

Globální analýza socioekonomických a zdravotních indikátorů a jejich vliv na úmrtnost novorozenců

1. Review

Na konci praci jste provedli hlubokou resersi problematiky a popsali jste souvisejici studia. I presto, ze dataset je docela maly, ten byl kombinaci unicef a world bank dat. Z me strany bych samozrejme takovou globalni analyzu provadel na vice datech, nejenom UNICEF a worldbank, a na vice prediktorech a vzorcich, napr., misto primocareho odstraneni zkusit zamenit chybejici data necim jinym. Celkove je videt, ze i kdyz data jsou docela znama a pouzivana v ruznych studiich, vase analyza se odlisuje od kazde z nej signifikantni zmenou. Takze, za prehled, resersi a originalitu praci dostavate plny pocet bodu, po prochazeni vseh vasich odkazu a i hledani podobnych praci na internetu se mi neprislo, ze by nekdo resil tuto otazku stejnym postupem jako Vy.

2. Dataset analysis

Analýza datasetu a jednotlivých prediktorů je videt, ze jste pouzili temer vseh technik, které jsem v tomto kurzu meli. Ale, musim zminit, ze kazda z pouzitych technik byla pouzita jen lehce a casto bez dodatecných analyz. Myslim napr. to, ze jste meli ploty s distribucemi pro kazdy sloupec, ale nezminili jste dulezita kriteria jako je, napr., typ promenne - diskretni nebo spojity, a tim padem jste i neuvedli typ distribuce, z které tyto promenne pochazi. Jako dusledek, napr., v dalsi casti jste aplikovali Poissonuv glm model. Ale cilova promenna infant mortality rate je ale spojita a ve skutečnosti sleduje log normalni distribuci. Dokonce jsem nasei i clanek který tuto skutečnost dokazuje pro Infant mortality rate.

<https://benthamopen.com/contents/pdf/TODEMOJ/TODEMOJ-4-34.pdf>

Dalsi poznamka, Mam podezreni, ze prediktor "Life expectancy" ma v sobe zahrnutou informaci o umrtnosti novorozencu, a proto neni uplne idealni volbou jako kritérium pro posuzovani o umrtnosti novorozencu. Proto musim strhnout par bodu za nedostatecnou analyzu, která Vas i vedla k chybe u Poisson GLM.

Navic -- krome analyzy distribuci a vlastnosti dat ve sve praci resite i ploty s mapovanim urovni prediktoru podle jednotlivých zemi, což beru jako hledani patternu nad ramec hlavní otázky.

3. Usage of SAN methods

Ve sve praci jste pouzili vsechny metody, které jsme v meli. Takze ano, 3 minimalni metody jste urcite splnili, ale mam komentare k aplikovani techto metod. Napr. u lineární regrese jste pouzili polynom, ale, jednak, v reportu nebylo uvedeno, ze polynom je stupne 2, a za druhé -- ani nebyla snaha zkusit aplikovat slozitejsi polynomy jako je piecewise nebo spliny. Podobna situace je i s glm, jak jsem zminil predem, aplikujete Poisson glm bez ohledu na to, ze family musi odpovidat count datum, které nemate.

U ostatních metod ale takove problémy nevidim a jste to pouzili naprosto spravne a dokonce jste i zavedli kriteria, které jsme v tomto kurzu nemeli. Proto jedina chyby které bych mohl u metody lineární regrese a glm.

4. Main question

Je videt, ze otazka zavislosti socioekonomickych faktoru a jejich vyvoj v case je aktualni problem, který je resen ve mnoha vedeckych pracich a dokonce je i soucasti praci UNICEF, od nichz jste data prevzali. S vyjimkou zahrnuti podezreleho kriteria *Life expectancy*, který by už mohl v sobe obsahovat vypocet i umrtnosti, ostatni kriteria a algoritmy, které jste zvolili, plne odpovidaji resene otazce a na konci jste spravne shrnuli sve vysledky.

5. Technical part of the report

Jak jsem rikal pred tim, malym problemem bylo to, ze jste nezahrnuli technicke informace o nekterych svych metodach (stupen a typ polynomu) a chtel bych videt vice interpretaci vysledku, ale tyto drobnosti nehrajou velkou roli. Kvalita reportu je vysoka, zkusili jste vice metod a praci povazuji za vynikajici.