



## A multivariate analysis of social security costs of occupational accidents and diseases

*Uma análise multivariada dos custos previdenciários de acidentes e doenças ocupacionais*

*Un análisis multivariado de los costos de seguridad social por accidentes y enfermedades laborales*

**Liandra dos Santos Jesus<sup>1</sup>, Eliél Batistão Freitas<sup>2</sup>, Marcello Coelho Ciorlia<sup>3</sup>, Guilherme Neto Ferrari<sup>4\*</sup>, Edwin Vladimir Cardoza Galdamez<sup>5</sup>, Gislaíne Camila Lapasini Leal<sup>6</sup>, & Paulo César Ossani<sup>7</sup>**

<sup>1</sup>Department of Production Engineering, Federal University of Santa Catarina<sup>4,5,6</sup> Department of Production Engineering, State University of Maringá, <sup>2,3,7</sup> Department of Statistics, State University of Maringá

<sup>1</sup>liandrasj@gmail.com <sup>2</sup>marcello.ciorlia@gmail.com <sup>3</sup>elielbaptistao@gmail.com <sup>4\*</sup>guinetoferrari@gmail.com <sup>5</sup>evcgaldamez@uem.br <sup>6</sup>gclleal@uem.br <sup>7</sup>pccossani@uem.br

### ARTICLE INFO.

**Received: February 4, 2026**

**Approved: March 10, 2026**

**Released: May 5, 2026**

**KEYWORDS:** Clustering, multivariate analysis, occupational health, occupational safety, social security.

**PALAVRAS-CHAVE:** Clustering, análise multivariada, saúde do trabalhador, segurança do trabalhador, previdência social.

**PALABRAS CLAVE:** Clustering, Análisis Multivariado, Salud Ocupacional, Seguridad Ocupacional, Seguridad Social.

**\*Correspondent Author: Ferrari, G. N.**

### ABSTRACT

The objective of this study is to identify and characterize patterns in granting social security benefits related to occupational accidents and diseases in Brazil. A quantitative approach was adopted, analyzing a comprehensive dataset of 185,138 occupational accident benefits granted by the INSS between 2019 and 2020. The methodology combined hierarchical and non-hierarchical clustering techniques to identify distinct beneficiary groups, followed by Multiple Correspondence Analysis (MCA) to explore associations among categorical variables. The results revealed three clusters with distinct profiles regarding injury types, wage levels, and demographic characteristics. Cluster 1 is characterized by repetitive strain and manual injuries among low-wage workers; Cluster 2 features musculoskeletal and psychological disorders in higher-wage groups; and traumatic injuries among the lowest wage earners dominate Cluster 3. The MCA highlighted cohesive and isolated associations among benefit types, diseases, and geographic regions. This research provides a robust, data-driven profile of the social security costs of occupational accidents in Brazil, offering valuable insights for policymakers and OSH professionals. The identification of distinct risk groups and the mapping of associated factors support the development of more targeted and equitable prevention and compensation policies.

### RESUMO

O objetivo deste estudo é identificar e caracterizar padrões na concessão de benefícios previdenciários relacionados a acidentes e doenças ocupacionais no Brasil. Um conjunto abrangente de dados composto por 185.138 benefícios concedidos pelo INSS entre 2019 e 2020 foi explorado com técnicas de agrupamento hierárquico e não hierárquico para identificar grupos distintos de beneficiários, seguida

de Análise de Correspondência Múltipla (ACM) para explorar associações entre variáveis categóricas. Os resultados revelaram três clusters, com perfis diferenciados quanto aos tipos de lesões, faixas salariais e características demográficas. O Cluster 1 é caracterizado por lesões manuais e por esforços repetitivos entre trabalhadores de baixa remuneração; o Cluster 2 reúne distúrbios musculoesqueléticos e psicológicos entre trabalhadores de salários mais elevados; e o Cluster 3 é predominado por lesões traumáticas entre os trabalhadores de menor renda. A ACM evidenciou associações coesas e isoladas entre tipos de benefícios, doenças e regiões geográficas. Este estudo fornece um perfil robusto e baseado em dados sobre os custos previdenciários dos acidentes de trabalho no Brasil, oferecendo subsídios relevantes para formuladores de políticas e profissionais. A identificação de grupos de risco distintos e o mapeamento de fatores associados fortalecem o desenvolvimento de políticas de prevenção e compensação direcionadas e equitativas.

### RESUMEN

El objetivo de este estudio es identificar y caracterizar patrones en la concesión de beneficios de seguridad social relacionados con accidentes y enfermedades ocupacionales en Brasil. Un conjunto de datos de 185.138 beneficios otorgados por el INSS entre 2019 y 2020 fue explorado con técnicas de agrupamiento jerárquico y no jerárquico para identificar grupos distintos de beneficiarios, seguida de un Análisis de Correspondencia Múltiple (ACM) para explorar asociaciones entre variables categóricas. Los resultados revelaron tres clusters, con perfiles diferenciados en cuanto a los tipos de lesiones, rangos salariales y características demográficas. El Cluster 1 se caracteriza por lesiones manuales y por esfuerzos repetitivos entre trabajadores de baja remuneración; el Cluster 2 reúne trastornos musculoesqueléticos y psicológicos entre trabajadores de salarios más elevados; y el Cluster 3 está predominado por lesiones traumáticas entre los trabajadores de menores ingresos. El ACM evidenció asociaciones cohesivas y aisladas entre tipos de beneficios, enfermedades y regiones geográficas. Este estudio proporciona un perfil robusto y basado en datos sobre los costos de seguridad social de los accidentes de trabajo en Brasil, ofreciendo insumos relevantes para formuladores de políticas y profesionales. La identificación de grupos de riesgo distintos y el mapeo de factores asociados fortalecen el desarrollo de políticas de prevención y compensación direcionadas y equitativas.

## INTRODUCTION

Work-related accidents, occupational diseases, and fatalities represent a significant challenge for organizations across all economic sectors. Occupational Safety and Health (OSH) has become a critical factor in fostering responsible labor relations with employees, society, and regulatory agencies tasked with monitoring workplace conditions. Ensuring workers operate safely, with minimal exposure to occupational risks and eliminating hazardous sources, is essential for sustainable operations.

However, despite these growing concerns and regulatory frameworks, persistent negligence and inadequate commitment to OSH practices continue to result in alarming statistics at both global and national levels. More than 370 million non-fatal workplace accidents and nearly 2.8 million fatalities are recorded worldwide each year (ILO, 2020). In Brazil, the situation mirrors this global trend: in 2022 alone, out of more than 44 million formal jobs, 619,920 occupational accidents were reported, as documented by the Information Panel of the New CAGED and the Occupational Health and Safety Observatory (SMARTLAB) (Brasil, 2025). These figures suggest that one in every 80 Brazilian workers may experience a work-related accident each year, considering both reported and unreported cases of Work Accident Report (*Comunicação de Acidente de Trabalho, CAT*).

