

PROIECT PROBABILITATE SI STATISTICA

ECHIPA

APOSTU ROBERT – GRUPA 232

PUNCTUL 1

Am efectuat operatiile statistice asupra setului de date “attenu” numai pentru magnitudinea, distanta si acceleratia fiecarui cutremur, numarul cutremurului si statia fiind considerate informatii irelevante.

Magnitudine

S-a calculat o medie de 6 a magnitudinii ce, impreuna cu variatia (0.52), ne indica un numar crescut de cutremure ce au avut magnitudinea in jur de 6 pe scara Richter.

Din plot se observa in continuare numarul sporit al cutremurelor de scara 6 insa nu se poate deduce nici o regula dupa care s-au petrecut.

Din reprezentarea quartilelor se observa o tendinta a cutremurelor de a avea valori mai joase de 6.5, majoritatea fiind intre 6.5 si 5.0. Deoarece quartila a 3-a este mai mult lunga decat prima se deduce o variatie crescuta cu cat magnitudinea cutremurului este mai mare.

Distanta

Din reprezentarea plot-ului distantelor cutremurelor fata de cea mai apropiata statie putem observa o bine impartire a statiilor astfel incat majoritatea cutremurelor se afla la cel mult o distanta de 100 m fata de cea mai apropiata statie.

Din outlierii reprezentati de graficul quartilelor se pot estima viitoare posibile locatii pentru statii de masurare, in cazul in care numarul outlierilor devine prea mare.

Datorita naturii aleatoare a cutremurelor media si variatia sunt ridicate.

Acceleratia

Acceleratia este in general constanta insa se observa in plot o crestere subita a sa in jurul masuratorii numarului 100 si o descrestere rapida.

Reprezentarea quartilelor ne indica un numar de outlieri ce se pot identifica drept anomalia din plot.

PUNCTUL 2

Pentru regresia simpla am ales sa folosesc distanta si acceleratia pentru a putea avea o imagine mai buna a acceleratiei in timp, pe distanta pana la cea mai apropiata statie. Se observa in ploturi forme convexe in functie de repartitia distantei.

Pentru regresia multipla am ales sa exprim evenimentul cu magnitudinea, distanta si acceleratie pentru a caracteriza fiecare cutremur in functie de toate caracteristicile lui inregistrate. Se observa in ploturi forma unei functii la inceput convexe, apoi concave.

Consider ca regresia multipla este mai potrivita pentru setul de date deoarece are un p-value mai mic decat regresia simpla ($<2.2e-16$ vs $1.864e-11$) si deoarece criteriile Akaike si Bayesian sunt mai mici tot la ce-a multipla decat la regresia liniara (1135.513 si 1151.533 vs 1979.031 si 1988.643).

PUNCTUL 3

Am folosit repartitia Rayligh pentru a reprezenta distantele. Repartitia Rayligh este de obicei folosita pentru a afla modulul unui vector bidimensional ce are elementele intr-o distributie normala, necorelata si cu varianta egala, precum viteza vantului.

In cazul current prin functia de densitate se observa 2 cazuri de extrema , mai precis cand un cutremur se afla la aproximativ 5000m de statie si la aproximativ 12750m.