**ACADEMIA DE STUDII ECONOMICE, BUCUREȘTI**

**FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ**

**SPECIALIZAREA – INFORMATICĂ ECONOMICĂ**

**AN DE STUDIU: III GRUPA: 1098 SERIA: E**

O imagine care conține siglă, Grafică, Font, design grafic

Descriere generată automat

**ANALIZA NUMĂRULUI DE DECESE PE FACTORI DE RISC LA NIVELUL ȚĂRILOR LUMII**

**Coordonator științific**

Conf. univ. dr. Claudiu VINȚE

**Student**

Liviu-Ioan ZECHERU

Gabriel ZĂVOIANU

Andrei-Constantin-Ștefan VASILESCU

**BUCUREȘTI**

**2023**

Cuprins

[Sursa de date 2](#_Toc152276507)

[Referință de timp privind datele 2](#_Toc152276508)

[Descrierea variabilelor 2](#_Toc152276509)

[Descrierea observațiilor 3](#_Toc152276510)

[Abordarea folosind analiza datelor 3](#_Toc152276511)

[Motivația de a opta pentru o anumită metodă de analiză a datelor 3](#_Toc152276512)

[Prezentarea rezultatelor 4](#_Toc152276513)

[Interpretare 4](#_Toc152276514)

[Discuții și concluzii 4](#_Toc152276515)

# Sursa de date

Am ales să folosim un set de date pus la dispoziție de un utilizator de pe platforma Kaggle. Acest set de date poate fi accesat la adresa de [aici](https://www.kaggle.com/datasets/varpit94/worldwide-deaths-by-risk-factors/).

# Referință de timp privind datele

Am curățat setul de date astfel încât să-l transformăm din serii cronologice în date nestructurate, alegând anul 2017.

# Descrierea variabilelor

Variabilele setului de date sunt factori ce au dus la deces în rândul populației. Vom enumera fiecare factor cu denumirea sa aferentă în setul de date.

* sursă îndoielnică de apă potabilă (Unsafe water source)
* salubrizare necorespunzătoare (Unsafe sanitation)
* lipsă de acces la spălatul pe mâini (No acces to handwashing facility)
* poluarea aerului la nivelul gospodăriei cauzată de combustibili solizi (Household air pollution from solid fuels)
* alăptare la comun (Non-exclusive breastfeeding)
* discontinuitatea alăptării (Discontinued breastfeeding)
* malnutriție low weight-for-height (Child wasting)
* malnutriție low height-for-age (Child stunting)
* hipogreutate pentru perioada gestațională (Low birth weight for gestation)
* fumatul pasiv (Secondhand smoke)
* consumul de alcool (Alcohol use)
* consumul de droguri (Drug use)
* dietă săracă în fructe (Diet low in fruits)
* dietă săracă în legume (Diet low in vegetables)
* sex neprotejat (Unsafe sex)
* sedentarism (Low physical activity)
* nivelul de glucoză la nivelul jejunului (High fasting plasma glucose)
* nivelul total de colesterol (High total cholesterol)
* BMI mare (High body-mass index)
* presiune sistolică mare (High systolic blood pressure)
* fumat (Smoking)
* deficiență de fier (Iron deficiency)
* deficiență de vitamina A (Vitamin A deficiency)
* demineralizare osoasă (Low bone mineral density)
* poluarea aerului în interior (Air pollution)
* poluarea aerului în exterior (Outdoor air pollution)
* dietă bogată în săruri (Diet high in sodium)
* dietă săracă în cereale integrale (Diet low in whole grains)
* dietă săracă în nuci și semințe (Diet low in nuts and seeds)

# Descrierea observațiilor

Observațiile reprezintă numărul total de decese (estimate, de aceea sunt în virgulă mobilă) pe factorii descriși mai sus la nivel de țară sau entitate pentru anul 2017.

# Abordarea folosind analiza datelor

În cadrul ACP (analizei în componente principale) ...

În cadrul [PLACEHOLDER pentru voi, Gabi și Ardeu] …

# Motivația de a opta pentru o anumită metodă de analiză a datelor

În ceea ce privește ACP, am decis să alegem această metodă deoarece duce la reducerea dimensiunii setului de date, transformând variabilele inițiale într-un set de componente principale, menținând în același timp cea mai mare parte a variației din date. Acest aspect poate face mai ușoare înțelegerea și interpretarea rezultatelor. Este cunoscut faptul că această metodă identifică corelațiile între diferiți factori, ceea poate dezvălui legături subiacente între factorii de risc și numărul total de decese. Componentele principale produc o reprezentare vizuală facilă a datelor, permițând o bună înțelegere a distribuției relațiilor dintre țări și factorii de risc principali. De asemenea, prin această metodă este eliminată redundanța datelor inițiale, evidențiindu-se doar informațiile semnificative. În ultimul rând, această metodă poate gestiona existența multicoliniarității între factorii de risc.

În ceea ce privește [PLACEHOLDER pentru voi, Gabi și Ardeu] …

# Prezentarea rezultatelor

TBA

# Interpretare

TBA

# Discuții și concluzii

TBA