SECRETARIA DE SEGURANÇA PÚBLICA
UNIVERSIDADE ESTADUAL DE GOIÁS – UEG
COORDENADORIA DE ENSINO – COE
COORDENAÇÃO DE ENSINO PRESENCIAL E DE PÓS-GRADUAÇÃO
ESPECIALIZAÇÃO EM ALTOS ESTUDOS DE SEGURANÇA PÚBLICA

THYAGO SOUSA MENDES

DIMENSIONAMENTO DA FORÇA DE TRABALHO PARA A SEÇÃO ESPECIALIZADA EM PERÍCIAS DE CRIMES CONTRA O PATRIMÔNIO DA POLÍCIA CIENTÍFICA DO ESTADO DE GOIÁS: MODELAGEM COM SÉRIES TEMPORAIS E AGRUPAMENTO

GOIÂNIA – GO

2025

THYAGO SOUSA MENDES

DIMENSIONAMENTO DA FORÇA DE TRABALHO PARA A SEÇÃO ESPECIALIZADA EM PERÍCIAS DE CRIMES CONTRA O PATRIMÔNIO DA POLÍCIA CIENTÍFICA DO ESTADO DE GOIÁS: MODELAGEM COM SÉRIES TEMPORAIS E AGRUPAMENTO

Artigo apresentado como exigência parcial para conclusão do Curso de Especialização em Altos Estudos de Segurança Pública - CAESP, pela Secretaria de Segurança Pública do Estado de Goiás - SSP e pela Universidade Estadual de Goiás - UEG, sob a orientação do(a) Prof(a). Me. Celso Faria de Souza.

GOIÂNIA – GO

2025

DIMENSIONAMENTO DA FORÇA DE TRABALHO PARA A SEÇÃO ESPECIALIZADA EM PERÍCIAS DE CRIMES CONTRA O PATRIMÔNIO DA POLÍCIA CIENTÍFICA DO ESTADO DE GOIÁS: MODELAGEM COM SÉRIES TEMPORAIS E AGRUPAMENTO

SIZING THE WORKFORCE FOR THE SECTION SPECIALIZED IN EXPERTISE OF CRIMES AGAINST PROPERTY OF THE SCIENTIFIC POLICE OF THE STATE OF GOIÁS: MODELING WITH TIME SERIES AND GROUPING

Aluno (a): Thyago Sousa Mendes l[1*]
Orientador (a): Celso Faria de Souza [2**]

Resumo: Este estudo desenvolve uma metodologia para dimensionar a força de trabalho da Seção Especializada em Perícias de Crimes Contra o Patrimônio da Polícia Científica de Goiás, com base em modelagem estatística. O problema central reside na alocação inadequada de peritos, fundamentada em critérios históricos e dissociada de previsões que considerem a sazonalidade e variabilidade da demanda. O objetivo é propor um modelo orientado por dados, utilizando séries temporais e análise de agrupamento para prever solicitações de perícias e classificar a carga de trabalho por plantão. A metodologia adotada é quantitativa, aplicada e exploratória-descritiva, empregando registros institucionais de ocorrências patrimoniais entre janeiro de 2020 e maio de 2025. Utilizou-se a modelagem GAMLSS, com distribuição Binomial Negativa, para prever a demanda diária, incorporando sazonalidade semanal e autocorrelação. Em paralelo, o algoritmo *k-means* foi aplicado a dados simulados computacionalmente para uma pesquisa de esforço percebido segundo a escala NASA-TLX. Os resultados do trabalho identificaram padrões temporais, como maior demanda às segundas-feiras e picos em janeiro, fevereiro e novembro. O agrupamento definiu três regimes de carga laboral. Como produto, propõe-se um indicador de alocação baseado na razão entre demanda prevista e número ideal de atendimentos, otimizando recursos e mitigando sobrecargas. Conclui-se que a abordagem substitui critérios subjetivos por uma base empírica, sendo replicável em outros contextos periciais.
Palavras-chave: Dimensionamento da força de trabalho; Perícia criminal; Crimes contra a propriedade; Séries temporais; Análise de agrupamentos.

Abstract: This study develops a methodology to size the workforce of the Specialized Section for Expertise of Crimes Against Property of the Scientific Police of Goiás, based on statistical modeling. The central problem lies in the inadequate allocation of experts, based on historical criteria and dissociated from forecasts that consider the seasonality and variability of demand. The objective is to propose a data-driven model, using time series and cluster analysis to predict requests for expert assessments and classify the workload by shift. The adopted methodology is

[1]* Perito Criminal da Polícia Científica de Goiás.
[2]** Formação acadêmica e vínculo institucional do orientador, indicando titulação, área de atuação, cargo ocupado e instituição de origem.

quantitative, applied and exploratory-descriptive, using institutional records of property occurrences between January 2020 and May 2025. GAMLSS modeling, with Negative Binomial distribution, was used to predict daily demand, incorporating weekly seasonality and autocorrelation. In parallel, the k-means algorithm was applied to computationally simulated data for a survey of perceived effort according to the NASA-TLX scale. The results of the study identified temporal patterns, such as higher demand on Mondays and peaks in January, February and November. The grouping defined three workload regimes. As a product, an allocation indicator is proposed based on the ratio between expected demand and the ideal number of appointments, optimizing resources and mitigating overloads. It is concluded that the approach replaces subjective criteria with an empirical basis, being replicable in other expert contexts.
Keywords: Workforce sizing; Criminal expertise; Property crimes; Time series; Cluster analysis.

1. INTRODUÇÃO

A administração pública enfrenta desafios crescentes na alocação eficiente de força de trabalho. Na segurança pública, esse desafio é agravado por demandas operacionais elevadas, limitação de efetivo e necessidade de respostas ágeis e qualificadas.

