



# ConBRepro

X CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



EVENTO  
ON-LINE

02 a 04  
de dezembro 2020

## Doenças e acidentes de trabalho no Brasil: uma análise exploratória de dados

**Beatriz Lavezo dos Reis**

Programa de Pós-Graduação em Engenharia de Produção - UEM

**Camila Rafael**

Departamento de Engenharia de Produção - UEM

**Gislaine Camila Lapasini Leal**

Programa de Pós-Graduação em Engenharia de Produção – UEM

**Rodrigo Clemente Thom de Souza**

Programa de Pós-Graduação em Engenharia de Produção – UEM, UFPR

**Edwin Vladimir Cardoza Galdramez**

Programa de Pós-Graduação em Engenharia de Produção – UEM

**Resumo:** A saúde e segurança ocupacional tem sido tema de alerta tanto na indústria como na academia, devido ao crescente número de casos de doenças e acidentes registrados mundialmente, que estão vinculados às funções executadas pelos trabalhadores. No Brasil, esse avanço nos registros de acidentes e doenças também é constatado e formalizado por meio de alguns conjuntos de dados. Um desses conjuntos são os registros de acidentes e doenças com abertura de Comunicação de Acidentes de Trabalho (CAT), disponibilizados publicamente pelo governo do país, com abrangência nacional. Empregando este banco de dados, o objetivo desta pesquisa é apresentar um panorama geral sobre os casos de acidentes, doenças e óbitos ocupacionais no país, utilizando o horizonte temporal de julho de 2018 a março de 2020. Para atingir esse objetivo foi realizada uma análise descritiva dos 797.688 registros do conjunto de dados, avaliando cada um dos atributos, seus *outliers* e comportamentos. A partir das análises foi possível representar o cenário da saúde e segurança do trabalho no Brasil, assim como, se aprofundar no conjunto de dados da CAT. Os registros de doenças ou acidentes de trabalho foram agrupados em cinco categorias de variáveis e são associados predominantemente ao gênero masculino, atuando na produção de bens e serviços industriais. Com empresas localizadas em sua maioria no estado de São Paulo e capitais do país, executando atividades da indústria de transformação. Com o estudo também foi possível identificar variáveis com um menor nível de confiabilidade, *insights* para estudos futuros e aplicações neste tema.

**Palavras-chave:** Comunicação de acidentes de trabalho, Saúde e segurança do trabalho, Acidentes de trabalho, Doenças ocupacionais.

## Occupational diseases and accidents in Brazil: an exploratory analysis of data

**Abstract:** Occupational safety and health has been an issue of alert both in industry and academia, due to growing number of diseases and accidents cases registered worldwide, which are linked to functions performed by workers. In Brazil, this advance in records of accidents and diseases is also

verified and formalized through some data sets. One of these sets is records of accidents and diseases with the opening of Work Accidents Communication (CAT), made publicly available by the country government, with national coverage. Using this database, objective of research is to present an overview of accidents, diseases and occupational deaths cases in the country, using the time horizon from July 2018 to March 2020. To achieve this objective, a descriptive analysis is performed of 797,688 records in the data set, evaluating each of attributes, their outliers and behaviors. From the analyzes it was possible to understand occupational safety and health scenario in Brazil, as well as to deepen CAT data set. The records of occupational diseases or accidents were grouped into five categories of variables and are predominantly associated with male gender, acting in production of industrial goods and services. With companies located mostly at São Paulo state and capitals carrying out activities in manufacturing industry. With the study it was also possible to find less reliable variables, insights for future studies and applications in this theme.

**Keywords:** Work Accidents Communication, Occupational safety and health, Work accidents, Occupational diseases.

## 1. Introdução

A Organização Internacional do Trabalho (2020) entende como saúde e segurança do trabalho (SST) a proteção dos trabalhadores contra doenças gerais ou profissionais e contra acidentes de trabalho. Há normas e leis que garantem ao trabalhador o direito de ter acesso a um ambiente profissional apropriado, em condições de liberdade, segurança e dignidade. Embora a SST seja um requisito social (RACHID *et al.*, 2015) essas ocorrências ainda representam números significativos. No mundo, de acordo com a OIT, estima-se que um funcionário vem a óbito em decorrência de acidente ou doença ocupacional a cada 15 segundos (WANG *et al.*, 2020). No Brasil, os dados também são alarmantes, segundo o Observatório Digital de Saúde e Segurança no Trabalho ocorrem 150 casos a cada 10 mil trabalhadores. Além disso, o país possui uma média de 2 mil óbitos por ano.

Para diminuir os incidentes é indispensável que as empresas e organizações governamentais se coordenem para lidar com essas ocorrências. Com relação às corporações, é necessário assumir os custos dos acidentes, lidar com perdas na produção e muitas vezes, lidar com prejuízos à reputação da organização (CIARAPICA; GIACCHETTA, 2009). Em relação ao governo, os altos gastos com benefícios concedidos aos trabalhadores que sofreram algum acidente ou doença representam um problema significativo, que exige atenção (COLNAGO; SIVOLELLA, 2019). No Brasil, todas as ocorrências ocupacionais, independente da gravidade, devem ser registradas com a abertura de uma Comunicação de Acidente de Trabalho (CAT), armazenadas e disponibilizadas em bancos de dados de domínio público (BATISTA *et al.*, 2019).