Workplace accidents incur both direct and indirect costs. Direct costs are explicitly tied to the incident and typically involve immediate monetary outlays, such as medical expenses, compensation, and legal fees. Indirect costs, however, are more insidious and often go unaccounted for in formal reporting. These include productivity losses, retraining expenses, workflow disruption, and reputational damage. Prior research has shown that indirect costs may be up to four times higher than direct costs (Hämäläinen, Takala, & Kiat, 2017; Paez et al., 2006; SWA, 2017), amplifying the overall impact of occupational accidents.

In Brazil, Law N.º 8.213 of July 24, 1991, regulates OHS and outlines the framework of the General Social Security System (*Regime Geral da Previdência Social, RGPS*) (Brasil, 1991). Under this system, the National Institute of Social Security (*Instituto Nacional do Seguro Social, INSS*) covers direct costs associated with occupational accidents. The INSS manages the registry of benefits granted to injured workers, whether the outcome involves death (triggering pension payments to dependents) or temporary or permanent incapacity to work (Santana et al., 2006).

The benefits of workplace accidents and occupational diseases have been steadily increasing, leading to a corresponding rise in social security expenditures. In 2019, Brazil's Social Security system granted a total of 5.19 million benefits, amounting to an expenditure of R\$ 7.56 billion; this reflects an increase of approximately 1.28% in the number of benefits granted and a 6.57% rise in spending compared to the previous year (Brasil, 2020). The funding of these benefits falls under the responsibility of system contributors, as there is no individual allocation mechanism for collecting these resources. Consequently, the balance between contributions and expenditures remains structurally unequal (Botelho et al., 2018).

Therefore, occupational accidents and diseases impose significant costs on workers, employers, the government, and society (Botelho et al., 2018; Gouveia, Horsth, & Faroni, 2017). In Brazil, the situation is financially unsustainable, as the revenue generated from occupational accident insurance falls short of covering the benefits paid to injured workers or their dependents (Sá, Gomide, & Sá, 2016). According to Botelho and Costa (2020), social

security expenditures account for approximately 12% of the national GDP and are expected to rise.

Given the substantial economic and social impacts caused by work-related accidents, diseases, and fatalities (ILO, 2020), it becomes imperative to thoroughly understand and analyze the data and repercussions associated with these incidents (Sanni-Anibire et al., 2020). Although "accident" often implies an unforeseeable or unavoidable event, several preventive measures can be implemented to avoid such occurrences or mitigate their consequences (Rajabi et al., 2025).

Thus, the costs associated with workplace accidents should be analyzed from two perspectives: (i) economic cost, which is subdivided into direct cost (insured) - encompassing expenses related to medical treatment and rehabilitation - and indirect cost (uninsured), referring to lost opportunities due to the accident, such as reduced productivity; (ii) human cost, which pertains to the decline in the quality of life of the worker and their family. In addition to these costs, workplace accidents result in human fatalities, material damage, decreased productivity, harm to the company's image, and psychological effects on the workforce (Sá; Gomide; Sá, 2016).

In this context, the available databases used to measure workplace accidents record only formal employment incidents, and even then, not all accidents are accounted for due to underreporting, which remains a persistent issue in workplace accident statistics (Menegon, Menegon, & Kupek, 2021; Sá, Gomide, & Sá, 2016). The main reasons for underreporting include: lack of awareness regarding the legal obligation to report accidents, time constraints due to heavy workloads, workers' fear of being dismissed, and limitations of the registration system, which only accepts records from formally employed individuals (Sá, Gomide, & Sá, 2016).

In addition, a systematic mapping conducted by Reis et al. (2021) highlights the difficulty in finding research that addresses data mining and relates the costs of workplace accidents to companies, government, and OSH. Various countries conducted research on the costs associated with workplace accidents, including the United States (Kakhki et al., 2019; Leigh, 2011), the United Kingdom (Kakhki et al., 2019), Australia (PAEZ et al., 2006), and the European Union (HSE, 2020). In Brazil, Sá, Gomide and Sá (2016) and Eyerkauffer et al. (2019) carried out studies estimating the occupational costs.

Therefore, analyzing records of occupational accident benefits enables the investigation of their characteristics and influencing factors, providing insights into the conditions leading to worker incapacitation. Consequently, numerous studies have been conducted to establish relationships between work conditions and their macro and microeconomic impacts, employing multivariate statistical tools. For instance, Ferrari et al. (2023) used Multiple Correspondence Analysis (MCA) to highlight associations between workplace accidents and heat stress indices, identifying worker profiles at higher risk and regions in Brazil more exposed to such conditions. Their findings indicated a positive correlation between workplace accidents and climatic factors.

Similarly, Cabello et al. (2021) applied Association Rule methods to identify predominant attributes and patterns of occupational accidents in the construction sector. They emphasized that subcontracting is a critical factor in workplace accidents within this sector and

recommended that companies require OSH certifications from freelance workers to ensure proper education and training. The researchers also noted that risk assessment is not an insurmountable barrier and suggested increasing evaluations at different stages of construction projects.

Even so, identifying patterns that accurately represent the reality of occupational accidents through relational analyses necessitates a substantial volume of historical data. For example, Ferrari et al. (2023) analyzed 404,355 occurrences, while Cabello et al. (2021) examined 5,495,609 instances. An agglomerative hierarchical clustering approach was initially employed to determine, a priori, the appropriate number of clusters, through the construction of a dendrogram, which allowed visualization of the clustering structure. Based on this initial estimate, the k-means algorithm, a non-hierarchical method, was then applied to form the final groups, as it provides greater robustness and precision in defining well-separated clusters, while the hierarchical method is suitable for suggesting the number of clusters. The approach involves clustering similar instances and distinguishing dissimilar ones using clustering techniques. Each group will be described through Multiple Correspondence Analysis (MCA), highlighting the most relevant characteristics to assist decision-making by industry and government.

We present our paper in five segments. The first is this introduction, followed by the theoretical framework, where we present and discuss the theoretical concepts essential to understanding our research. The third segment is the materials and methods, describing our methodological research approach. We present our results in the fourth segment and conclude our study with our considerations in the final segment.

## **THEORETICAL FRAMEWORK**

This section presents important theoretical concepts related to the methods used in this paper, including the clustering technique, which is divided into hierarchical and non-hierarchical clustering, and multiple correspondence analysis.

Clustering techniques can be used for various purposes, including grouping problems, dimensionality reduction, and profile identification (Fernandes, 2021). In grouping scenarios, clustering involves dividing observations in a dataset such that groups together those with similar measured characteristics, while observations in different groups exhibit heterogeneous traits (Rencher & Christensen, 2012; Xu & Tian, 2015).