Em Goiás, crimes patrimoniais (furtos, roubos e arrombamentos) representaram 92% das ocorrências em 2024[3] (SSP-GO, 2025), gerando não apenas prejuízos materiais, mas também impactos intangíveis como sensação de insegurança (PLASSA; CUNHA, 2016) com custos estimados em 5-10% do PIB nacional (CERQUEIRA et al., 2007).

Nesse contexto, a atuação da Seção Especializada em Perícias de Crimes Contra o Patrimônio da Polícia Científica de Goiás (SEPPATRI) na Região Metropolitana de Goiânia depende criticamente da distribuição adequada da força de trabalho. Contudo, a alocação atual ancora-se em critérios históricos e administrativos, desvinculados de análises preditivas.

A perícia criminal desempenha papel estratégico no sistema de justiça, fornecendo base técnica para elucidação de crimes e responsabilização penal, legitimando decisões judiciais e fortalecendo a confiança social (RODRIGUES et al., 2018; PRASAD, 2023).

Diante do exposto, este estudo tem como objetivo geral desenvolver uma metodologia de Dimensionamento da Força de Trabalho (DFT) baseada em dados, utilizando modelos de séries temporais e técnicas de agrupamento para prever a demanda pericial e classificar a carga de

[3] Número inferido das estatísticas de 16 tipos de crimes monitorados pela SSP-GO, cujos dados são disponibilizados no sítio https://goias.gov.br/seguranca/estatisticas/. Data de acesso: 21/06/2025.

trabalho por plantão, respectivamente. A proposta busca corrigir falhas do modelo atual, promovendo escalas mais racionais, evitando sobrecargas e fortalecendo a eficiência da perícia criminal como instrumento de justiça (DIAS et al., 2013; ROCHA et al., 2023).

Os objetivos específicos da pesquisa são: i) realizar análise exploratória das solicitações de perícia; ii) ajustar um modelo preditivo para esses dados; iii) aplicar técnica de agrupamento para identificar o número de ocorrências compatível com um nível razoável de esforço físico e mental dos peritos da SEPPATRI; e iv) propor um indicador para orientar a alocação de servidores por plantão.

A relevância científica deste trabalho está na adaptação de metodologias consolidadas em outras áreas ao contexto da perícia criminal, ainda pouco explorado sob a ótica da gestão da força de trabalho. Destacam-se o uso do modelo de regressão GAMLSS para prever o número de solicitações (STASINOPOULOS; RIGBY, 2007) e do algoritmo *k-means* para caracterizar plantões com base nas ocorrências e no esforço exigido dos peritos (KASSAMBARA, 2017).

Esta é uma pesquisa aplicada, documental, com abordagem quantitativa, caráter exploratório e descritivo, fundamentada no método hipotético-dedutivo. Parte de duas hipóteses centrais: (H1) a demanda por perícias apresenta um padrão sistemático ao longo do tempo, passível de modelagem; e (H2) há uma relação estatisticamente modelável entre o número de atendimentos por perito e a carga de trabalho percebida subjetivamente.

A justificativa da pesquisa está ancorada na relevância institucional de otimizar os recursos humanos da perícia, preservar a qualidade de vida dos servidores ao se evitar sobrecarga de trabalho, bem como na necessidade de promover maior eficiência.

Os dados analisados foram extraídos de registros institucionais da Polícia Científica, abrangendo atendimentos entre janeiro de 2020 e maio de 2025, em 38 cidades goianas. O período foi selecionado por oferecer uma amostra ampla, representativa e também atual.

Além de oferecer um modelo técnico de alocação alinhado à realidade da SEPPATRI, espera-se que os resultados sejam replicáveis e sirvam de base para aprimorar políticas públicas da Polícia Científica de Goiás, fortalecendo a parceria entre SSP-GO e UEG.

Este trabalho está estruturado da seguinte forma: i) o Capítulo 2 apresenta a revisão bibliográfica; ii) o Capítulo 3 descreve a metodologia adotada; iii) o Capítulo 4 expõe os resultados e suas interpretações; e iv) o Capítulo 5 traz a síntese final e as conclusões.

2. REVISÃO DA LITERATURA

Esta Seção apresenta uma análise crítica da literatura teórica e empírica sobre o tema, com base em fontes acadêmicas consolidadas.

2.1. Crimes Patrimoniais, Teoria Criminológica e o Papel da Perícia Criminal na Justiça

Crimes contra o patrimônio, como furtos, roubos e arrombamentos, representam uma das tipologias mais recorrentes no cenário brasileiro. A Teoria das Atividades Rotineiras, proposta por Cohen e Felson (1979), fornece uma estrutura analítica para a compreensão desses delitos, sugerindo que a convergência entre um infrator motivado, um alvo desprotegido e a ausência de vigilância são condições propícias à ocorrência do crime. Esse modelo se aplica com particular precisão a centros urbanos, como a Região Metropolitana de Goiânia, onde a circulação de pessoas e bens amplia o risco de vitimização.

A atuação pericial assume papel central na produção de provas técnicas e na legitimação do sistema de justiça penal. Para Bitencourt (2012), a perícia criminal substitui a subjetividade do testemunho por dados empíricos que resistem ao contraditório. A imparcialidade dos peritos e a adoção de tecnologias modernas fortalecem a credibilidade das investigações, como destaca Prasad (2023).

2.2. Modelagem com Séries Temporais no Contexto Criminal

Chen, Yuan e Shu (2008) utilizaram o modelo ARIMA (*AutoRegressive Integrated Moving Average*) para prever crimes patrimoniais em uma cidade chinesa, com base em dados semanais de 50 semanas. O modelo foi treinado para estimar a criminalidade na 51ª semana e superou os métodos de suavização exponencial simples e de Holt.