Neste contexto, o presente estudo realizou uma análise exploratória dos dados da CAT entre o período de julho de 2018 a março de 2020, visando apresentar um panorama geral sobre os casos de acidentes, doenças e óbitos ocupacionais no país. Para isso, o estudo está dividido em cinco seções, iniciando com uma introdução ao tema, seguida de uma fundamentação teórica para contextualizar a pesquisa, e também, a metodologia aplicada. A quarta sessão disponibiliza os resultados da pesquisa por meio da análise exploratória e por fim, a pesquisa apresenta suas considerações finais.

## 2. Fundamentação teórica

### 2.1 Saúde e segurança do trabalho

Desde o início da industrialização, a automatização de serviços e avanços tecnológicos gerou grande impacto na sociedade, tornando a vida mais ágil e eficiente (BADRI *et al.*, 2018). No entanto, também elevaram os riscos de acidentes e doenças ocupacionais através do aumento de exposição a maquinários e computadores, e dessa forma, o

ambiente tornou-se propício ao desenvolvimento de doenças cardiovasculares e relacionadas ao estresse (RUSO *et al.*, 2012).

Apesar de existirem normas e leis que fiscalizem os procedimentos de saúde e segurança do trabalho (SST), os acidentes e doenças ocupacionais possuem índices alarmantes de ocorrência, tornando-se um problema enfrentado mundialmente (RACHID *et al.*, 2015). Este impasse tem recebido atenção especial na maioria dos países, já que estimativas nacionais registram despesas substanciais para o governo e empresas (JABBARI; GHORBANI, 2016). Para Sanmiquel *et al.* (2015), não é coincidência que os países com condições de trabalho mais seguras relatam menos acidentes de trabalho, juntamente com melhores percentuais financeiros.

De acordo com uma estimativa de 2017 da OIT, as perdas econômicas relacionadas a acidentes, doenças ocupacionais e óbitos chegaram a US\$ 3 trilhões, o que representa 4% do PIB mundial (CHEN *et al.*, 2020; WANG *et al.*, 2020). Além das perdas financeiras, os acidentes, doenças e óbitos contribuem para o afastamento de milhares de pessoas do mercado de trabalho, evidenciando a urgência em resolver o problema motivado pelo desenvolvimento econômico global (COLNAGO; SIVOLELLA, 2019).

Uma das formas de controlar e evitar os riscos ocupacionais é aderir a uma cultura de segurança no local de trabalho. Para Rachid *et al.* (2015), a cultura de segurança é representada pelo conjunto de práticas desenvolvidas e implementadas pela autoridade da empresa, a fim de melhorar as condições de trabalho dos colaboradores. Este gerenciamento torna possível a fiscalização do uso correto de equipamentos de proteção, por exemplo, diminuindo as chances de ocorrer algum acidente ou contrair alguma doença e evitando possíveis prejuízos (CIARAPICA; GIACCHETTA, 2009).

Outro método de contribuição para a saúde e segurança dos trabalhadores é o estudo de casos que já aconteceram, e que muitas vezes estão armazenados em bancos de dados. Sarkar *et al.* (2019) realizaram um estudo de caso em uma indústria siderúrgica na Índia, e após a coleta de dados por três anos, foi possível realizar um cálculo de previsões de acidentes de trabalho. Já os autores Cheng *et al.* (2010) exploraram dados de acidentes na indústria de construção em Taiwan. Neste estudo, foram aplicadas técnicas de mineração de dados a fim de encontrar padrões e levantar relações de causa-efeito entre os acidentes.

## **2.2 Acidentes de Trabalho no Brasil**

Com os avanços da tecnologia da informação e computadores é possível a coleta, o armazenamento, a transferência e análise de grandes quantidades de dados (RAHMAN *et al.*, 2010). Han e Pei (2012) afirmam que a sociedade está vivendo na era dos dados, já que houve um explosivo aumento no volume de informações armazenadas devido a informatização da sociedade.

Neste contexto, os acidentes, doenças ocupacionais e óbitos produzem um grande volume de dados. Os dados ocupacionais informam não apenas que uma decisão precisa ser tomada, mas também qual decisão deve ser tomada. A análise do histórico de ocorrências ocupacionais é importante pois fornece uma visão direcionada para as atividades de prevenção de acidentes e doenças, podendo ser uma ferramenta poderosa no foco de redução de riscos (ZHANG; JIANG, 2012).

O Brasil ocupa o 4º lugar entre os países que registram o maior número de óbitos durante as atividades laborais, atrás apenas da China, Estados Unidos e Rússia (FERREIRA *et al.*, 2020). Segundo o Observatório de Segurança e Saúde do Trabalho, no período de 2012 a 2018, constatou-se que a cada três horas um trabalhador morre devido a acidentes e doenças ocupacionais. Além disso, estima-se que foram gastos com benefícios concedidos pelo Instituto Nacional do Seguro Social (INSS) mais de 26 bilhões de reais no mesmo período.

Comumente, as ocorrências relacionadas à SST são armazenadas em bancos de dados, podendo ser privados, que correspondem aos registros das próprias empresas, ou de domínio público, que armazenam informações nacionais ou mundiais. Como exemplo, no Brasil, o Departamento de Informática do Sistema Único de Saúde (DATASUS) disponibiliza dados publicamente a respeito de informações epidemiológicas, sanitárias e de origem administrativa dos hospitais do país (SALDANHA *et al.*, 2019).