They rely on distance measurements to determine the degree of similarity between two observations in a dataset. There are numerous ways to compute these distances, such as Euclidean, Manhattan, Minkowski, Canberra, and Maximum distances (Rencher & Christensen, 2012).

Consequently, there are various clustering algorithms, each with different starting points and distinguishing criteria. A simple and widely accepted classification separates these techniques into hierarchical and non-hierarchical methods (Xu & Tian, 2015).

Hierarchical clustering is typically represented by a diagram known as a dendrogram, which helps summarize and visualize data. Within this classification, there is a subclassification into agglomerative and divisive algorithms. The former assumes that each observation is a cluster and progressively merges them to form groups. The latter assumes a single cluster that contains all the data and then subdivides it into smaller groups (Zubaroglu & Atalay, 2020).

Specifically, agglomerative clustering uses a linkage criterion to cluster similar observations. Ward minimizes the sum of squared differences within all clusters and, in this sense, is similar to the k-means objective function, but follows a hierarchical agglomerative approach; Complete linkage minimizes the maximum distance between observations of pairs of clusters; Average linkage minimizes the average of the distances between all observations of pairs of clusters; and single linkage minimizes the distance between the closest observations of pairs. This technique can be scaled to a large dataset when used in conjunction with a connectivity matrix. However, it becomes computationally expensive when adding no connectivity constraints, as it considers all possible merges at each step (Varoquaux et al., 2015).

Correspondence Analysis is an exploratory and multivariate technique whose result is the graphical representation of the rows and columns of a contingency table. In this context, the contingency table is the intersection of various categories in which the number of instances in each "cell" is counted (Souza, 1990).

It can also be understood as an algorithm for reducing qualitative or categorical variables, aiming to represent data sets through a graphical model that facilitates interpreting relationships between the variables and observations that make up such sets (Rencher & Christensen, 2012; Souza, 1990). Furthermore, not adopting any probability distribution model can reveal relationships that would not have been perceived if the analysis were conducted on pairs of variables (Rencher & Christensen, 2012).

This technique can be either simple or multiple. The former represents information from a two-way contingency table, which contains the frequencies of items from a cross-classification of two categorical variables (Souza, 1990). However, when there are three or more entries, the method is called MCA, as it produces a two-dimensional graph of the instances in the multi-entry contingency table. In this case, an indicator matrix is analyzed, and each categorical variable is transformed into a dummy variable that is expanded into several vectors equal to the number of categories. Which 1 represents the presence of the category and 0 represents the absence of the category in the respective vectors (Rencher & Christensen, 2012).

While simple correspondence analysis deals with the observed frequencies of each categorical variable, multiple correspondence analysis transforms the categorical variables into dummy variables to represent the original variables.

## **MATERIALS AND METHODS**

The INSS makes various labor-related data available following Decree Nº. 8.777/16 (Brasil, 2016) and the Access to Information Law Nº. 12.527/2011 (Brasil, 2011). Among these are data on granted benefits, which, according to Article 27 of Law Nº. 8.213, include the provision of maternity leave, sick pay, and disability retirement benefits for insured employers and casual workers (Brasil, 1991). We extract these data from SUIBE, a Unified Benefit Information System, a management platform that centralizes and stores information related to beneficiaries and their benefits.

In this sense, we analyzed benefits granted between 2019 and 2020. Since they were available monthly, it was necessary to merge the 24 files to perform data preprocessing, during which instances with missing data were first excluded. Subsequently, the categorical attributes were standardized.

Initially, the database contained 13 attributes. However, the model did not use the granting period, International Statistical Classification of Diseases and Related Health Problems (ICD) name, concession type, beneficiary group, dependents' relationship, and affiliation type. We converted dates of birth into age and grouped them, aggregated municipality of residence into states, and reclassified ICD codes using the category level of the ICD structure. We retained benefit types, gender, the number of minimum wages in the Initial Monthly Income, and state (Table 1).

**Table 1.** Description of the attributes used in the model

Attributes	Class / Description
Gender	Masculine, Feminine
Age Groups (years)	15 - 25, 25 - 35, 35 - 45, 45 - 55, 55 - 65, 65 - 75, 75 - 85, 85 - 95
Occupational Accident Benefits	B91: Work-Related Sickness Benefit B92: Work-Related Disability Pension B93: Death Pension due to Work-related Accident B94: Disability Benefit due to Work related Accident
International Statistical Classification of Diseases and Related Health Problems (ICD)	The ICD codes have been truncated, leaving only the first two characters
State of Residence	State where the beneficiary resides
State	State where the accident occurred
Competence Granting	Date which the benefit was granted
Number of minimum wages used in the Initial Monthly Income	Number of minimum wages

Source: Authors.

Since the INSS manages all social security-related benefits, this study focused only on occupational accident benefits, which comprise four specific types, as outlined in Table 1.

The first, Work-Related Sickness Benefit (B91), is granted to workers who become temporarily unable to perform their duties due to a work-related accident or occupational disease. No minimum contribution period is required. Payment begins on the 16th day of work absence. While receiving the benefit, the employer must continue to deposit into the worker's Severance Indemnity Fund. Upon returning to work, the employee is entitled to 12 months of job stability, meaning they can only be dismissed for good cause.

The Work-Related Disability Pension (B92) is intended for insured individuals who become permanently unable to work due to a work-related accident or occupational disease. Like the B91, it requires no minimum contribution period. The benefit amount corresponds to 100% of the average salary of contributions, with no reductions. This pension is discontinued if the insured person regains the ability to work or resumes any professional activity.

The Death Pension due to Work-related Accident (B93) is paid to the dependents of the insured individual who dies as a result of a work-related accident. This benefit aims to ensure financial support for the worker's family.

Lastly, Disability Benefit due to Work-related Accident (B94) is an indemnity benefit granted to insured individuals who, after stabilizing injuries (whether work-related or of any nature), are left with a permanent condition that reduces their work capacity. The worker may continue performing their professional activities, and the benefit amount corresponds to 50% of the compensation. This benefit is terminated in cases of retirement, death, or upon request for a Certificate of Contribution Time (CTC). Eligibility requires medical verification of reduced work capacity by an INSS medical examiner.

The database comprises 185,138 granted benefits, which were converted into dummy variables, binary representations commonly used in statistical analyses, to enable clustering and multiple correspondence analysis. An agglomerative hierarchical clustering approach was initially applied, using Ward's method as the linkage criterion and Manhattan distance to determine the optimal number of clusters ( $k$ ). We followed this step by applying the  $k$ -means algorithm, a non-hierarchical method generally considered more effective for forming distinct clusters. Finally, we performed a Correspondence Analysis to identify associations between the variables. We conducted all analyses using scripts developed in R software (R Core Team, 2023), employing the publicly available MVar package (version 2.2.1) (Ossani & Cirillo, 2025).

