

11 Lineární diskriminační analýza (LDA)

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`). Pomocí lineární diskriminační analýzy sestrojte funkci, která bude 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`) rozlišovat muže a ženy. Všechny rozměry byly měřeny v milimetrech.

1. Zjistěte počty pozorování, vektory průměrů a varianční matice pro obě pohlaví.
2. Orientačně ověřte linearitu vztahů mezi proměnnými. U žen je jedno extrémně odlehle pozorování, najděte ho a odstraňte ho.
3. Pro jednodušší práci s funkcí `mvn` si odstraňte sloupec `id`, který k analýze nepotřebujeme.
4. Ověřte předpoklady lineární diskriminační analýzy - normalitu dat ve skupinách a shodnost variančních matic.
5. Otestujte hypotézu o shodnosti vektorů středních hodnot. Co vám výsledek říká?
6. Sestavte lineární diskriminační funkci. Jaké jsou hodnoty koeficientů pro dané proměnné?
7. Vypočítejte podíl správně zařazených objektů.
8. Vypočítejte podíl mylně zařazených objektů. Srovnajte s podílem mylně zařazených objektů, pokud bychom případy rozřazovali pouze na základě odhadnutých pravděpodobností (tj. bez využití lineární diskriminační analýzy)
9. Na základě sestavené funkce zařaďte neznámé pozorování s hodnotami `mc2.L` 64, `mc2.CW` 14, `mc2.MW` 8, `mc2.BW` 10.
10. Vyberte proměnné pomocí dopředné krokové metody.

Příklad 2. V souboru `Howell.csv` máme k dispozici kranio-metrické údaje z různých populací. Nás zajímají muži (kategorie `M` proměnné `Sex`) ze 3 populací (proměnná `Population`) - BERG, BURIAT a PERU. Konkrétně máme tyto kranio-metrické rozměry (vše v milimetrech):

- XFB - maximální transversální šířka čela,
- NPH - výška horní části obličejového skeletu,
- NLH - výška nosu,
- OBH - výška očníce levé strany,
- OBB - šířka očníce,
- MAB - šířka patra,
- EKB - biorbitální šířka.

1. Zjistěte počty pozorování, vektory průměrů a varianční matice pro všechny populace.
2. Orientačně ověřte linearitu vztahů mezi proměnnými.
3. Ověřte předpoklady lineární diskriminační analýzy - normalitu dat ve skupinách a shodnost variančních matic.
4. Otestujte hypotézu o shodnosti vektorů středních hodnot. Co vám výsledek říká?
5. Sestavte lineární diskriminační funkci. Jaké jsou hodnoty koeficientů pro dané proměnné?

6. Vypočítejte podíl správně zařazených objektů.
7. Na základě sestavené funkce zařadte neznámé pozorování s hodnotami XFB 124, NPH 70, NLH 53, OBH 34, OBB 41, MAB 64, EKB 100.
8. Vyberte proměnné pomocí dopředné krokové metody.