Além do banco de dados do DATASUS, o Brasil possui um conjunto de dados que agrupa as ocorrências com abertura de Comunicação de Acidentes de Trabalho (CAT). Neste documento, são registrados todos os acidentes, doenças e óbitos que ocorrem no local de trabalho ou trajeto, independente da gravidade. Este registro serve para informar a previdência social do país sobre a ocorrência de algum acontecimento ocupacional, além de ser uma forma de garantir, à vítima ou a seus familiares, os benefícios decorrentes do acontecimento (BATISTA *et al.*, 2019; MINISTÉRIO DA FAZENDA, 2017).

Segundo Pinto (2017), a CAT possui uma limitação de abrangência dos dados, já que os trabalhadores que não estão vinculados ao Regime Geral de Previdência Social (RGPS), como os autônomos, não preenchem o documento. Entretanto, continua sendo uma fonte essencial e de grande abrangência de informações a respeito das ocorrências ocupacionais no país.

### **3. Metodologia**

O estudo foi realizado em três fases, representadas por: planejamento; coleta e pré-processamento dos dados; desenvolvimento e avaliação. No planejamento foi estruturada uma contextualização do tema, buscando na literatura por estudos e referências na área de SST. Ainda nesta fase, foi realizada uma busca por conjuntos de dados, onde determinou-se que seria utilizado o da CAT, devido à sua abrangência nacional e disponibilização pública.

Em seguida iniciou-se a etapa de coleta e pré-processamento dos dados, com a união de todos os conjuntos na mesma base, estudo das variáveis, adaptações e agrupamentos. Esta etapa foi realizada utilizando o software Microsoft Excel<sup>®</sup>. Com os dados estruturados, a última fase foi de desenvolvimento da análise exploratória e avaliação. Nesta etapa, com as variáveis organizadas e auxílio do software Tableau<sup>®</sup>, foram gerados gráficos dos atributos, que possibilitaram discussões e conclusões encontradas nas seções seguintes.

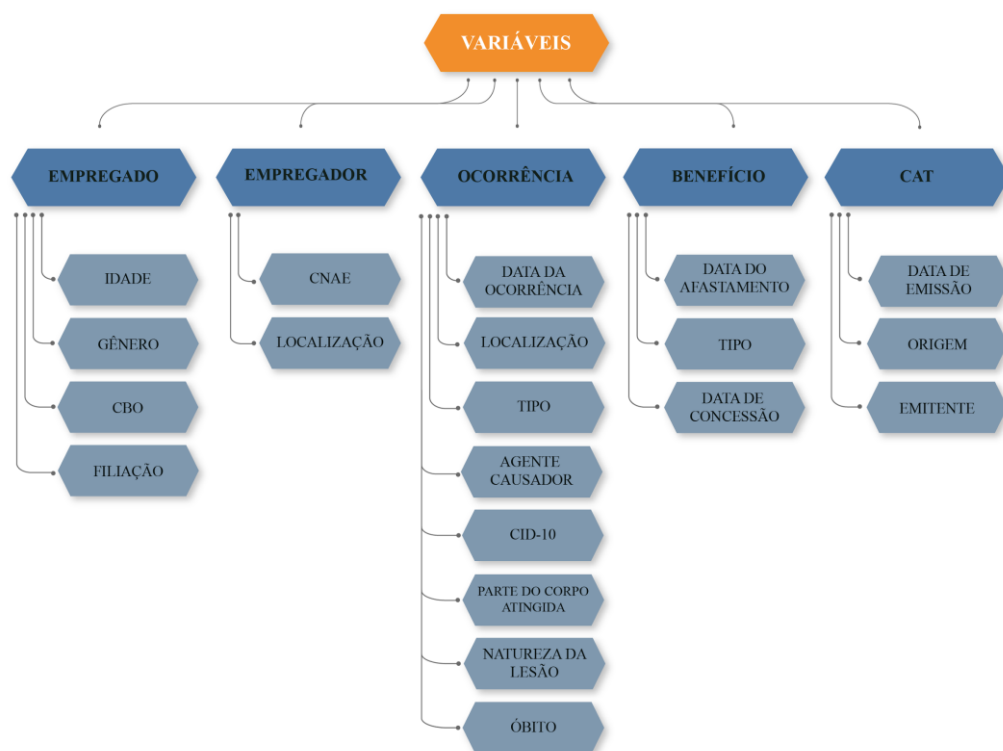
### **4. Resultados e discussões**

#### **4.1 Coleta e pré-processamento dos dados**

A fonte de dados selecionada para a pesquisa é divulgada publicamente por meio do Portal Brasileiro de Dados Abertos, disponibilizada pela Previdência Social do Brasil. Esse conjunto de dados apresenta todas as doenças e acidentes do trabalho que possuem abertura de CAT. As ocorrências são agrupadas por trimestre e disponibilizadas em documentos de formato CSV, do período de julho de 2018 a março de 2020, correspondendo à sete conjuntos de dados disponíveis.

Para o presente estudo, foram selecionados todos os conjuntos de dados divulgados até o momento e agrupados em um único banco. Este apresentava em sua totalidade 797.688 registros de acidentes e doenças, dispostos em 25 colunas. A partir de uma análise inicial dos dados, foi observado que algumas dessas variáveis apresentavam informações repetidas, possibilitando a redução à 20 atributos representados no conjunto de dados selecionado, conforme apresentado pela Figura 1.