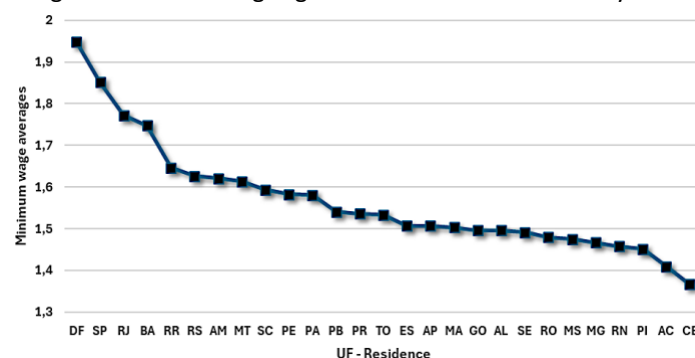
## RESULTS AND DISCUSSIONS

### DESCRIPTIVE ANALYSIS

The data revealed a disparity in the average benefits received by men and women. On average, men receive benefits equivalent to 1.6 times the minimum wage. In contrast, women receive 1.5 times, a difference that may be attributed to the broader social phenomenon of gender-based wage inequality in the labor market (Madalozzo & Artes, 2017; Tenoury, Madalozzo, & Martins, 2021). A detailed statistical analysis using data from the Annual Social Information Report (*Relação Anual de Informações Sociais*, RAIS) confirms that women continue to receive lower wages than men, regardless of educational level, sector of employment, or age. This inequality is even more pronounced in the higher salary ranges. While the distribution between men and women is relatively balanced in the brackets up to two minimum wages, only 31.4% are women compared to 68.6% of men above twenty minimum wages. This pattern persists over the years and is not explained by differences in qualifications, as women exhibit higher average educational attainment, particularly in the highest salary ranges, where the proportion of women holding a doctoral degree surpasses that of men (Bolsoni, 2024).

A noticeable difference is also evident at the state level, as Figure 1 presents the average minimum wages granted to beneficiaries across each state and federal unit. Among these, the Federal District (DF), São Paulo (SP), Rio de Janeiro (RJ), Bahia (BA), and Roraima (RR) stand out for having the highest average benefit amounts.

**Figure 1.** The average of minimum wages granted to the beneficiaries by state and federal unit

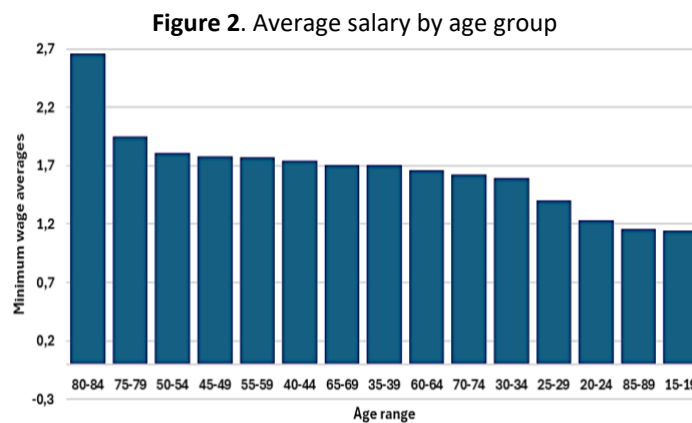


Source: Authors.

This pattern holds for the Federal District, São Paulo, and Rio de Janeiro when considering the national average remuneration. For the Federal District, this prosperity is primarily attributed to the region's concentration of administrative and political functions, which drives up salaries, particularly in the public sector (Perez, 2025). In the case of São Paulo and Rio de Janeiro, this is linked to the fact that a significant portion of Brazil's industrial production chain is concentrated in the country's southeastern region. However, this pattern does not apply to

Bahia and Roraima, as Minas Gerais (MG) has a higher number of benefits granted and reports a larger average remuneration in the country, which could be related to the work activities performed. However, the database does not provide information on the workers' areas of employment, preventing further analysis of this factor.

Regarding the age groups of the beneficiaries, as Figure 2 shows that individuals aged between 80 and 84 years have the highest average wages, while those between 15 and 19 years have the lowest. For older beneficiaries, higher average wages can be attributed to increased life expectancy and the need to supplement income during retirement, often resulting in longer careers and greater accumulation of experience and qualifications. Additionally, older workers are more likely to have reached advanced positions in their professional trajectories, which tend to offer higher salaries (Sindeaux et al., 2025).



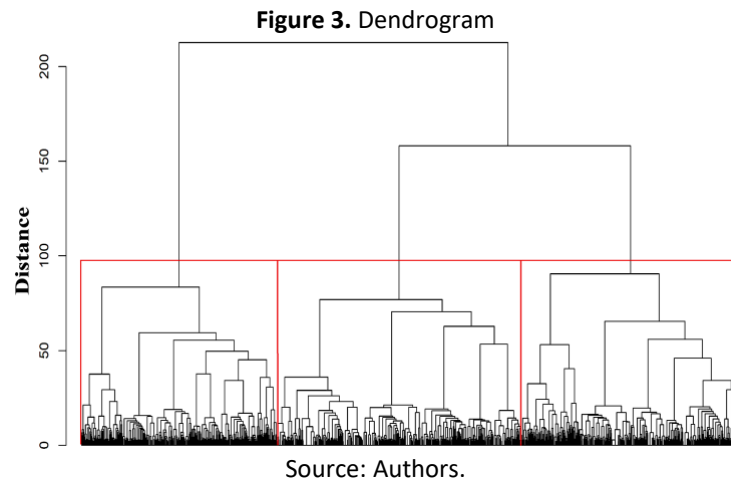
In contrast, the lower average wages observed among the youngest group (15 to 19 years) are directly related to Brazilian labor legislation. Individuals under 18 are subject to strict limitations regarding working hours and the types of activities they can perform. For example, those aged 14 to 15 may only work as apprentices, with a maximum of 6 hours per day and 30 hours per week, while those aged 16 and 17 are limited to 8 hours per day and 44 hours per week, and are prohibited from working in hazardous, unhealthy, or night-time conditions. These legal restrictions, combined with the fact that young workers are typically at the beginning of their careers and often lack experience or formal qualifications, contribute to their lower average wages (Sindeaux et al., 2025).

Thus, the observed wage distribution by age group reflects both structural factors- career progression and accumulated experience for older workers- and regulatory protections that limit the working conditions and earnings of younger individuals, aiming to safeguard their development and educational opportunities (Sindeaux et al., 2025).

### CLUSTERING

Determining the optimal number of clusters is a crucial step in cluster analysis, as it directly impacts the quality and interpretability of the results. Initially, an agglomerative hierarchical clustering method was employed, using the Manhattan distance combined with Ward's linkage criterion. This combination showed better performance in indicating the number of clusters compared to other distance metrics and linkage criteria, while also promoting the formation of more homogeneous groups by minimizing within-cluster variability. This procedure aimed to obtain an appropriate estimate of the number of clusters through the construction of a dendrogram, which allowed visualization of the data's clustering structure.

