

# Zadání domácího úkolu ze 4st416 (za max. 10 bodů)

## 1 Data

- Pracujte s následujícím datovým souborem: <http://www1.aucegypt.edu/faculty/hadi/RABE5/Data5/P155b.txt> (Zdroj: Chatterjee & Hadi, Regression Analysis by Example, 5th Edition).
- K dispozici jsou údaje o automobilech různých značek (Car). Pro každou značku je uvedena:
  - **cena** (Y),
  - **země původu** (Country): kategoriální proměnná se čtyřmi kategoriemi (USA, Japan, Germany, Other),
  - **koňská síla automobilu** (Horsepower); pro význam koňské síly automobilu viz např. <http://en.wikipedia.org/wiki/Horsepower>.

## 2 Úkol

- S využitím regresní analýzy se pokuste zodpovědět na tyto dvě otázky:
  1. Je vztah mezi vysvětlovanou proměnnou (Y) a vysvětlujícími proměnnými (Horsepower, Country) aditivní nebo neaditivní? Pokud je neaditivní, pokuste se popsat charakter této neaditivity.
  2. Jak *prakticky* významná je – po *očistění* o koňskou sílu (Horsepower) – závislost ceny automobilu (Y) na zemi původu (Country)? (Pojmem “očistění o koňskou sílu” máme na mysli to, že nás zajímá závislost ceny automobilu na zemi původu při konstantní koňské síle.)

## 3 Doplnující poznámky

- Cílem úkolu *není mechanické* testování hypotéz apod. Cílem je, abyste se snažili *co nejlépe porozumět datům* a pokusili se co nejlépe zodpovědět na otázky položené výše. K nalezení odpovědí můžete stejně dobře a možná i lépe než testy hypotéz použít mimo jiné např. jen “*popis*” závislostí, jejich *vizualizaci* apod. Můžete samozřejmě použít i testy hypotézy, pokud to uznáte za vhodné a oprávněné.
- Přesný postup vypracování a zodpovězení otázek tohoto domácího úkolu nechávám na Vás. Budu rád za Vaši vlastní invenci a nápad.:-) Nechci proto předkládat žádné další návody, jak postupovat při zodpovězení otázek. Více ocením skutečné porozumění datům než pouhé “mechanické” testování hypotéz (které třeba ani nemusí být v některých modelech a kontextech oprávněné z důvodu nesplnění potřebných předpokladů).

## 4 Forma vypracování domácího úkolu

- V kontextu domácího úkol jsou důležité *především* odpovědi na otázky položené výše. Další jiné věci, které byste v kontextu regresní analýzy mohli potenciálně zjistit (pokud tedy tyto věci nesouvisí úzce s položenými otázkami), do vypracování úkolu proto prosím neuvádějte.
- Text proložte výstupy (tabulkami) a obrázky (ale prosím jen těmi relevantními v kontextu daných otázek). Doplňte veškeré komentáře, úvahy a argumenty, abyste podpořili správnost Vašich odpovědí. Pokuste se doplnit i *věcné* vysvětlení pro vztahy, které jste v kontextu hledání odpovědí na položené otázky našli. Vypracování úkolu by mělo být *minimálně* na jednu standardní stránku A4, *maximálně* na stránky čtyři (tím mám na mysli text + obrázky + tabulky/výstupy, počítáno *bez* kódu).

## 5 Další pokyny

- Maximální počet bodů, které za úkol můžete obdržet, je 10.
- Úkol odevzdejte do odevzdávárny v ISISu. Pro deadline odevzdání úkolu viz odevzdávárna v ISISu.
- Vlastní vypracovaný domácí úkol odevzdejte ve formátu *.doc(x)* (Word) nebo *.odt* (OpenOffice Writer apod.) nebo jako *pdf* apod. Do odevzdávárny též prosím odevzdejte **Rkovský skript**. Tento skript by měl obsahovat jen příkazy a kód, jež je relevantní v kontextu položených otázek. Kód je nutné okomentovat.
- K vypracování domácího úkolu můžete použít libovolné vědomostní zdroje (včetně konzultací se mnou, s kolegy apod.), avšak úkol ve finále musíte vypracovat *sami*. Za úkol, na kterém je vidět velký kus odvedené práce, můžete získat vysoký počet bodů, i kdybyste nedospěli ke zcela správným výsledkům.
- **Důležité upozornění:** Vše, co odevzdáte, musíte být schopni obhájit, pokud bych Vás k tomu vyzval. Tj. měli byste být schopni (v případné diskuzi se mnou) vysvětlit vše, co jste v domácím úkolu udělali a napsali a případně Váš postup analýzy zreprodukovat/zopakovat. Pokud byste nebyli schopni Váš domácí úkol v případě potřeby obhájit, výsledkem by bylo 0 bodů za tento domácí úkol.