Figura 1 – Variáveis do conjunto de dados



Fonte: Autores (2020)

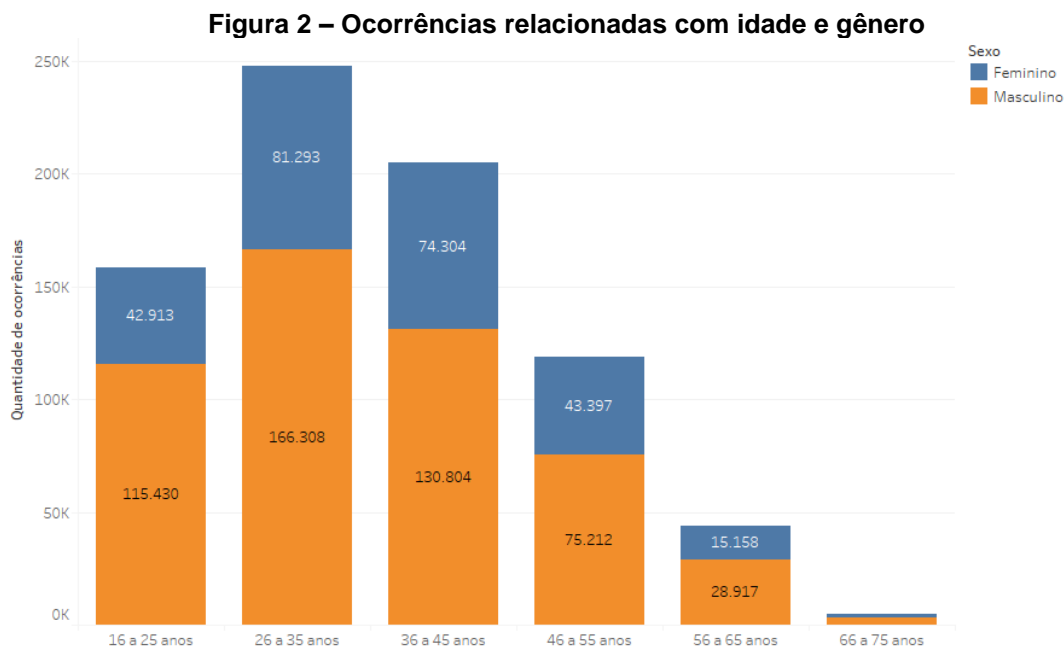
Esses atributos foram divididos em cinco grupos, sendo associados: ao empregado, ao empregador, à ocorrência, ao benefício e à CAT. A iniciativa de dividir as variáveis em conjuntos pode auxiliar na organização e associação dos dados, assim como as discussões e conclusões desenvolvidas ao longo da pesquisa. Em seguida foi realizada uma análise aprofundada para cada um dos grupos e suas respectivas variáveis, buscando entender o cenário de acidentes do trabalho e doenças no Brasil.

## 4.1 Análise exploratória dos dados

### 4.1.1 Variáveis associadas ao empregado

As variáveis que estão associados ao empregado ou vítima do acidente, são: idade, gênero, CBO e filiação. A idade do trabalhador não é uma variável apresentada no conjunto de dados da CAT, mas pode ser calculada a partir dos campos de data de nascimento do trabalhador e data de acidente. Ao realizar esse cálculo foram encontrados valores de idade abaixo do permitido pela lei e outros muito altos, como empregados menores de 10 anos ou maiores de 90. Para reduzir esses *outliers* foram consideradas apenas idades entre 16 e 75 anos. Além disso, foram removidas também datas de nascimento e acidente com valores inválidos, resultando em 537 ocorrências retiradas nesta variável.

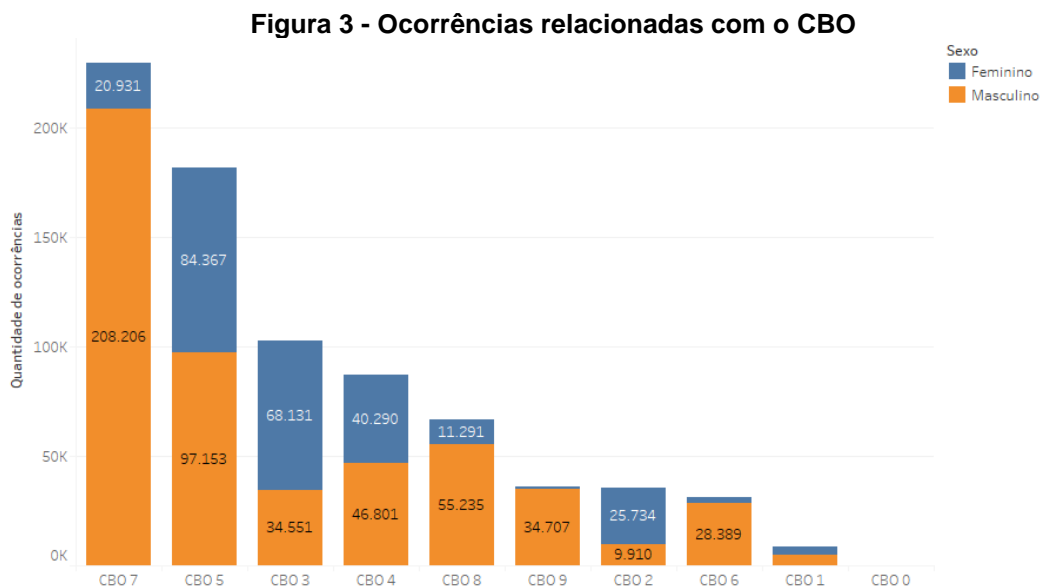
Quanto à variável gênero, os valores são definidos por “feminino” e “masculino”, no entanto alguns registros apresentavam a categoria “Indeterminado”, com 12 registros, e “Não informado” com 293, que também foram considerados *outliers* e retirados. A Figura 2 apresenta uma relação entre a idade e gênero dos trabalhadores e a quantidade de ocorrências registradas com CAT.