Wardhani et al. (2020) realizaram um estudo sobre a previsão de criminalidade por furto na região da polícia distrital de Surabaya (na ilha de Java na Indonésia), utilizando dois modelos

de séries temporais para dados de contagem: o modelo Poisson-GARMA (*Poisson Generalized Autoregressive Moving Average*) e o modelo Binomial Negativa-GARMA (*Negative Binomial Generalized Autoregressive Moving Average*). O objetivo foi comparar a capacidade preditiva dos dois modelos nos dados criminais. Os autores realizaram previsões para 5 meses adiante!

Cortes et al. (2018) desenvolveram um índice geral de criminalidade para os municípios do Rio Grande do Sul com base em dados de sete tipos de crimes reportados pela Secretaria de Segurança Pública, oferecendo uma métrica interpretável para análise e gestão da segurança pública.

Ainda é possível utilizar modelos mais gerais no contexto criminal que permitem analisar tendências, sazonalidades e realizar previsões de demanda, como é o caso do modelo GAMLSS (*Generalized Additive Models for Location, Scale and Shape*), que destaca-se por sua flexibilidade em modelar simultaneamente múltiplos parâmetros da distribuição de probabilidade da variável resposta (STASINOPOULOS; RIGBY, 2007; RIGBY et al., 2019). O modelo GAMLSS utilizado neste trabalho é dado por:

$$Y_t \sim \text{Binomial Negativa}(\mu_t, \sigma), \quad log(\mu_t) = \beta_0 + \gamma t + \sum_{j=1}^{6} \beta_j D_{j,t} + \sum_{k=1}^{q} \alpha_k y_{t-k} \qquad (1),$$

sendo $D_{j,t}$ variáveis *dummies* que representam o dia da semana, y_{t-k} são valores da variável resposta com *delay*, t é o tempo e σ, γ, β_0, β_j e α_k são parâmetros do modelo.

2.3. Análise de Agrupamento

A análise de agrupamento é uma técnica exploratória da estatística que visa identificar grupos homogêneos em um conjunto, onde as observações de um mesmo grupo são mais semelhantes entre si do que em relação a outros (HAIR et al., 2009).

No estudo de Costa e Silva (2022), o algoritmo *k-means* foi aplicado para agrupar municípios de Pernambuco com perfis criminais semelhantes, visando apoiar políticas públicas de segurança mais direcionadas.

2.4. Métodos de Dimensionamento da Força de Trabalho

A discussão sobre o dimensionamento da força de trabalho no setor público é recorrente na literatura. Destacam-se métodos baseados em dados históricos, percepção de esforço e tempo de atendimento. A Escala de Borg (BORG, 1982) é amplamente usada para mensurar percepção subjetiva de esforço, enquanto a Teoria das Filas fornece bases para análise de capacidade e tempo médio de espera (OLIVEIRA; BIANCHINI; ABBADE, 2007).

Outros métodos de dimensionamento também se destacam. O método de Gaidzinski, voltado à enfermagem, baseia-se no tempo médio de cuidado por paciente, jornada efetiva e taxa de absenteísmo (VIANNA et al., 2013). Modelos de Programação Inteira são aplicados em setores com rotinas padronizadas, buscando otimizar a alocação de recursos considerando custos e produtividade (ROSA; FILHO, 2008). Para o planejamento macro, o método das Componentes Demográficas e Goic estima estoques futuros de profissionais com base em taxas de formação, migração e aposentadoria (RODRIGUES, 2008). O método WISN (*Workload Indicators of Staffing Need*), proposto pela OMS, calcula a necessidade ideal de pessoal com base na carga de trabalho e tempo disponível (WHO, 2010).

3. METODOLOGIA

Nesta Seção são apresentados o tipo de estudo, a abordagem utilizada, os métodos de coleta e análise dos dados.

3.1. Classificação da Abordagem e da Natureza da Pesquisa

Essa pesquisa possui natureza quantitativa, pois, de acordo com Günther (2006), essa modalidade caracteriza-se pela objetividade e estruturação metodológica, utilizando métodos padronizados como questionários fechados, experimentos controlados e análises estatísticas para testar hipóteses.

Por sua vez, trata-se de uma pesquisa de natureza aplicada, conforme as classificações propostas por Gil (2002) e Lakatos e Marconi (2003). Pois seu objetivo é aplicar modelos estatísticos já consolidados para desenvolver um método de dimensionamento de força de trabalho.

3.2. Classificação dos Objetivos e Procedimentos Técnicos Utilizados

Quanto aos objetivos, esta pesquisa é classificada como exploratória e descritiva, conforme Lakatos e Marconi (2003). É exploratória por empregar métodos inovadores, como a análise de séries temporais e de agrupamento no dimensionamento da força de trabalho da perícia criminal. É também descritiva ao aplicar um modelo preditivo para investigar relações entre variáveis observáveis e compreender padrões da demanda pericial.

Do ponto de vista dos procedimentos técnicos, trata-se de uma pesquisa documental, segundo a tipologia de Severino (2014), uma vez que se baseia em registros institucionais da Polícia Científica de Goiás. Esses dados, extraídos de bancos de ocorrências periciais, não haviam sido previamente analisados sob a perspectiva adotada neste estudo, o que reforça seu caráter documental.

3.3. Etapas da Pesquisa

A pesquisa será estruturada em três etapas principais: i) coleta e limpeza dos dados; ii) análise exploratória e modelagem da série temporal; iii) análise de agrupamento para levantamento do número de solicitações associadas a um nível de esforço moderado dos servidores lotados na SEPPATRI; e iv) dimensionamento da força de trabalho, por meio da construção de uma métrica de alocação.

