

7 Binární logistická regrese

Příklad 1. V souboru `logistic-metacarpals.txt` máme k dispozici údaje o rozměrech druhé kosti zápěstí pravé ruky u vzorku z řecké populace. Známe také pohlaví zaznamenaných jedinců (proměnná `sex`). Sestrojte model pro odhad pohlaví neznámého případu z řecké populace na základě rozměrů kosti zápěstí: délky (proměnná `mc2.L`), šířky hlavičky (proměnná `mc2.CW`), šířky středu těla (proměnná `mc2.MW`) a šířky báze (proměnná `mc2.BW`). Všechny rozměry byly měřeny v milimetrech.

1. Zjistěte počty pozorování, výběrové průměry a výběrové směrodatné odchylky všech rozměrů zápěstní kosti zvlášť pro ženy a pro muže.
2. Vykreslete krabicové diagramy všech rozměrů zápěstní kosti zvlášť pro ženy a pro muže. Pokud vidíte nějaké extrémně odlehle pozorování, identifikujte ho a odstraňte.
3. Prozkoumejte vztah mezi vysvětlovanými proměnnými a vysvětlující proměnnou. Protože vysvětlující proměnné jsou spojité, budete používat dvouvýběrové t -testy. Nezapomeňte ověřit předpoklady.
4. Sestrojte model logistické regrese s těmi proměnnými, u kterých dvouvýběrové testy vycházely významné.
5. Proveďte celkový test významnosti modelu.
6. Na základě dílčích t -testů postupně vynechávejte nevýznamné vysvětlující proměnné a sestavte nový model. Při každém vynechání proměnné otestujte hypotézu, že je tento model dostačující.
7. Zkuste najít nejlepší model pomocí STEPWISE procedury (zkuste všechny tři varianty - `backward`, `forward` i `both`). STEPWISE procedura vybírá model na základě Akaikeho informačního kritéria (AIC).
8. Vypočítejte Negelkerkeho koeficient determinace. Pokud jste v bodě 6 došli k jinému modelu než v bodě 7, vypočítejte koeficient pro oba modely a pomocí něj je srovnajte.
9. Sestavte klasifikační tabulku a vypočítejte relativní četnost správně určených jedinců. Pokud jste v bodě 6 došli k jinému modelu než v bodě 7, vytvořte tabulku pro oba a srovnajte je pomocí relativní četnosti správně určených jedinců.
10. Vykreslete ROC křivku a vypočítejte hodnotu AUC. Pokud jste v bodě 6 došli k jinému modelu než v bodě 7, vykreslete ROC křivku s hodnotou AUC pro oba a srovnajte je pomocí AUC.
11. Na základě předchozích srovnání vyberte model, který je podle vás nejlepší. Pro něj si vypište podrobnosti modelu, vypočítejte intervaly spolehlivosti pro odhadnuté parametry, a interpretujte odhadnuté parametry.