**Fonte: Autores (2020)**

O grupo com maior número de registros (31%) são trabalhadores entre a faixa etária de 26 a 35 anos, tanto para o sexo masculino quanto feminino, seguidos pela faixa de 36 a 45 anos (26%). Juntas, as duas faixas etárias correspondem a mais da metade das ocorrências do conjunto de dados.

A Classificação Brasileira de Ocupações (CBO), também é uma variável relacionada ao empregado e descreve qual a profissão associada ao trabalhador no momento do acidente. O CBO apresenta 10 grandes grupos de profissões, que são divididas em mais cinco subníveis. Para esta pesquisa foram analisados apenas os grandes grupos, conforme apresentado pela Figura 3. Dentre os valores da variável CBO, 342 registros não apresentavam classificação.



**Fonte: Autores (2020)**

Analisando as classes de ocupações, a maior responsável por doenças e acidentes do trabalho é o grupo CBO 7, que se caracteriza por “trabalhadores da produção de bens e serviços industriais”, sendo majoritariamente composta por trabalhadores do sexo masculino. As classes CBO 2 e CBO 3 são as únicas que apresentam maior percentual

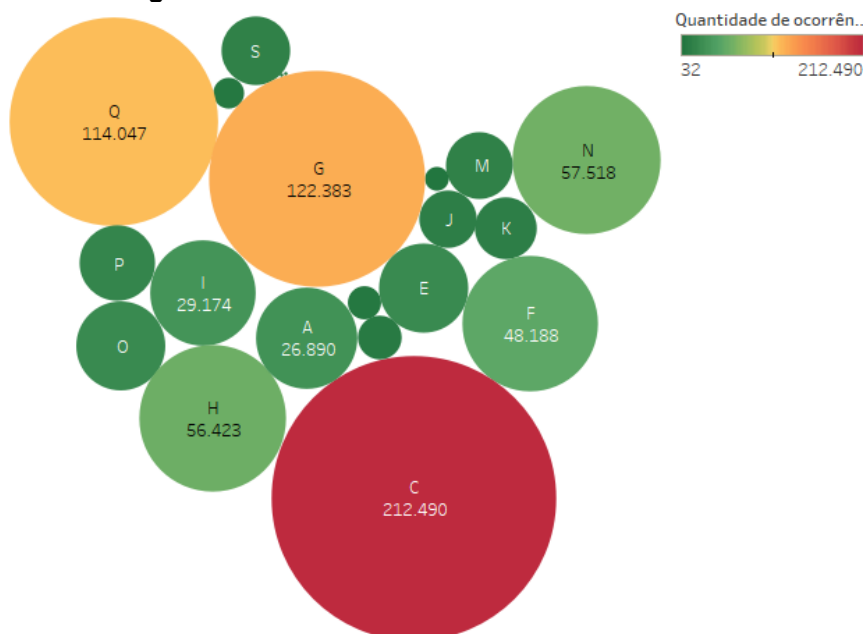
para o sexo feminino e descrevem, respectivamente, “profissionais das ciências e das artes” (66% sexo feminino) e “técnicos de nível médio” (72%).

A última variável associada ao empregado é sua filiação, que retrata se a vítima é “empregado”, “trabalhador avulso” ou “segurado especial”. A categoria “empregado” é responsável por 99,8% dos registros do conjunto de dados, “trabalhador avulso” soma 1.392 casos e, por fim, “segurado especial”, com 95 registros. Para a variável filiação, apenas 12 registros apresentavam resultado “*ñ class*”.

#### 4.1.2 Variáveis associadas ao empregador

As variáveis associadas ao empregador são CNAE e localização. O CNAE representa a Classificação Nacional de Atividades Econômicas, que apresenta 21 grupos para descrever os setores industriais em que as empresas atuam. Esses 21 grupos possuem subníveis, mas para esta pesquisa foram adotados apenas os grandes grupos representados pelas letras A à U, conforme detalhado na Figura 4.

Figura 4 – Ocorrências relacionadas com o CNAE



Fonte: Autores (2020)

O grupo com maior número de registros é o C, com 27% de incidência, representando a classe de “indústria de transformação”. Essa categoria é seguida pelos grupos G e Q, que representam, respectivamente, “comércio, reparação de veículos automotores e motocicletas” e “saúde humana e serviços sociais”. Esses três grupos acumulam juntos mais de 56% dos registros no conjunto de dados. Também são destaque neste atributo os grupos N, H e F, descritos por “Atividades administrativas e serviços complementares”, “Transporte, armazenagem e correio” e “Construção”. Neste atributo foram encontradas 5.798 ocorrências que apresentavam valor “*ñ class*”.