Resulting dendrogram, Figure 3 highlights the presence of 3 main clusters, particularly evident when a cut-off point is considered around 100. This visual analysis was essential to guide the preliminary definition of the number of clusters, which was subsequently used in the application of non-hierarchical methods to consolidate the groupings.



Based on this structure suggested by the hierarchical method, the non-hierarchical k-means algorithm was applied, fixing  $k=3$ . This combined approach is recommended in the literature, as the hierarchical method helps identify the appropriate number of clusters. At the same time, k-means, due to its iterative nature, refines the allocation of elements, resulting in more well-defined clusters with more precise boundaries.

Therefore, considering that the total number of beneficiaries was 185,138, Cluster 1 consisted of 60,557 individuals (32.7%), Cluster 2 had 54,165 individuals (29.3%), and Cluster 3 included 70,416 individuals (38.0%), showing a balanced division. Regarding the ICDs, Table 2 presents the distribution of ICD codes by cluster, highlighting those representing 70% of each group.

**Table 2.** Distribution of ICDs by cluster

Cluster 1		Cluster 2		Cluster 3	
ICD	Proportion	ICD	Proportion	ICD	Proportion
S6	25.92 (14,040)	M6	12.7 (7,692)	S6	28.02 (19,728)
S8	13.63 (7,382)	S9	12 (7,269)	S8	13.16 (9,264)
S9	11.88 (6,433)	S8	9.67 (5,857)	S9	11.61 (8,178)
M4	9.09 (4,924)	S6	8.52 (5,157)	S5	8.31 (5,852)
S5	8.47 (4,587)	M4	7.84 (4,750)	M4	8.14 (5,731)
S4	8.45 (4,576)	F4	6.31 (3,819)	S4	7.88 (5,547)
M6	8.23 (4,460)	G5	6.25 (3,785)	M6	3.86 (2,717)
F4	4.05 (2,195)	S5	6.25 (3,784)	S3	2.02 (1,423)
S3	2.28 (1,235)	F3	4.23 (2,562)	S7	1.82 (1,284)
S2	1.88 (1,020)	S4	3.84 (2,328)	S0	1.77 (1,249)
M2	1.82 (987)	M2	2.05 (1,244)	S2	1.57 (1,107)
T0	1.72 (934)	S3	1.32 (799)	T0	1.55 (1,094)
S7	1.71 (927)	T2	1.04 (628)	M2	1.49 (1,050)
S0	1.61 (872)	T0	0.8 (484)	T2	1.19 (835)
T2	1.55 (840)	S7	0.77 (465)	T9	0.93 (658)
F3	1.54 (834)	I8	0.69 (415)	F4	0.87 (613)
G5	1.49 (805)	S2	0.67 (404)	K4	0.81 (571)
K4	1.07 (577)	S0	0.58 (352)	G5	0.57 (402)
T9	0.96 (521)	T9	0.45 (270)	F3	0.41 (290)
				S1	0.36 (257)

Source: Authors.

The analysis of the distribution of injury types across the identified clusters reveals distinct patterns of occupational health risks. In Cluster 1, there is a predominance of wrist and hand injuries (S6), accounting for 25.92% of cases, followed by knee and lower leg injuries (S8) at 13.63%, ankle and foot injuries (S9) at 11.88%, as well as dorsopathies (M4) and elbow and forearm injuries (S5), each representing 9.09%. This profile suggests greater exposure to activities that require repetitive movements or intensive manual tasks, often associated with work environments that demand manual dexterity or localized physical effort.

In Cluster 2, the distribution is more diverse. Disorders of soft tissues (M6) lead with 12.7%, followed by ankle and foot injuries (S9) at 12.0%, knee and lower leg injuries (S8) at 9.67%, and wrist and hand injuries (S6) at 8.52%. Dorsopathies (M4) account for 7.84%, while neurotic, stress-related, and somatoform disorders (F4) represent 6.31%. Injuries to the elbow and forearm (S5) and nerve root and plexus disorders (G5) each account for 6.25%, and mood disorders (F3) represent 4.23%. This cluster is characterized by a significant combination of musculoskeletal disorders and psychological conditions, indicating the presence of both psychosocial and ergonomic risk factors in the workplace.

Cluster 3 shows an even stronger concentration of traumatic injuries, with wrist and hand injuries (S6) accounting for 28.02% of cases, followed by knee and lower leg injuries (S8) at 13.16%, ankle and foot injuries (S9) at 11.61%, dorsopathies (M4) at 8.19%, and shoulder and upper arm injuries (S4) at 7.88%. This profile suggests involvement in activities with higher physical impact and accident risks, likely related to industrial or operational sectors.

Overall, the high incidence of wrist and hand injuries across all three clusters highlights persistent ergonomic challenges. At the same time, the significant presence of psychological disorders in Cluster 2 emphasizes the importance of psychosocial factors in the work environment. These findings underscore the need for targeted and integrated prevention strategies that address both physical risks and organizational and emotional factors in the workplace.

Table 3 shows the distribution of concession types across the three clusters. Each row corresponds to a cluster, while each column represents a specific concession type: administrative concessions, meaning benefits granted directly by the INSS without the need for special procedures (Admin.); benefits granted based on Article 27 of Law No. 8,213/1991, which addresses specific concession situations (Article 27); benefits granted based on Article 35 of the same law, typically related to exceptional cases such as rehabilitation (Article 35); benefits granted through judicial proceedings following a court decision (Judicial); standard grants for occupational accident benefits (Regular); and benefits granted following an administrative appeal (Appeal).

**Table 3.** Distribution of concession type by cluster

Cluster	Adm.	Article 27	Article 35	Judicial	Regular	Appeal
1	0.30% (183)	0.04% (24)	0	0.07% (42)	99.6% (60,302)	0.01% (6)
2	0.42% (228)	0.08% (42)	0.26% (139)	0.13% (70)	99.1% (53,676)	0.02% (10)
3	0.69% (485)	0.12% (85)	0.49% (345)	0.07% (49)	98.6% (69,427)	0.04% (25)

Source: Authors.

Across all clusters, the predominant concession type is the regular grant, accounting for 99.1% in Cluster 1, 99.6% in Cluster 2, and 98.6% in Cluster 3; this indicates that most benefits were granted through the standard procedure without the need for administrative, judicial, or appeal processes.

The other types of concessions appear at significantly lower rates. Administrative concessions, for instance, range from 0.3% to 0.7% across clusters, while concessions based on Articles 27 and 35 of Law No. 8,213/1991 and judicial concessions show even lower percentages, generally below 0.5%. There were no records of appeal concessions in any of the groups.

The table highlights that regular concession is the standard and overwhelmingly predominant procedure across the three analyzed clusters. In contrast, other concession forms are rare, occurring only in exceptional cases.