3.3.1. Coleta e Limpeza dos Dados

Foi realizado um levantamento documental dos registros de solicitações periciais atendidas pela SEPPATRI no período compreendido entre os dias 01/01/2020 e 01/05/2025. A coleta foi feita por meio de solicitação formal à Gerência de Inovação da Secretaria de Segurança Pública de Goiás. Os dados apresentam as seguintes variáveis: i) número da ocorrência; ii) ano da ocorrência; iii) RAI (Registro de Atendimento Integrado); iv) data do fato; v) data da solicitação; vi) perito responsável; vii) cidade do fato; viii) solicitante (delegacia afeta); e ix) naturezas (tipificações preliminares do delito).

A limpeza dos dados envolveu a padronização de datas e nomes de variáveis. Foram removidos registros com dados faltantes em campos essenciais, como data da solicitação e

número da ocorrência. Também foram eliminadas duplicidades e registros fora do período analisado (anteriores a 01/01/2020 ou posteriores a 01/05/2025). Ao final, obteve-se um conjunto de dados limpo e organizado, pronto para as análises estatísticas.

3.3.2. Análise Exploratória para Modelagem da Série Temporal

Esta etapa contempla a análise exploratória da série temporal referente ao número diário de solicitações de perícias criminais atendidas pela SEPPATRI. Inicialmente, serão elaboradas representações gráficas — como gráficos de linha e *boxplots*, por dia da semana e por mês — com o objetivo de identificar padrões de tendência, sazonalidade, variabilidade e eventuais anomalias. Serão também calculadas estatísticas descritivas, como média, desvio padrão e variância.

A dependência temporal e a estabilidade da série serão examinadas por meio das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), bem como pelo gráfico de controle CUSUM, visando detectar a estrutura temporal e possíveis mudanças de regime.

Com base nos padrões observados, serão considerados modelos estatísticos adequados à natureza discreta dos dados. Será ainda investigada a distribuição de probabilidade mais apropriada para os dados, com ênfase nas distribuições Poisson e Binomial Negativa, observando-se a presença de sobredispersão (dispersão dos dados é menor do que a esperada pelo modelo) ou superdispersão (dispersão dos dados é maior do que a esperada pelo modelo) .

A escolha do modelo final será guiada pela comparação de critérios de informação, pela significância estatística dos parâmetros estimados e pela verificação dos pressupostos do modelo.

3.3.3. Análise de Agrupamento

Para a adequada alocação da força de trabalho, é essencial avaliar a carga de trabalho direcionada ao servidor. Com base nos métodos analisados na revisão bibliográfica, compreende-se que a natureza do trabalho pericial exige a consideração conjunta do esforço físico, mental e emocional. Por isso, é necessário estabelecer uma relação estatística entre a carga de trabalho e o número de ocorrências atendidas, de modo a evitar que o servidor ultrapasse, de forma recorrente, o limite do razoável, comprometendo sua saúde e sua performance.

A proposta metodológica deste trabalho consiste em definir um modelo estatístico F, por:

$$Carga\ de\ Trabalho\ =\ F(Número\ de\ Atendimentos\ por\ Perito\ e\ Por\ Plantão), \quad (2)$$

$$Número\ Ótimo\ de\ Atendimentos\ por\ Perito\ =\ F^{-1}(Carga\ de\ Trabalho\ Moderada\ ) \quad (3).$$

Para investigar a carga de trabalho, é necessário realizar uma pesquisa de opinião com os trabalhadores lotados na SEPPATRI, utilizando uma escala validada para esse fim, como a NASA-TLX (*Task Load Index*). Esta (ver ANEXO A) é uma das ferramentas mais utilizadas para esse tipo de avaliação, por se tratar de um instrumento subjetivo que considera diferentes dimensões do esforço exigido: demanda mental, demanda física, demanda temporal, desempenho percebido, esforço e nível de frustração (HART; STAVELAND, 1988).

Como não foi possível aplicar a escala NASA-TLX no prazo disponível, optou-se por utilizar dados simulados computacionalmente (ver ANEXO B). Cada ponto simulado representa um plantão, em que o eixo x corresponde ao número de solicitações atendidas por perito e o eixo y ao score de carga de trabalho percebida (0 a 100).

Com esses dados, aplicou-se a seguinte abordagem: primeiro, utilizou-se o algoritmo k-means para identificar três grupos de plantões (com o intuito de representar diferentes regimes, excluir outliers e mitigar o risco moral[4]). Em seguida, ajustou-se um modelo de regressão aos pontos do grupo associado a uma carga de trabalho moderada. Por fim, a partir da curva ajustada, identificou-se o número de ocorrências correspondente ao score 50 da escala NASA-TLX, estimando-se assim o valor ideal de solicitações por perito e por plantão.

3.3.4. Dimensionamento da Força de Trabalho

Ao concluir as etapas descritas nas subseções 3.3.1, 3.3.2 e 3.3.3, terá sido ajustado um modelo de regressão capaz de fornecer previsões do número de solicitações periciais para um horizonte temporal futuro (por exemplo, o dia seguinte). Além da estimativa pontual, o modelo permitirá a construção de um intervalo de confiança, definido por limites inferior e superior, que expressam a incerteza associada à predição. Os parâmetros estimados também possibilitam a identificação de tendências na série histórica, permitindo compreender se há crescimento, estabilidade ou redução na demanda ao longo do tempo, elemento fundamental para o

[4] Risco moral é a possibilidade de que um agente aja de forma estratégica ou irresponsável quando não assume integralmente as consequências de suas ações.

planejamento estratégico da força de trabalho. Paralelamente, por meio da análise de agrupamento, serão definidos valores de referência que caracterizam o esforço considerado moderado. A partir desses elementos (predição futura e limiar operacional) será construído um indicador de apoio à alocação de peritos por plantão. Este indicador é definido pela Equação 4:

$$Indicador\ de\ Alocação\ = \frac{Previsão\ da\ Demanda}{Número\ Ótimo\ de\ Atendimentos\ por\ Perito\ e\ Por\ Plantão} \quad (4).$$

3.4. Instrumentos de coleta e processamento

As etapas de manipulação, visualização, modelagem e análise dos dados foram realizadas integralmente na linguagem R (R CORE TEAM, 2024), versão 4.2.