Outra variável associada ao empregador é a localização, expressa pela cidade e estado onde está localizada a sede da empresa. As cinco cidades que são frequentemente associadas às ocorrências são capitais, sendo: São Paulo (80.877), Rio de Janeiro (30.014), Belo Horizonte (16.857), Curitiba (15.696) e Porto Alegre (14.045). Apenas 33 registros não apresentavam classificação para cidade. Quanto ao estado, ocorre uma concentração nos estados das regiões sul e sudeste (Figura 5).

**Figura 5 – Ocorrências relacionadas com o estado do empregador**



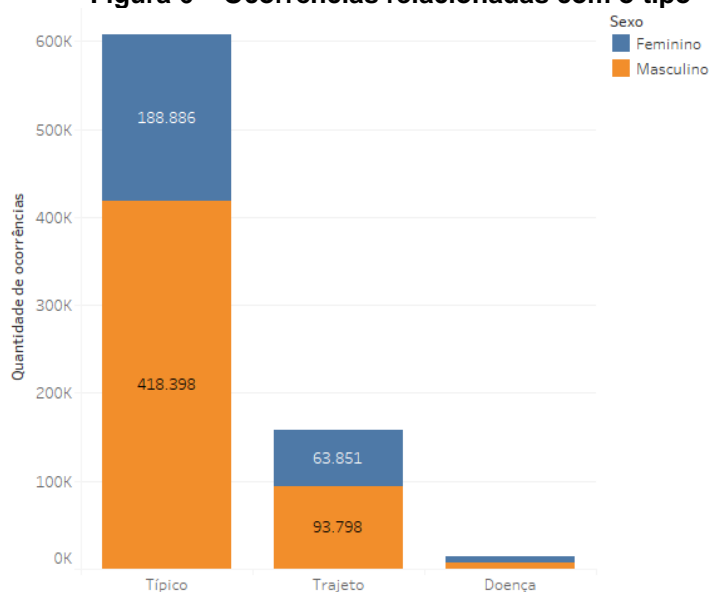
Fonte: Autores (2020)

O maior número de registros está associado a empresas com sede em São Paulo, representando aproximadamente 37% dos casos. Também são destaque os estados de Minas Gerais (9,9%), Rio Grande do Sul (8%), Paraná (7,9%), Rio de Janeiro (6,5%) e Santa Catarina (6%).

#### 4.1.3 Variáveis associadas à ocorrência

No conjunto de variáveis relacionadas a ocorrências estão: data do acidente, localização, tipo, agente causador, CID-10, parte do corpo atingida, natureza da lesão e óbito. Dentre os dados selecionados 3.405 estavam ligados a mortes do trabalhador, equivalente à 0,43% dos dados do conjunto. Em relação à localização, essa variável é descrita pelo estado onde o acidente efetivamente aconteceu. No entanto, mais de 30% dos registros apresentam classificação “*ñ class*” neste aspecto, além de não apresentar registros vinculados aos estados da região sul, sudeste, centro-oeste e ao estado da Bahia. Quanto à variável de tipo da ocorrência, existem três classificações que podem ser atribuídas: acidente típico, acidente de trajeto e doença, como apresentado na Figura 6.

**Figura 6 – Ocorrências relacionadas com o tipo**



Fonte: Autores (2020)



Os acidentes típicos são mais evidentes neste contexto, representando mais de 76% dos casos, seguido por acidentes de trajeto com quase 20%. A doença é a classe com menor número de registros, 5.800 casos para vítimas do sexo feminino e 7.782 masculino. Apenas 16 registros apresentavam resposta como “ignorado”.

Agente causador representa aquilo que foi responsável por provocar o acidente na vítima, podendo ser uma coisa, uma substância ou uma condição ambiental. Segundo as instruções para preenchimento da CAT existem 241 classes para registrar acidentes e mais 58 para doenças. Nesta pesquisa essas classes foram agrupadas em 31 tipos de acidentes e 18 doenças, onde os cinco mais recorrentes de cada tipo são apresentados pela Tabela 1.

**Tabela 1 – Classes de agente causador com maior número de registros**

	<b>Classes agente causador</b>	<b>Quantidade</b>
Acidentes	Veículos e meios de transporte	142916
	Superfícies e equipamentos utilizados para sustentar pessoas	136429
	Máquinas	74306
	Ferramentas manuais sem força motriz	71732
	Metal e produtos metálicos	41349
Doenças	Esforço excessivo	5450
	Movimentos involuntários	3247
	Ataque de ser vivo	1411
	Atrito ou abrasão	509
	Impacto	389