Table 4 presents individuals' percentage and absolute distribution by age group within the three clusters identified in the occupational accident benefit data analysis. The columns represent different age ranges (in years), while the rows correspond to the groups formed through the clustering method. The percentage value is shown in each cell, followed by the absolute number of beneficiaries in parentheses for that specific age group and cluster.

**Table 4.** Distribution of age groups by cluster

Cluster	15-25	25-35	35-45	45-55	55-65	65-75	75-85	85-95
1	2.9% (1,766)	22.1% (13,382)	33.8% (20,471)	27.5% (16,632)	12.6% (7,611)	0,57 (687)	0.008% (5)	0,005% (3)
2	5.3% (2,870)	23.7% (12,847)	31.7% (17,172)	26.7% (14,443)	12.0% (6,524)	1.1% (307)	0.004% (2)	0
3	13.3% (9,366)	30.8% (21,712)	27.0% (19,037)	18.4% (12,977)	9.6% (6,760)	0.8% (564)	0	0

Source: Authors.

In Group 1, there is a higher concentration of individuals aged between 35 and 45 (33.8%), followed by those aged 45 to 55 (27.5%) and 25 to 35 (22.1%). The extreme age groups, such as 15 to 25 (2.9%) and over 65 (less than 1%), have a much lower representation; this indicates that the group predominantly comprises adults in the mid-productive age range.

Group 2 shows a similar pattern, with the most significant shares in the 35 to 45 years (31.7%), 25 to 35 years (23.7%), and 45 to 55 years (26.7%) brackets. Younger individuals (15 to 25 years, 5.3%) and older (65 to 75 years, 1.1%) are even less represented than Group 1.

Group 3, however, shows a higher proportion of younger individuals: 13.3% are in the 15 to 25-year range, significantly more than in the other groups. Still, the majority falls between 25 and 45 years (30.8% for 25 to 35 years and 27.0% for 35 to 45 years). The representation of individuals over 55 years is very low and practically nonexistent beyond 75 years.

Overall, the table reveals that most occupational accident benefits are granted to workers in the productive age range, particularly between 25 and 55 years old. Group 3 includes more young individuals, while Groups 1 and 2 are more concentrated in intermediate age ranges. The participation of older adults is minimal across all groups. These patterns reflect the age profile of workers most exposed to accidents or work-related illnesses, predominantly composed of adults in active employment.

Table 5 presents the distribution of beneficiaries across clusters based on the range of minimum wages received. The rows correspond to the three clusters identified in the analysis, while the columns represent different intervals of minimum wage multiples (MWM) received by the beneficiaries: 0.9 to 1.5, 1.5 to 2.5, 2.5 to 3.5, 3.5 to 4.5, and 4.5 to 6.0 times the minimum wage.

**Table 5.** Distribution of minimum wage multiples by cluster

Cluster	0.9 -1.5	1.5-2.5	2.5-3.5	3.5-4.5	4.5-6.0
1	0	74.4% (45,069)	16.2% (9,829)	5.9% (3,576)	3.4% (2,083)
2	69.9% (37,858)	18.8% (10,181)	5.4% (2,934)	3.6% (1,923)	2.3% (1,269)
3	100% (70,416)	0	0	0	0

Source: Authors.

Cluster 1 has no beneficiaries in the 0.9 to 1.5 minimum wage multiple range. The majority (74.42%, or 45,069 individuals) receive between 1.5 and 2.5 minimum wages. Additionally, 16.23% (9,829) fall into the 2.5 to 3.5 MWM range, 5.91% (3,576) between 3.5 and 4.5 MWM, and 3.44% (2,083) between 4.5 and 6.0 MWM. This pattern shows that the beneficiaries in this group tend to earn slightly higher average wages than those in Cluster 2.

In Cluster 2, most beneficiaries (69.89%, or 37,858 individuals) receive between 0.9 and 1.5 times the minimum wage. Another 18.8% (10,181 individuals) fall within the 1.5 to 2.5 MWM range. Smaller percentages appear in the higher brackets: 5.42% (2,934 individuals) between 2.5 and 3.5 MWM, 3.55% (1,923) between 3.5 and 4.5 MWM, and 2.34% (1,269) between 4.5 and 6.0 MWM; this indicates that most beneficiaries in this group receive wages close to the minimum wage.

Cluster 3, on the other hand, is composed exclusively of beneficiaries receiving between 0.9 and 1.5 times the minimum wage multiple, totaling 100% (70,416 individuals) within this range, with no records in the other salary brackets. Also, the table shows that the clusters have distinct salary profiles: Cluster 1 is concentrated around values close to the minimum wage, Cluster 2 around slightly higher salary ranges, and Cluster 3 exclusively among beneficiaries between 0.9 and 1.5 times the minimum wage.

#### **MULTIPLE CORRESPONDENCE ANALYSIS**

The Multiple Correspondence Analysis (MCA) was used to deepen the understanding of associations among categorical variables within clusters identified by cluster analysis. MCA is an exploratory multivariate technique that graphically represents the relationships between categories of different variables, making it easier to identify patterns and latent structures in the data. By transforming categorical variables into a binary matrix, MCA projects the categories into a low-dimensional space where associated categories tend to be placed close together. It is important to note that while this method reveals statistical associations, it does not establish causal relationships. The findings should be interpreted as patterns of co-occurrence within the dataset, which are influenced by the available variables and their categories.

This study applied MCA to each of the three identified groups to uncover patterns of association among the variables that could help characterize workplace accidents and the social security benefits granted. This approach enabled visualization of key associations among variables such as gender, age group, federal unit, type of illness (ICD), and benefit type using perceptual maps.

In the MCA chart, each point represents a category of a variable (e.g., an ICD code, a state, or a type of decision). The position of the points in the two-dimensional plane is determined by the first two dimensions extracted by the MCA. These explain the most considerable proportions of variability in the associations between the categories. Categories that appear close to each other in the chart tend to be more strongly associated in the data, while categories farther apart indicate little or no association.

Thus, in Figure 4, related to cluster 1, it can be observed that the Concession by Article 27 is positioned isolated on the MCA chart, suggesting a lack of significant association with the other variables analyzed. Similarly, the categories "state" Ceará (CE), "residential state" CE, "state" Bahia (BA), and "residential state" BA also appear distanced from the central cluster,





worker population at risk, nor does it include detailed occupational variables (e.g., industry, occupation, specific tasks), which limits the ability to generalize these findings to broader populations and to establish causal links between work conditions and the observed health outcomes. Therefore, this study presents an exploratory characterization of patterns among benefit recipients, and these results can provide hypotheses for future research.

### FINAL CONSIDERATIONS

The results of this study provide a comprehensive and nuanced understanding of the social security costs associated with occupational accidents and diseases in Brazil. By applying advanced multivariate techniques, clustering and multiple correspondence analysis (MCA), to a large national dataset of INSS benefit records, the research uncovers distinct beneficiary profiles, reveals patterns of injury and illness, and highlights significant socioeconomic disparities.