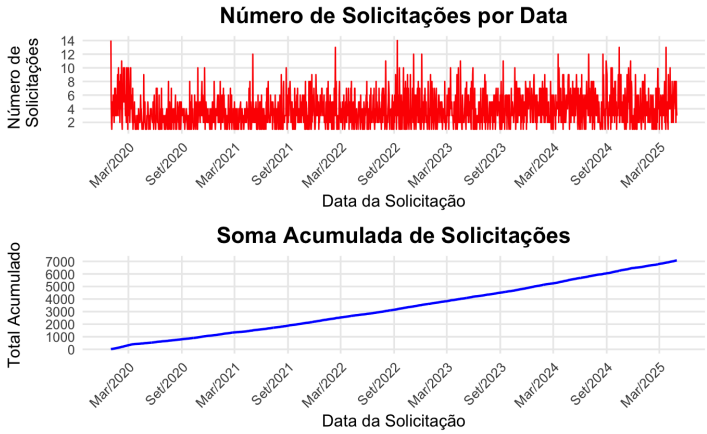
4. RESULTADOS E DISCUSSÕES

Esta Seção apresenta os principais resultados obtidos no estudo, analisando-os de forma crítica à luz da fundamentação teórica (ver Seção 2).

4.1. Análise Exploratória

A Figura 1 apresenta dois gráficos com as solicitações de perícias recebidas pela SEPPATRI. O gráfico superior mostra o número diário de solicitações, evidenciando oscilações regulares com média estável e variância aparentemente constante. O gráfico inferior exibe o total acumulado no tempo, indicando uma tendência linear de crescimento, o que reflete a constância da demanda ao longo do período analisado.

Figura 1 – Dados completos solicitações por dia (parte superior) e total de ocorrências acumuladas (parte inferior)

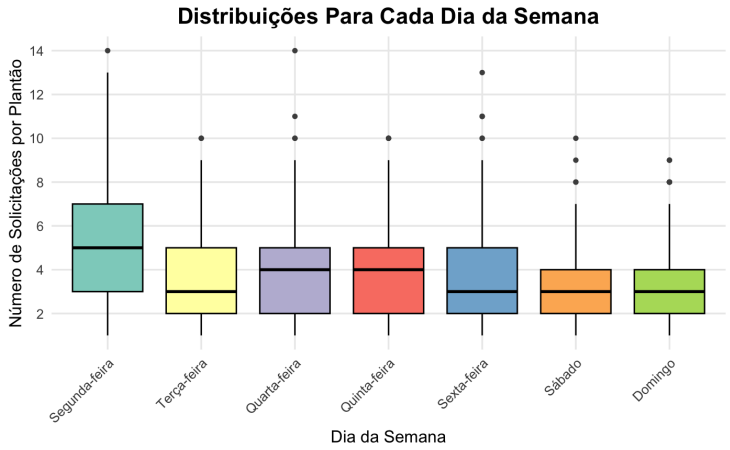


Fonte: Elaborado pelo autor (2025)

4.1.1. Distribuição Por Dia da Semana e Por Mês

Na Figura 2 temos as distribuições, representadas por *boxplots*, por dia da semana. Nota-se que a segunda-feira apresenta maior variabilidade e mediana. A Tabela 1 confirma o destaque das segundas-feiras (maior média aritmética de ocorrências e maior desvio padrão). Este padrão pode ser explicado pelo fato de que muitas vítimas de crimes patrimoniais ocorridos nos finais de semana decidem procurar uma delegacia na segunda-feira, seja por desconhecerem a existência de centrais de plantão, seja por comodidade.

Figura 2 – *Boxplots* representando as distribuições do número de solicitações de perícias por plantão para cada dia da semana



Fonte: Elaborado pelo autor (2025)

O teste de Wilcoxon-Mann-Whitney (com ajuste para comparações múltiplas) revelou diferenças significativas ($p < 0,05$) entre os seguintes pares de dias da semana: i) segundas-feiras versus todos os outros dias; ii) terças-feiras versus sábado e domingo; iii) quartas-feiras versus sábado e domingo; iv) quintas-feiras versus sábado; e v) sextas-feiras versus sábado.

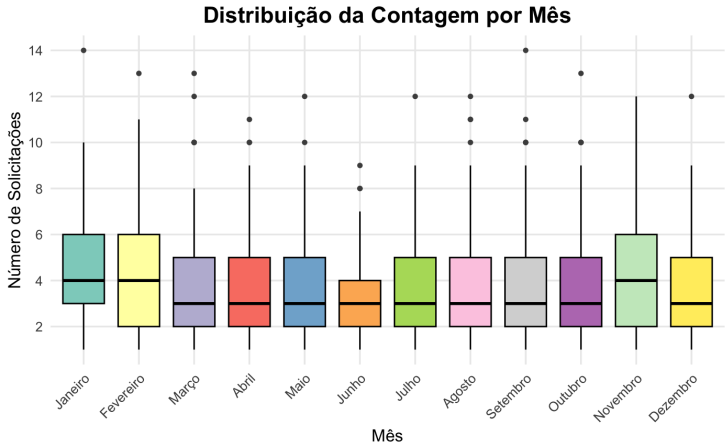
Tabela 1 – Média, desvio padrão e número total de ocorrências por dia da semana

Dia da Semana	Média	Desvio Padrão	Nº de Ocorrências
Segunda-feira	5,25	2,70	1455
Terça-feira	3,74	2,03	992
Quarta-feira	3,80	2,15	1010
Quinta-feira	3,92	2,04	1016
Sexta-feira	3,83	2,37	991
Sábado	3,04	1,75	778
Domingo	3,20	1,90	834

Fonte: Elaborado pelo autor (2025).

