Cvičení z NSTP129

Základy lineární regrese

12. 1. 2012

Úvodní nastavení. Budeme pracovat s daty `ojetiny.dat` (stáhněte si je z internetu). Spusťte si `R Commander` (`Packages`→`Load Package`→`Rcmdr`), změňte si svůj pracovní adresář (`File`→`Change working directory`) a načtěte si do `R` data `ojetiny.dat` (`Data`→`Import Data`→`from text file, clipboard or URL`).

Popis dat: Soubor obsahuje data o cenách náhodně vybraných ojetých automobilů. Pro každý automobil máme k dispozici informaci o ceně, počtu najetých kilometrů a stáří. Proměnné v datech:

`cena`	cena v tis. Kč,
`stari`	stáří automobilu (v měsících)
`najeto`	počet najetých km (v tis.)

1. Pomocí vhodných charakteristik a obrázků popište zkoumaná data.
2. Popište vztah mezi cenou automobilu a počtem najetých kilometrů pomocí vhodného obrázku. Interpretujte tvar závislosti. Spočtěte také vhodnou číselnou charakteristiku.
3. Popište vztah mezi cenou automobilu a jeho stářím pomocí vhodného obrázku. Interpretujte tvar závislosti. Spočtěte také vhodnou číselnou charakteristiku.
4. Uvažujte model lineární závislosti ceny vozu na počtu najetých kilometrů, tj. model regresní přímky.
   (a) Napište rovnici modelu a vysvětlete jej.
   (b) Odhadněte parametry uvažovaného modelu.

   `Statistics`→`Fit models` →`Linear regression`. Zde vhodně zvolte závislou a nezávislou proměnnou.
   (c) Prohlédněte si dobře výstup z `R`. Napište rovnici odhadnutého modelu. Diskutujte statistickou významnost jednotlivých parametrů. Interpretujte.
   (d) Jaké je koeficient determinace R^2 tohoto modelu? Co nám tato hodnota vyjadřuje?
   (e) Jaké jsou předpoklady lineární regrese? Jsou v našem případě splněny (ověřte graficky)? `Models` →`Graphs` →`Basic diagnostic plots`

   Je lineární závislost (přímka) dostačující pro popsání studovaného vztahu?
5. Rozšiřme předchozí model o kvadratický člen, tj. uvažujme kvadratickou závislost ceny na počtu najetých kilometrů.
   (a) Napište rovnici modelu.

(b) Odhadněte parametry tohoto modelu.

`Statistics→Fit models →Linear model`. Zde zadejte model
`cena~najeto + I(najeto^2)`

(c) Napište rovnici odhadnutého modelu.

(d) Podívejte se na statistickou významnost jednotlivých parametrů. Speciálně se zaměřte na koeficient kvadratického členu. Interpretujte.

(e) Jak se změnilo R^2? Dala se tato změna očekávat?

(f) Posuďte splnění předpokladů pomocí vhodných obrázků.

(g) Který z uvažovaných dvou modelů je „lepší“ a proč?
Odhadněte na základě toho lepšího typickou cenu automobilu s 30 000 najetými km.

6. Uvažujme navíc v modelu i lineární závislost ceny na stáří vozu.

(a) Napište rovnici modelu.

(b) Opět odhadněte koeficienty v tomto modelu.

`Statistics→Fit models →Linear model`. Zde zadejte model
`cena~najeto + I(najeto^2)+stari`.

(c) Závisí cena automobilu na jeho stáří (uvažujeme-li kvadratickou závislost na najetých km)?

(d) Na základě uvažovaného modelu odhadněte cenu 1 rok starého automobilu s 30tis najetými km.