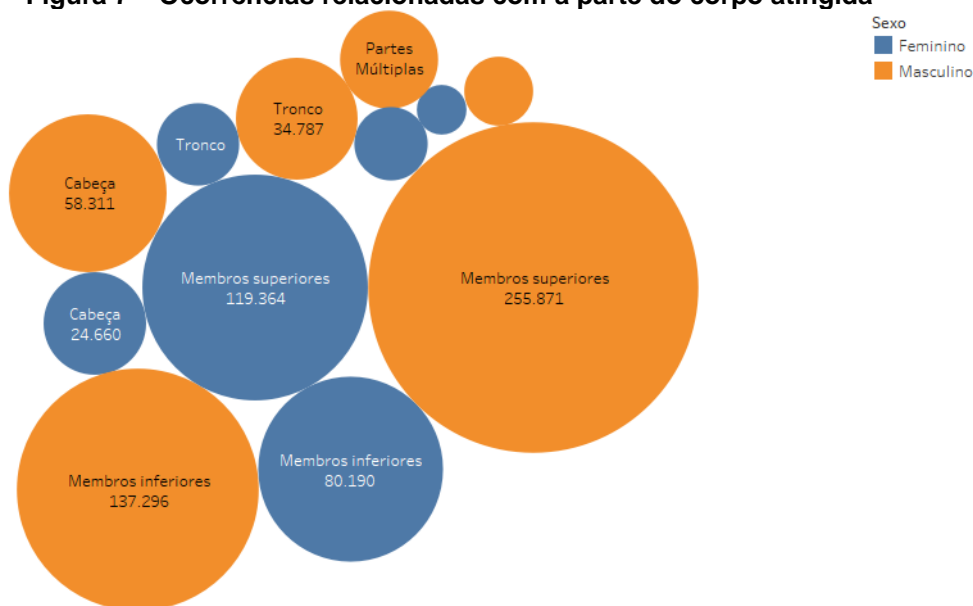
**Fonte: Autores (2020)**

Considerando as classes de agente causador para acidentes, os veículos e meios de transporte ocupam quase 19% dos registros. Em relação às doenças, o esforço excessivo é maior agente causador, acumulando 40% dos casos. Para esta variável os *outliers* foram 3.416 registros categorizados como “*ñ class*”.

A variável CID-10 apresenta a classificação de doenças e lesões e é dividida em 22 capítulos. A maior parte dos capítulos apresenta poucos registros, que se acumulam, principalmente no capítulo XIX. Este, representa as ocorrências associadas a “lesões, envenenamentos e algumas outras consequências de causas externas” e é responsável por quase 76% dos dados. Também são representativos no CID-10 os capítulos XIII (52.299), XXI (47.531) e XX (47.142), descritos, respectivamente por “Sintomas, sinais e achados anormais de exames clínicos e de laboratório”, “Fatores que influenciam o estado de saúde e o contato com os serviços de saúde” e “Causas externas de morbidade e de mortalidade”. Nesta variável houve grande incidência de valores não classificados, onde 4.273 eram “brancos” e 4.632 “*ñ class*”, somando 8.905 *outliers*.

Em relação a parte do corpo atingida, o conjunto de dados apresenta 41 classificações, que foram agrupadas em seis divisões (Figura 7): cabeça, membros superiores, membros inferiores, tronco, partes múltiplas e sistemas e aparelhos.

**Figura 7 – Ocorrências relacionadas com a parte do corpo atingida**



**Fonte: Autores (2020)**

O grupo mais atingido, tanto em homens quanto mulheres são os membros superiores (47%), seguidos por membros inferiores (27%). Neste atributo 147 registros foram considerados *outliers*, pois apresentavam respostas “*ñ class*” (108) e “Localizacao da lesão” (39). A última variável relacionada à ocorrência é a natureza da lesão, expressa pela forma com que o trabalhador foi lesionado. O conjunto de dados original apresentava 28 classificações, que foram agrupadas em 23 a fim de facilitar a análise. Dentre essas classes de lesões a mais recorrente é “Corte, laceração, ferida contusa e punctura” (20,6%), seguida de “Fratura” (16,5%), “Contusão e esmagamento” (14,2%) e “Lesões” (14,1%).

#### **4.1.4 Variáveis associadas ao benefício**

Os atributos associados ao benefício são: data do afastamento, tipo e data da concessão, no entanto, poucas informações são apresentadas no conjunto de dados associadas a essas variáveis. Em relação a data de concessão do benefício expedido, apenas 1,46% das respostas são diferentes de zero, e para o tipo de benefício o mesmo percentual é diferente de “Pa”. Considerando a data do afastamento, as informações são apresentadas em função do mês e ano do acontecimento, o que impede uma análise mais aprofundada. Dessa forma, a análise realizada foi em função de quantos meses deu-se o afastamento do indivíduo após seu acidente, onde 20% não se afastou após a ocorrência, 79% se afastou no mesmo mês e o restante distribuído no prazo de até um ano após (1%).

#### **4.1.5 Variáveis associadas à CAT**

No último grupo de atributos estão aqueles associados à CAT, que são: data de emissão, origem e emitente. Considerando a data de emissão e a data do acidente, é possível identificar quanto tempo foi transcorrido desde a ocorrência até a notificação formalizada. Neste aspecto, mais de 96% dos registros foram notificados no mesmo mês de ocorrência do acidente, 2,5% no mês seguinte e o restante (1,5%) no período de até um ano após o acidente.

Considerando a origem do cadastramento da CAT, as opções são “internet” ou “prisma”, porém, apenas 3 registros têm como fonte a segunda opção. Por fim, quanto ao emitente, a CAT por ser emitida por: empregador; médico; segurado ou seu dependente; sindicato; e autoridade pública. Dentre essas possibilidades, o empregador é o maior responsável pelas emissões de CAT, em mais de 98,5% dos casos.

## 5. Considerações finais

A análise realizada no conjunto de 797.688 registros de acidentes e doenças com abertura de CAT, possibilitou não só o entendimento do panorama geral de saúde e segurança do trabalho no Brasil, mas também a avaliação de cada variável. Deste estudo foi identificado que a maioria dos acidentes e doenças acontecem com trabalhadores do sexo masculino (67%), na faixa etária de 26 a 45 anos (57%). A maior parte dos trabalhadores está associada ao setor de indústrias de transformação (27%), com as ocorrências principalmente classificadas como lesões, envenenamentos e algumas outras consequências de causas externas (76%).