A Figura 3 exibe a distribuição das ocorrências por mês, destacando janeiro, fevereiro e novembro com as maiores medianas. A Tabela 2 reforça esse padrão, apontando maiores médias nesses mesmos meses, enquanto junho apresenta os menores valores médio e de dispersão. Apesar disso, o teste de Wilcoxon-Mann-Whitney indicou diferença estatisticamente significativa apenas entre janeiro e junho ($p < 0,05$).

Figura 3 – *Boxplots* representando as distribuições do número de solicitações de perícias por plantão para cada mês do ano



Fonte: Elaborado pelo autor (2025)

Tabela 2 - Média, Desvio Padrão e Contagem de Solicitações por Mês

Mês	Média	Desvio Padrão	Nº de Ocorrências
Janeiro	4,16	2,14	749
Fevereiro	4,15	2,51	660
Março	3,74	2,23	655
Abril	3,89	2,40	658
Maio	3,66	2,12	527
Junho	3,31	1,93	470
Julho	3,72	2,10	536
Agosto	3,57	2,25	528
Setembro	3,94	2,41	556
Outubro	3,96	2,32	598
Novembro	4,14	4,14	749
Dezembro	3,71	3,71	660

Fonte: Elaborado pelo autor (2025).

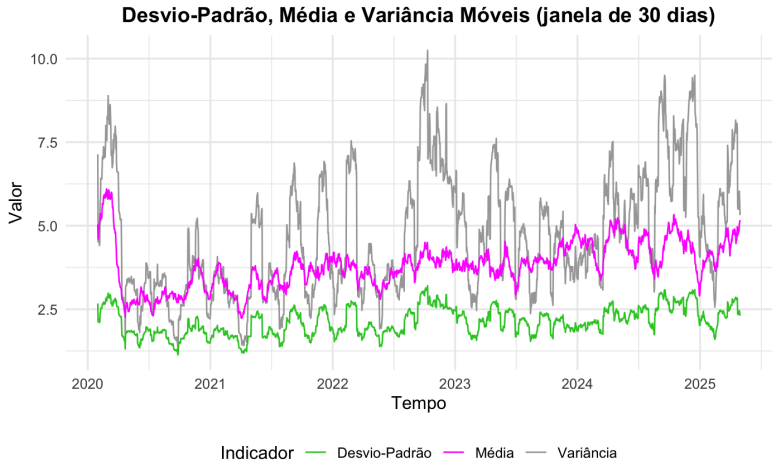
Uma possível explicação para esse comportamento é que em janeiro e fevereiro há o aumento da circulação de pessoas devido às férias escolares, que favorece crimes como furtos. Em novembro, as compras de fim de ano elevam os roubos a lojas e cargas. Já em junho, a

criminalidade cai, pois o frio e as férias de inverno reduzem a movimentação nas ruas, diminuindo oportunidades para crimes.

4.1.2. Análise da Tendência e da Sazonalidade

Na Figura 4 temos os gráficos da média aritmética, desvio padrão e variância móveis, com período de 30 dias. Observamos: i) sazonalidade; ii) tendência crescente na média móvel (não estacionariedade); iii) variância maior que média (sinal de superdispersão); e iv) ausência de tendência no desvio-padrão.

Figura 4 – Média, desvio-padrão e variância móveis em janelas de 30 dias



Fonte: Elaborado pelo autor (2025)

Embora no projeto de pesquisa tenha se falado em empregar os modelos PAR (*Poisson AutoRegressive Model*) ou INGARCH (*Integer-valued Generalized AutoRegressive Conditional Heteroskedasticity*), a Figura 5 aponta para a necessidade da utilização de uma classe mais geral de modelos, pois temos evidência de não estacionariedade, superdispersão e sazonalidade Deste modo, será empregado o modelo GAMLSS descrito em 2.2.

Para investigar a dependência temporal, foram ajustados dois modelos GAMLSS: um nulo (apenas com interceptos) e outro incluindo o tempo como covariável da média. O segundo modelo apresentou AIC inferior ao primeiro, com diferença de 44,889 unidades, indicando melhor ajuste. Além disso, o coeficiente associado ao tempo mostrou-se altamente significativo (p = 8,95e-12), confirmando a influência temporal sobre a média. De forma semelhante, foi analisada a variância. Não foi identificada evidência de que esta depende do tempo, pois a diferença entre AIC foi de apenas 1,726 unidades e a estimativa do coeficiente do tempo não se mostrou significativa (p = 0,667) .

