10^{-11}$
Special	$-9.25 \times 10^{-13}$
S	$-3.38 \times 10^{-12}$
RS	$-1.78 \times 10^{-11}$
Turbo	$-4.16 \times 10^{-12}$
GT	$-3.47 \times 10^{-11}$
Targa	$3.20 \times 10^{-12}$
Cabriolet	$7.20 \times 10^{-12}$
Automat	$1.54 \times 10^{-11}$
Power	$-4.23 \times 10^{-11}$
Mileage	$8.34 \times 10^{-11}$
CylCap	$-1.42 \times 10^{-12}$



Obrázek 2: Korelační matice proměnných

Tabulka 6: Finální model

Proměnná	Koeficient	Směrodatná chyba	P-hodnota
const	11.4855	0.100	0.000
IS911	0.4178	0.067	0.000
IS964	0.4863	0.058	0.000
IS993	0.4223	0.051	0.000
IS996	-0.4288	0.044	0.000
IS997	-0.3016	0.032	0.000
IS991	-0.1936	0.017	0.000
Special	0.3401	0.032	0.000
S	0.1192	0.015	0.000
RS	0.5208	0.031	0.000
Turbo	0.2188	0.042	0.000
GT	0.3128	0.028	0.000
Targa	0.0727	0.021	0.000
Cabriolet	-0.0703	0.011	0.000
Automat	-0.1830	0.016	0.000
Power	0.0017	0.000	0.000
Mileage	$-3.379 \times 10^{-6}$	$1.58 \times 10^{-7}$	0.000
CylCap	-0.1076	0.020	0.000
<b>Koeficient determinace:</b>			0.857
<b>Adj. koeficient determinace:</b>			0.856
<b>Poznámka:</b> Směrodatné chyby jsou robustní vůči heteroskedasticitě (HC1).			