Estas e as demais conclusões da pesquisa, podem ser utilizadas por organizações privadas e governamentais para auxílio na tomada de decisão. As informações podem servir de subsídio para criação de novas normas e políticas para mitigação de acidentes e doenças ocupacionais. Também indicam qual setor industrial, região do país ou faixa etária requerem maior atenção, auxiliando na criação de programas e investimento em prevenção.

Também foi possível identificar com esse estudo algumas variáveis não confiáveis no conjunto de dados, sendo: estado de ocorrência do acidente, variáveis associadas ao benefício e a CAT. A localização do acidente, descrita pelo estado onde ocorreu, apresentou mais de 30% dos seus valores sem classificação e não relacionava nenhum registro com o estado da Bahia e regiões sul, sudeste e centro-oeste. Os atributos ligados ao benefício foram de difícil avaliação e aplicabilidade, pois apresentavam poucos valores diferentes de zero (1,46%) em seus registros. De maneira análoga as variáveis relacionadas a CAT não apresentavam muita variação, chegando à apenas três registros classificados como “prisma” na origem de cadastramento.

Além disso, a partir da análise exploratória foi possível remover os *outliers* dos dados, ou seja, valores atípicos, diferentes do padrão de resposta. Essa remoção resultou em 778.475 registros finais, preparados para aplicações. Como sugestões para pesquisas futuras estão estudos que se aprofundem nas variáveis menos confiáveis do modelo, buscando entender seu comportamento. Também como recomendação, realizar aplicações estatísticas ou de mineração de dados, buscando prever comportamentos futuros e tendências no cenário de SST no Brasil.

### Agradecimentos

Este estudo foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código Financeiro 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq).

### Referências

BADRI, A.; BOUDREAU-TRUDEL, B.; SOUISSI, A. S. Occupational health and safety in the industry 4.0 era: A cause for major concern?. **Safety Science**, v.109, p. 403-411, 2018.

BATISTA, A. G.; SANTANA, V. S.; FERRITE, S. Registro de dados sobre acidentes de trabalho fatais em sistemas de informação no Brasil. **Revista Ciência e Saúde Coletiva**, v.24, p. 693-704, 2019.

CHEN, H.; HOU, C.; ZHANG, L.; LI, S. Comparative study on the strands of research on the governance model of international occupational safety and health issues. **Safety Science**, v.22, p. 104513, 2020.

CHENG C.W.; LIN C.C.; LEU S.S. Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry. **Safety Science**, v.48, p. 436–444, 2010.

CIARAPICA, F.E.; GIACCHETTA, G. Classification and prediction of occupational injury risk using soft computing techniques: An Italian study. **Safety Science**, v.47, p. 36–49, 2009.

COLNAGO, L. M.; SIVOLELLA, R. F. Convenção 187 da OIT: promoção da saúde e segurança do trabalho no Brasil e a viabilidade de sua ratificação. **Juslaboris**, 2019.

HAN, J; KAMBER, M.; PEI, J. Data Mining Concepts and Techniques. **Morgan Kaufmann Publishers**. Third Edition, 2012.

JABBARI, M.; GHORBANI, R. Developing techniques for cause-responsibility analysis of occupational accidents. **Accident Analysis and Prevention**, v.96, p. 101-107, 2016.

MINISTÉRIO DA FAZENDA. Anuário Estatístico de Acidentes do Trabalho, v.1, p.1-996 2017.

PINTO, J. M. Tendência na incidência de acidentes e doenças de trabalho no Brasil: aplicação do filtro Hodrick-Prescott. **Revista Brasileira de Saúde Ocupacional**, v.42, p. 1-12, 2017.

RACHID, C.; ION, V.; IRINA, C.; MOHAMED, B. Preserving and improving the safety and health at work: Case of Hamma Bouziane cement plant (Algeria). **Safety Science**, v.76, p. 145–150, 2015.

RAHMAN, S; AHSAN FERROZ, M. A; KAMRUZZAMAN, M; FARUQUE, N. M. Analyze Database Optimization Techniques. **International Journal of Computer Science and Network Security**, v.10, n.8, p. 275-279, 2010.

RUSO, J.; STOJANOVIĆ, V. Occupational health and safety using data mining. **International Journal for Quality Research**, v.6, n.4, p. 355-364, 2012.

SALDANHA, R. F; BASTOS, R. B; BARCELLOS, C. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). **Revista Ciência e Saúde Coletiva**, v.35, p. e00032419, 2019.

SANMIQUEL, L; ROSSELL, J. M; VINTRÓ, C. Study of Spanish mining accidents using data mining techniques. **Safety Science**, v.75, p. 49–55, 2015.

SARKAR, S.; LODHI, V.; MAITI, J. Text-clustering based deep neural network for prediction of occupational accident risk: A case study. In: **2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)**. IEEE, p. 1-6, 2018.

WANG, Y.; HONG, C.; BEI, L.; MENGHUA, Y.; QIANYI, L. A Systematic Review on the research progress and evolving trends of occupational health and safety management: a bibliometric analysis of mapping knowledge domains. **Frontiers in Public Health**, v.8, p. 81, 2020.

ZHANG, D; JIANG, K. Application of Data Mining Techniques in the Analysis of Fire Incidents. **Procedia Engineering**, v.43, p. 250–256, 2012.