

Úkol 1 – Explorace dat a lineární regrese – Point data

Cíl

Cílem prvního úkolu je seznámit se s daty z Point studie a modelovat koncentraci kreatininu v moči pacientů jako funkci hmotnosti pacienta a dalších proměnných.

Data

Point data byla sesbírána v průběhu několika let na více než 2000 pacientech podstupujících náročné lékařské zákroky. Cílem studie bylo posoudit výhody a nevýhody přesné kontroly glykémie pacienta (koncentrace glukózy v krvi) během operace. Velmi nízká koncentrace glukózy (hypoglykémie) stejně jako velmi vysoká koncentrace glukózy (hyperglykémie) mohou být pro pacienty velmi nebezpečné. V rámci studie byly použity dva protokoly při provádění operace, v prvním případě byla hladina glukózy velmi přesně regulována, zatímco v druhém případě nebyla hladina glukózy kontrolována, ale pouze měřena. Z důvodu posouzení výhod byly pacienti sledováni při pobytu na anesteziologicko - resuscitačním oddělení a veškeré pooperační komplikace byly zaznamenány. Pro účely prvního úkolu budete používat podmnožinu původních dat.

Explorace dat

Data z point studie obsahují 29 příznaků, jejichž stručný popis je v tabulce níže. V rámci explorace dat prozkoumejte všechny příznaky, které máte k dispozici a seznamte s typem dat v jednotlivých příznacích, rozsahem jejich hodnot, jejich rozdělením, případně počtem chybějících a odlehlých hodnot. Zvažte, jestli některé z příznaků není potřeba transformovat pro potřeby dalších analýz. Prozkoumejte závislosti mezi jednotlivými příznaky a zajímavé závislosti vizualizujte.

Modelování pomocí lineární regrese

Koncentrace kreatininu v moči by mimo jiné měla indikovat množství svalové hmoty. Pomocí modelu lineární regrese prozkoumejte závislost koncentrace kreatininu na tělesné hmotnosti, zvažte zda je možné modelovat koncentraci kreatininu pouze pomocí hmotnosti a jaký je vztah tělesné hmotnosti a hmotnosti svalové hmoty. Najděte proměnnou, která je úměrná množství svalové hmoty a přidejte ji do modelu. Některé z proměnných je vhodné transformovat.

Popište výsledný model a interpretujte jeho koeficienty. Jak dobrý je model z hlediska vysvětlení odezvy a jak významná je nalezená závislost.

Požadované kroky analýzy

Průzkumová analýza. [4 body]

- Dimenze dat (počet příznaků, počet instancí)?
- Chybějící a odlehlé hodnoty, transformace dat
- Vizualizace

Modelování vztahu mezi koncentrací kreatininu a tělesné hmotnosti [6 bodů]

- Formální zápis modelu
- Interpretace koeficientů modelu
- Přesnost modelu
- Statistická významnost

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**.

vek	věk pacienta v letech
pohlavi	pohlaví pacienta (0 - muž/ 1 - žena)
vyska	výška pacienta v cm
vaha	váha pacienta v kg
bmi	body mass index v kg/m ²
kurak	kouření (0 - nekuřák/ 1 - kuřák)
exkurak	kouření v historii (0 - vždy nekuřák/ 1 - bývalý kuřák)
dm	Diabetes mellitus (0 - ne/ 1 - ano)
neurolog_preop	Neurologické předoperační komplikace (0 - ne/ 1 - ano)
chopn	Chronická obstrukční plicní nemoc (0 - ne/ 1 - ano)
kreatinin	Koncentrace kreatininu v μmol/l
ledviny_preop	Urologické předoperační komplikace (0 - ne/ 1 - ano)
dialyza	Pacient dialyzován (0 - ne/ 1 - ano)
ef	Ejekční frakce v %
obeh_stabilni	Kardiovaskulární oběh stabilní (0 - ne/ 1 - ano)
mo	Mimotělní oběh (0 - ne/ 1 - ano)
mo_min	Doba v min, po kterou byl pacient na MO
komplikace	Pooperační komplikace (0 - ne/ 1 - ano)
peritgc	Perioperační těsná kontrola glykémie (0 - ne/ 1 - ano)
empc	Optimalizovaný algoritmus pro regulaci glykémie (0 - ne/ 1 - ano)
min	Minimální naměřená glykémie v mmol/l
max	Maximální naměřená glykémie v mmol/l
Δ	Rozdíl mezi maximální a minimální naměřenou glykemií
proc_hypo_mensi_4.4	Doba, kdy glykémie byla menší než 4.4 mmol/l v %
proc_tgc_4.4_6.1	Doba, kdy glykémie byla v oblasti 4.4 - 6.1 mmol/l v %
proc_risk_hyper_6.1_8.3	Doba, kdy glykémie byla v oblasti 6.1 - 8.3 mmol/l v %
proc_hyper_vetsi_8.3	Doba, kdy glykémie byla větší než 8.3 mmol/l v %
Ø_glykemie	Průměrná glykémie v průběhu operace
sd	Směrodatná odchylka glykémie v průběhu operace