Identifying three well-defined clusters demonstrates the heterogeneity among beneficiaries of occupational accident benefits. Cluster 1 is characterized by repetitive strain and manual injuries, such as those affecting the wrist, hand, knee, and lower limbs, reflecting the risks inherent in manual and repetitive work environments. Cluster 2 stands out for its combination of musculoskeletal and psychological disorders, including a notable presence of soft tissue injuries and stress-related conditions, highlighting the relevance of both ergonomic and psychosocial risks in specific occupational contexts. Cluster 3 is marked by a concentration of traumatic injuries, particularly among low-wage workers, suggesting heightened exposure to high-impact physical labor and potentially limited access to preventive measures.

The analysis of wage distribution across clusters and demographic groups reveals persistent inequalities. Men receive higher average benefits than women, reflecting broader gender-based wage disparities in the labor market. Regional disparities are also evident, with states like the Federal District, São Paulo, and Rio de Janeiro reporting higher average benefit values, likely linked to economic structure and labor market characteristics. However, the high average benefits in less industrialized states such as Bahia and Roraima suggest that local economic and occupational factors also play a significant role.

The age group analysis shows that older beneficiaries receive higher average benefits, likely due to accumulated experience and longer careers. In contrast, younger workers, constrained by labor legislation and limited experience, tend to receive lower benefits. The distribution of benefit types further indicates that most cases are processed through standard administrative procedures, with only a small fraction requiring judicial or exceptional interventions.

The multiple correspondence analysis further enriches these findings by mapping associations between variables such as state, type of illness, and benefit type. While some categories—such as specific states or rare diseases—appear isolated, most variables cluster together, indicating shared risk profiles and demographic characteristics. Notably, specific patterns emerge, such as the association between particular benefit types and certain diseases or regions, underscoring the complexity of occupational health risks and their management.

These findings are significant for several reasons. First, they provide empirical evidence supporting the need for targeted occupational health and safety interventions tailored to the identified risk profiles. Second, they highlight persistent gender and regional disparities that must be addressed to ensure greater equity in social security coverage and workplace protections. Third, the results underscore the importance of considering physical and psychosocial factors in workplace risk assessments and policy design.

Ultimately, the study advances the understanding of occupational accidents' multifaceted nature and social security costs, offering actionable insights for policymakers, employers, and

occupational safety and health (OSH) professionals. By revealing the interplay between demographic, economic, and health-related variables, the research contributes to more effective, data-driven strategies for reducing the incidence and impact of workplace accidents and diseases in Brazil.

## ACKNOWLEDGMENTS

The authors thank the financial support from the National Council for Scientific and Technological Development. (CNPq), Process Number 406387/2021-4.

## REFERENCES

- Bolsoni, L. C. (2024). Análise da disparidade salarial entre gêneros no Brasil: uma abordagem estatística [Master's Thesis]. Instituto de Ciências Matemáticas e de Computação. *Universidade de São Paulo, USP*. <https://doi.org/10.11606/d.55.2024.tde-17032025-192127>
- Botelho, L. H. F., Abrantes, L. A., Ferreira, M. A. M., & Costa, T. de M. T. da. (2018). Análise de fatores influentes na relação entre contribuições e benefícios previdenciários. *Revista de Ciências da Administração*, 1(3), 23-39. <https://doi.org/10.5007/2175-8077.2018v20n52p23>
- Botelho, L. H. F. & Costa, T. D. M. T. da. (2020). Análise financeira da seguridade e previdência social no Brasil em tempos de reformas fiscais. *Revista Catarinense da Ciência Contábil*, 19(1), 2922. <https://doi.org/10.16930/2237-766220202922>
- Brasil. (1991). Lei 8.213, de 24 de julho de 1991. Recovered from [https://www.planalto.gov.br/ccivil\\_03/leis/l8213cons.htm](https://www.planalto.gov.br/ccivil_03/leis/l8213cons.htm)
- Brasil. (2011). Lei 12.527, de 18 de novembro de 2011. Recovered from [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)
- Brasil. (2016). Decreto 8.777, de 11 de maio de 2016. Recovered from <https://legislacao.presidencia.gov.br/atos/?tipo=DEC&numero=8777&ano=2016&ato=c90ATSq1EeZpW T24e>
- Brasil. (2020). Boletim Estatístico da Previdência Social, BEPS (p. 58). *Secretaria da Previdência*. Recovered from [https://www.gov.br/previdencia/pt-br/assuntos/previdencia-social/copy\\_of\\_arquivos/beps122020\\_final-1.pdf](https://www.gov.br/previdencia/pt-br/assuntos/previdencia-social/copy_of_arquivos/beps122020_final-1.pdf)
- Brasil. (2025). SMARTLAB. Observatório digital de saúde e segurança no trabalho. *Ministério Público do Trabalho*. Recovered from <https://smartlabbr.org/sst>
- Cabello, A. F. T., Martínez-Rojas, M., Carrillo-Castrillo, J. A., & Rubio-Romero, J. C. (2021). Occupational accident analysis according to professionals of different construction phases using association rules. *Safety Science*, 144, 105457. <https://doi.org/10.1016/j.ssci.2021.105457>
- Eyerkaufner, M. L., Bonfante, E. A., Dallabona, L. F., & Fabre, V. V. (2019). Simulador de custos para gestão de riscos de acidentes de trabalho. *Revista Catarinense da Ciência Contábil*, 18, 1-16. <https://doi.org/10.16930/2237-766220192753>
- Fernandes, F. T. (2021). Machine learning em saúde e segurança do trabalhador: perspectivas, desafios e aplicações [Doctorate Thesis]. Faculdade de Saúde Pública. *Universidade de São Paulo, USP*. <https://doi.org/10.11606/t.6.2021.tde-27012022-140548>
- Ferrari, G. N., Ossani, P. C., de Souza, R. C. T., Leal, G. C. L., & Galdamez, E. V. C. (2023). Impact of rising temperatures on occupational accidents in Brazil in the period 2006 to 2019: a multiple correspondence analysis. *Safety Science*, 161, 106078. <https://doi.org/10.1016/j.ssci.2023.106078>
- Gouveia, A. G. P., Horsth, T. A., & Faroni, W. (2017). Análise quantílica dos fatores de influência na arrecadação do regime geral de previdência social em Minas Gerais. *Revista Brasileira de Políticas Públicas e Internacionais*, 2(2), 135-151. <https://doi.org/10.22478/ufpb.2525-5584.2017v2n2.35243>
- Hämäläinen, P., Takala, J., & Kiat, T. B. (2017). Global estimates of occupational accidents and work-related illnesses 2017 (p. 21). *WHS Institute*. Recovered from <https://www.icohweb.org/site/images/news/pdf/Report%20Global%20Estimates%20of%20Occupational%20Accidents%20and%20Work-related%20Illnesses%202017%20rev1.pdf>
- Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: a K-Means clustering algorithm. *Applied Statistics*, 28(1), 100. <https://doi.org/10.2307/2346830>
- Health and Safety Executive - HSE. (2020). Costs to Britain of workplace fatalities and self-reported injuries and ill health, 2018/2019. *Health and Safety Executive*. Recovered from <https://www.hse.gov.uk/statistics/>
- International Labour Organization - ILO. (2020). *Quick guide on sources and uses of statistics on occupational safety and health* (p. 58). Recovered from [https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40stat/documents/publication/wcms\\_759401.pdf](https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40stat/documents/publication/wcms_759401.pdf)
- Kakhki, F. D., Freeman, S. A., & Mosher, G. A. (2019). Evaluating machine learning performance in predicting injury severity in agribusiness industries. *Safety Science*, 117, 257-262. <https://doi.org/10.1016/j.ssci.2019.04.026>