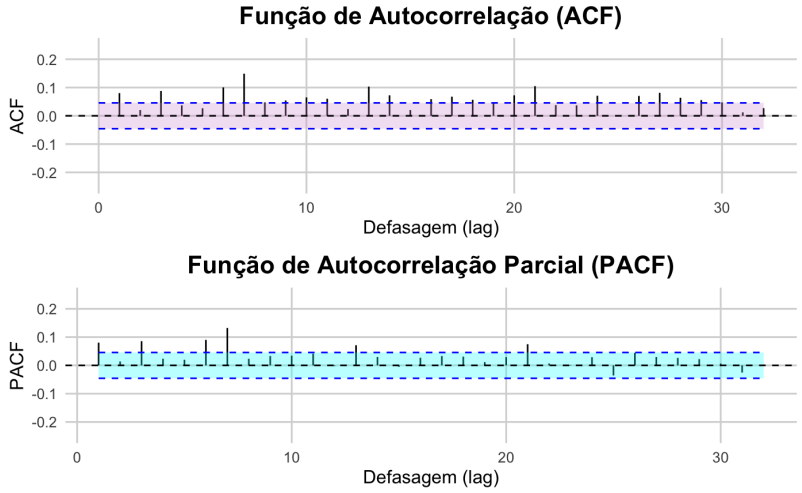
Para investigar padrões sazonais na média da variável resposta, foram consideradas como covariáveis o dia da semana e o mês do ano. Quatro modelos GAMLSS foram ajustados e comparados: (i) modelo nulo (apenas intercepto); (ii) com efeito do dia da semana; (iii) com efeito do mês; e (iv) modelo completo, com ambos os efeitos simultaneamente. A comparação entre os modelos revelou uma diferença significativa no AIC entre o modelo nulo (i) e o modelo completo (ii) (redução de 156,769 unidades). Adicionalmente, o intercepto, todos os dias da semana e 4 meses do ano mostraram significância estatística ($p < 0,05$).

Na análise da variância, a inclusão do dia da semana como covariável resultou em uma redução modesta do AIC (10,18 unidades). Diante da pequena melhora no ajuste, optou-se por um modelo mais simples, mantendo apenas o intercepto na estrutura da variância.

4.1.3. Análise de Termos Auto-regressivos e de Média Móvel

A Figura 5 apresenta, na parte superior, a Função de Autocorrelação (ACF), indicando correlação entre a série e suas defasagens. Embora alguns lags ultrapassem levemente o intervalo de confiança de 95%, os valores são baixos (< 0,1), sugerindo ausência de autocorrelação significativa. Na parte inferior, a Função de Autocorrelação Parcial (PACF) também não revela lags com correlações expressivas, exceto por um ou dois lags. Conclui-se que a série apresenta, no máximo, uma fraca estrutura AR ou MA de ordem 1 ou 2.

Figura 5 – Funções de autocorrelação e de autocorrelação parcial



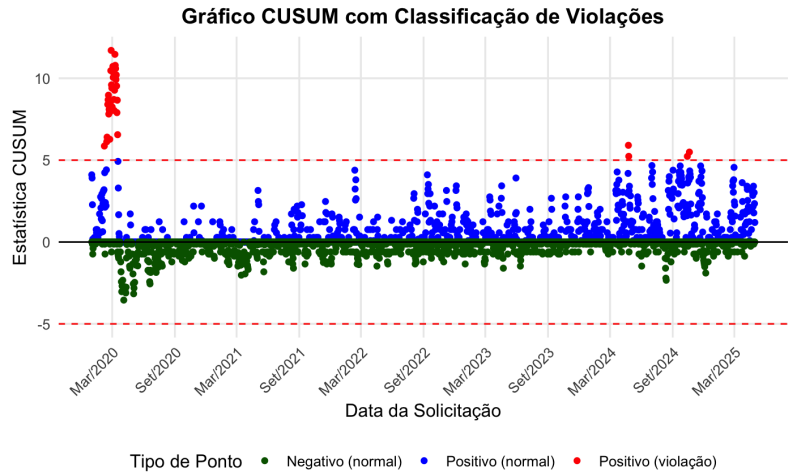
Fonte: Elaborado pelo autor (2025)

4.1.4. Análise da Estacionariedade

Até aqui foram reunidas evidências de que a série de dados utilizada no estudo não é estacionária, pois a distribuição da variável resposta depende do dia da semana (Seção 4.1.1) e demonstramos que a média depende do tempo (Seção 4.1.2). Para complementar, foi aplicado um teste não paramétrico de controle estaístico: Gráfico de Controle Cusum.

A Figura 6 exibe a soma acumulada das diferenças entre os valores observados e a média do processo em função dos dias. A linha preta indica a média esperada, e as linhas vermelhas tracejadas delimitam os limites de controle. Pontos vermelhos sinalizam violações desses limites, concentradas especialmente no início da série e no ano de 2024. Tais desvios sugerem instabilidades associadas a mudanças na demanda, rotinas operacionais ou impactos externos, como a pandemia da Covid-19.O padrão observado confirma que a série não é estacionária, apresentando variações da média e da variabilidade ao longo do tempo.

Figura 6 – Gráfico de controle da soma cumulativa



Fonte: Elaborado pelo autor (2025)

4.2. Ajuste do Modelo Final e Predição

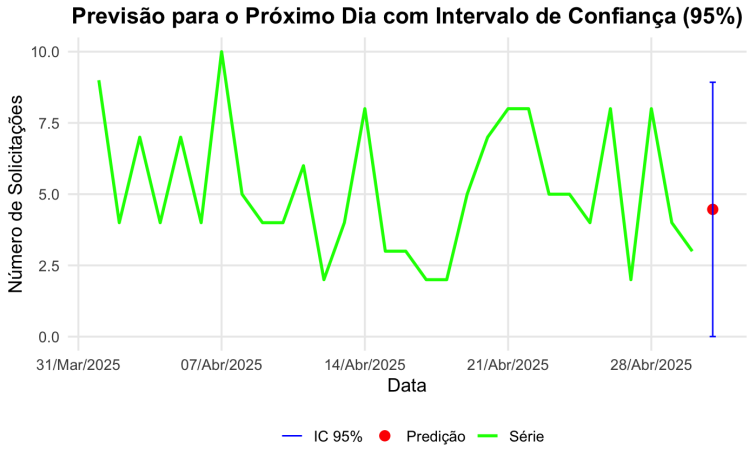
Da Seção 4.1 foi inferida uma estrutura para o modelo a ser ajustado aos dados, que corresponde a Equação 1 da Seção 2.2.

