

Proiect PS - Set de date Swiss

Subiectul 1 - Operatii de statistica descriptiva

- Sunt calculate media, varianta, quantilele.
- Este desenat un boxplot pentru setul de date.
- Putem observa ca exista o fertilitate mare, numarul de muncitori in agricultura este aproximativ jumatate din populatie, sunt putini oameni care iau nota maxima la examinarea fizica si putini oameni educati dincolo de scoala primara in armata. De asemenea mortalitatea infantila este mica.
- Numarul de catolici variaza mult pe regiune, fiind regiuni cu multi catolici si regiuni cu putini catolici. Totusi un numar foarte mare de regiuni au un numar mic de catolici.

Subiectul 2 - Modele de Regresie

- Am desenat grafice pentru mai multe perechi de variabile (predictor, raspuns)
 - (Education, Agriculture)
 - (Education, Catholic)
 - (Agriculture, Catholic)
- (Education, Agriculture) pare sa prezinte cea mai puternica legatura intre variabile.
- Am folosit 2 modele de regresie, Am vizualizat detalii despre ele folosind summary si lm
 - I - (Education, Agriculture)
 - II - (Education, Agriculture+Catholic)
- Dupa cum era de asteptat, valoarea p a primei regresii este mai mica si valoarea t este mai mare ceea ce inseamna ca prima regresie(cea simpla) este mai buna decat a doua (cea multipla). Mai multe detalii despre valorile p si t: <http://blog.minitab.com/blog/statistics-and-quality-data-analysis/what-are-t-values-and-p-values-in-statistics>.
- Era de asteptat ca alegerea buna a variabilelor sa fie (Education, Agriculture), din moment ce ar fi logic sa existe o legatura puternica intre cele 2. Acest lucru este vizibil si pe desen.
- In schimb nu exista vreun motiv pentru care sa presupunem ca Education ar influenta puternic Catholic.

Subiectul 3 - Repartitia Weibull

- Repartitie continua
- 2 parametri
- Folosita la a prezice rata de esuare a unor obiecte in raport cu timpul trecut de la fabricarea lor.
- k este parametrul de “shape”, determina ratele de esuare
 - $k < 1$, rata de esuare **scade** in timp, inseamna ca exista un numar mare de esuari imediat dupa fabricare.
 - $k = 1$, rata de esuare este **constanta**, inseamna ca cel mai probabil esuarile sunt cauzate de un factor extern. In acest caz este **echivalenta cu repartitia exponentiala**.
 - $k > 1$, rata de esuare **creste**, inseamna ca exista un factor de “imbatranire”, piesele au o durata de viata.
- Densitatea de probabilitate
 - Pentru $k > 1$ creste, apoi scade din nou, datorita faptului ca exista o durata de viata comuna pentru un numar foarte mare de componente.
 - Pentru $k \leq 1$ scade, deoarece “peak”ul ratei de esuare este atins la inceput.
- Functia de repartitie
 - Dupa cum ne-am astepta bazat pe cum arata densitatile, functiile de repartitie cresc abrupt pana in momentul in care este atins “peak”ul ratei de esuare. De acolo, cresterea scade drastic cu tot cu scaderea ratei de esuare.