- Leigh, J. P. (2011). Economic burden of occupational injury and illness in the United States. *Milbank Quarterly*, 89(4), 728-772. <https://doi.org/10.1111/j.1468-0009.2011.00648.x>
- Madalozzo, R & Artes, R. (2017). Escolhas profissionais e impactos no diferencial salarial entre homens e mulheres. *Cadernos de Pesquisa*, 47(163), 202-221. <https://doi.org/10.1590/198053143666>
- Menegon, L. da S., Menegon, F. A., & Kupek, E. (2021). Mortalidade por acidentes de trabalho no Brasil: análise de tendência temporal, 2006-2015. *Revista Brasileira de Saúde Ocupacional*, 46(e8). <https://doi.org/10.1590/2317-6369000025219>
- Ossani, P. C. & Cirillo, M. A. (2025). Multivariate analysis [R package MVar version 2.2.7]. *R-Project.org*. Recovered from <https://cran.r-project.org/package=MVar>
- Paez, O., Uahinui, T., Genaidy, A., Karwowski, W., Sun, L., & Daraiseh, N. (2006). Estimating uninsured costs of work-related accidents, part II: an incidence-based model. *Theoretical Issues in Ergonomics Science*, 7(3), 247-260. <https://doi.org/10.1080/14639220500090547>
- Perez, G. de S. (2025). Capital humano e desigualdade: um estudo sobre os salários no Distrito Federal [Professional Master's Degree in Economics]. *Universidade de Brasília*, Brasília. Recovered from <http://repositorio.unb.br/handle/10482/52062>
- R Core Team. (2023). A Language and environment for statistical computing. *R Foundation for Statistical Computing*. *R Core Team*. Recovered from <https://www.r-project.org/>
- Rajabi, F., Jahangiri, M., Alimohammadlou, M., & Kamalinia, M. (2025). Identify and classify common errors, antecedents, outcomes, and mitigation strategies in qualitative and semi-quantitative workplace safety risk management: Integrating grounded theory and systematic literature review. *Safety Science*, 187, 106851. <https://doi.org/10.1016/j.ssci.2025.106851>
- Reis, B. L. dos, Rosa, A. C. F. da, Machado, A. de A., Wencel, S. L. S. S., Leal, G. C. L., Galdamez, E. V. C., & Thom de Souza, R. C. (2021). Data mining in occupational safety and health: a systematic mapping and roadmap. *Production Journal*, 31(e20210048). <https://doi.org/10.1590/0103-6513.20210048>
- Rencher, A. C. & Christensen, W. F. (2012). *Methods of multivariate analysis*. Wiley.
- Sá, A. C. M. G. N. de, Gomide, M. H. M., & Sá, A. T. N. de. (2016). Acidentes de trabalho suas repercussões legais, impactos previdenciários e importância da gestão no controle e prevenção: revisão sistemática de literatura. *Revista Médica de Minas Gerais*, 26(e1825). <https://doi.org/10.5935/2238-3182.20160125>
- Sanni-Anibire, M. O., Mahmoud, A. S., Hassanain, M. A., & Salami, B. A. (2020). A risk assessment approach for enhancing construction safety performance. *Safety Science*, 121(2), 15-29. <https://doi.org/10.1016/j.ssci.2019.08.044>
- Santana, V. S., Araújo-Filho, J. B., Albuquerque-Oliveira, P. R., & Barbosa-Branco, A. (2006). Acidentes de trabalho: custos previdenciários e dias de trabalho perdidos. *Revista de Saúde Pública*, 40(6), 1004-1012. <https://doi.org/10.1590/s0034-89102006000700007>
- Sindeaux, R. V., de Souza, R. M. F., Júlia, A. J. V. G., Reis, B. S., Sobral, K. J. B. V., & Martins, K. R. (2025). Salários mais elevados: quem são, onde estão e o que fazem os que ocupam o topo da pirâmide salarial no norte de Minas Gerais. *Revista Foco*, 18(4), e8142-e8142. <https://doi.org/10.54751/revistafoco.v18n4-002>
- Souza, N. M. de. (1990). Análise de correspondência [Undergraduate Final Projects]. Instituto de Matemática. Estatística: Bacharelado. *Universidade Federal do Rio Grande do Sul*. Recovered from <https://lume.ufrgs.br/handle/10183/134295>
- Safe Work Australia - SWA. (2017). The cost of work-related injury and illness for Australian employers, workers and the community 2012-13. *Safe Work Australia*. Recovered from <https://www.safeworkaustralia.gov.au/doc/cost-work-related-injury-and-illness-australian-employers-workers-and-community-2012-13>
- Tenoury, G. N. C. da S., Madalozzo, R. C., & Martins, S. R. (2021). Diferença salarial e taxa de participação no mercado de trabalho brasileiro: uma análise a partir do sexo dos indivíduos. *Estudos Econômicos*, 51(01), 33-72. <https://doi.org/10.1590/0101-41615112grs>
- Tompa, E., Mofidi, A., Heuvel, S. van den, Bree, T. van, Michaelsen, F., Jung, Y., Porsch, L., Emmerik, M. van, IWH (Institute for Work & Health), TNO (The Netherlands Organisation for Applied Scientific Research), & VVA (Valdani Vicari & Associati). (2019). The value of occupational safety and health and the societal costs of work-related injuries and diseases. *European Agency for Safety and Health at Work - EU-OSHA*. <https://doi.org/10.2802/251128>
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn machine learning without learning the machinery. *GetMobile: Mobile Computing and Communications*, 19(1), 29-33. <https://doi.org/10.1145/2786984.2786995>
- Xu, D. & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165-193. <https://doi.org/10.1007/s40745-015-0040-1>
- Zubaroğlu, A. & Atalay, V. (2020). Data stream clustering: a review. *Artificial Intelligence Review*, 54. <https://doi.org/10.1007/s10462-020-09874-x>