**Analiza deceselor in razboi ( 1980-2010 )**

**Tema 1. Reducerea dimensionalității Metode: Analiza în componente principale**

**Student:** Flita-Vasile Adrian

**Grupa: 1084, Seria D**

**Cuprins**

**1. Prezentare generală**

**1.1. Definirea variabilelor**

**1.2. Sursă de date**

**1.3. Analiza componentelor principale (ACP)**

**2. Analiza în componente principale**

**2.1. Analiza variabilității datelor**

**2.1.1. Distributia variantei - Calcul varianta componente**

**2.1.2. Plot varianta**

**2.2. Calcul corelații componente-variabile**

**2.3. Calculul scorurilor**

**2.4. Valorile cosinus**

**2.5. Contribuţiile instanțelor la varianța componentelor**

**2.6. Comunalităţi**

**1.1. Definirea variabilelor**

**Index-** Index

**Calendar Year**- An calendaristic

**Active Duty**- Soldati activi

**Full-Time (est) Guard+Reserve**- Garzi + Rezerve

**Selected Reserve FTE**- Rezerve selectate

**Total Military FTE**- Militari in total

**Total Deaths**- Morti in total

**Accident**–Accidente

**Hostile Action**– Actiuni ostile

**Homicide –** Ucisi de alti soldati

**Illness–** Boli

**Pending–** In curs de decidere

**Self-Inflicted-** Sinucideri

**Terrorist Attack**- Atacuri teroriste

**Undetermined**- Nedeterminat

**1.2. Sursă de date**

Datele au fost extrase de pe platforma: https://www.kaggle.com/toomuchsauce/us-military-deaths-by-cause-19802010 , setul fiind selectat dupa criteriul datelor din SUA referitoare la statisticile deceselor dintre anii 1980-2010.

**Fisierul de intrare al datelor:** U.S. Military Deaths by cause 1980-2010.csv

**1.3.** **Analiza componentelor principale (ACP)**

Analiza componentelor principale ( ACP sau APC în limba engleză pentru analiza principală componentă ), sau în domeniul de aplicare al Karhunen-Loeve ( KLT ) este o metodă de familie de [analiză a datelor](https://ro.frwiki.wiki/wiki/Analyse_des_donn%C3%A9es) și , în general [statisticile multivariate](https://ro.frwiki.wiki/wiki/Statistique_multivari%C3%A9e) , care implică transformând variabile legate între ele ( numite „corelate” în statistici) în noi variabile necorelate în afară. Aceste noi variabile sunt numite „componente principale” sau axe principale. Permite practicantului să reducă numărul de variabile și să facă informațiile mai puțin redundante.

Aceasta este o abordare atât geometrică (variabilele fiind reprezentate într-un spațiu nou, conform direcțiilor de inerție maximă), cât și statistică (cercetarea axelor independente care explică cel mai bine variabilitatea - [varianța](https://ro.frwiki.wiki/wiki/Variance_(statistiques_et_probabilit%C3%A9s)) - datelor). Când doriți să comprimați un set de variabile aleatorii, primele axe ale analizei componentei principale sunt o alegere mai bună , din punct de vedere al inertiei sau al variatiei.

**2.1. Analiza variabilității datelor**

**2.1.1. Distributia variantei - Calcul varianta componente**

Potrivit criteriilor Cattel și Kaiser, primele patru componente sunt semnficative deoarece au o varianta mai mare decat 1. Primele doua componente acopera 98,75%, respectiv 1.44% din variabilitate, celelalte doua 0,07% si 0,0%, restul spre 100% fiind preluat cumulat de celelalte 14 componente.

Text

Description automatically generated

**2.1.2. Plot varianta**

Plotul valorilor propii este compus din primele n valori proprii si se reprezintă grafic celealte n valori proprii.Valoarea lui k este determinata prin efectuarea unei taieturi astfel încât la dreapta sa aiba forma de o porțiune cu panta apropiată de 0. Din interpretarea grafica rezulta un procent de acoperire mai mare de 80%, aplicandu-se criteriul Cattell, conform caruia sunt considerate importante grupul de componente aflate până la o componentă la care se înregistrează o diferenta de pantă în graficul varianței

Graphical user interface

Description automatically generated with medium confidence

**2.2. Calcul corelații componente-variabile**

Imaginea de ansamblu este constituia de corelatiile dintre variabilele analizate in tabel. Legaturile puternice intre indicatori genereaza corelatii stranse iar în acest mod se crează exces informaţionala în setul de date. Aceste variabile generează cele mai numerose corelaţii, deci ele vor aduce cel mai mare aport la contruirea componentelor principale. Componentele principale sunt evidentiate prin aceasta trasatura, astfel din figura de mai jos reies urmatoarele legaturi stranse:

- Rezerve selectate si Soldati activi;

-atacuri teroriste si decese totale;

-Sinucideri si accidente;

Corelatiile dintre variabilele observate și componente se calculează pornind de la vectorii proprii ai matricei de corelație. Ele infatiseaza coeficienții de corelație liniară dintre variabilele observate și componentele principale.

Timeline

Description automatically generated with medium confidence

Aceste corelații sunt semnificative în încercarea de a denomina componentele principale, adică de a corela o definitie concretă componentelor principale. Denominarea mai este denumită, eticheta fiind legată de ceea ce au în comun variabilele puternic corelate cu componenta respectivă. Astfel, prima componenta (anul de razboi) este major influentata de numarul de garzi si rezerve, iar a doua componenta (Garzi si rezerve) este afectata pozitiv rezervele selectate.

Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

**2.3. Calculul scorurilor**

Chart, scatter chart

Description automatically generated

**2.4. Valorile cosinus**

Valorile cosinus descriu importanța unui an de razboi pentru o componentă în raport cu celelalte componente. In fisierul „cosinusuri.csv” sunt subliniatevalorile cosinus pentru fiecare componenta, subliniandu-se astfel anii de razboi cu o importanta majora pentru alte variabile.

Text

Description automatically generated

**2.5. Contribuţiile instanțelor la varianța componentelor**

Contribuția instanțelor la varianța componentelor defineste importanța relativă a unui an de razboi pentru o componentă în raport cu ceilalți indivizi. Din fisierul de ouput „contributii.csv”

Text

Description automatically generated

**2.6. Comunalităţi**

Comunalitățile subliniaza "schimbul" gradual de informație din variabilele observate la componentele principale, comunalitățile sunt prezentate prin graficul de mai jos. Se pot observa variabilele care au legături mai puternice cu componentele mai puțin semnificative, garzi si rezerve,precum accidente, numar toatele de morti etc.

Timeline

Description automatically generated