Foram testados modelos com *lag* de 1 a 5, combinados com o dia da semana. O melhor resultado obtido foi para *lag* 1. Na Tabela 3 temos as estimativas de todos os parâmetros do modelo. Observe que todos são altamente significativos! Isso se dá devido ao fato da estrutura do modelo ter sido investigada antes do ajuste final. O parâmetro γ, embora significativo, apresentou

um valor muito baixo e os coeficientes dos dias da semana negativos, mostrando um padrão de redução da média em relação às segundas-feiras.

Para concluir esta Seção, apresenta-se a predição do número de solicitações de perícias para o dia seguinte ao final da série (dia 02/05/2025). A Figura 7 mostra o resultado: temos as solicitações dos 30 dias anteriores à referida data (linha em verde limão), a predição representada pelo ponto em vermelho (4,47 solicitações) e o intervalo de confiança em azul (de 0 a 13,4 solicitações). Observe que a previsão apresentou incerteza compatível com a variabilidade da série no período.

Figura 7 – Gráfico da predição para o final da série e do intervalo de confiança



Fonte: Elaborado pelo autor (2025)

Tabela 3 - Coeficientes do Modelo GAMLSS com *Lag* 1 e Efeito de Dia da Semana

Variável	Estimativa	Erro Padrão	t-valor	p-valor
Intercepto	1,436e+00	3,976e-02	36,129	< 2e-16
lag 1	2,061e-02	5,984e-03	3,444	0,000586
Terça-feira	-3,856e-01	4,615e-02	-8,355	< 2e-16
Quarta-feira	-3,478e-01	4,452e-02	-7,813	9,35e-15
Quinta-feira	-3,082e-01	4,426e-02	-6,964	4,61e-12
Sexta-feira	-3,328e-01	4,477e-02	-7,433	1,62e-13
Sábado	-5,599e-01	4,766e-02	-11,749	< 2e-16
Domingo	-4,974e-01	4,653e-02	-10,690	< 2e-16
log(sigma)	-3,3231	0,2617	-12,7	<2e-16
t	1.659e-04	2,437e-05	6,806	1,35e-11

Fonte: Output da biblioteca gamlss (2025).

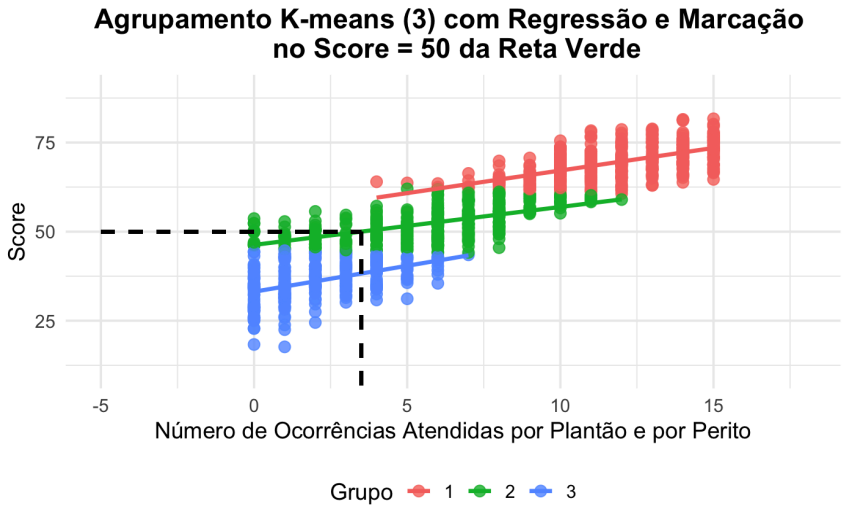
4.3. Análise de Agrupamento em Dados Simulados

A Figura 8 mostra os resultados da aplicação do algoritmo *k-means* sobre dados simulados de carga de trabalho de peritos criminais, considerando o número de ocorrências

atendidas por plantão e o respectivo *score* de esforço. A análise identificou três grupos distintos, cada um refletindo um padrão específico de regime de plantão em termos de volume de atendimentos e percepção de esforço.

O grupo 2 (verde) representa plantões com carga moderada, caracterizados por um número mediano de ocorrências e uma relação linear entre número de ocorrências e esforço percebido. A linha de regressão desse grupo cruza o valor de *score* igual a 50 (referência de esforço considerado aceitável) aproximadamente na marca de três atendimentos por plantão, sugerindo que esse é um ponto de equilíbrio operacional para esse regime. O grupo 1 (azul) agrupa plantões com baixa carga de trabalho, com poucos atendimentos e níveis reduzidos de esforço relatado, sendo compatível com escalas mais leves ou com menor pressão operacional. Já o grupo 3 (vermellho) representa plantões de alta demanda, nos quais há maior número de ocorrências e *scores* de esforço mais elevados, indicando um ambiente potencialmente mais exigente e com maior risco de sobrecarga.

Figura 8 – Gráfico de agrupamentos formados com o uso do *kmeans*



Fonte: Elaborado pelo autor (2025)

A segmentação em três grupos ajuda a mitigar o risco moral da pesquisa, ao limitar o impacto de servidores que, por diferentes motivos, possam superestimar ou subestimar subjetivamente sua carga de trabalho. Além disso, ao permitir a análise de padrões internos a cada grupo, torna-se possível detectar casos fora da curva, como servidores que, mesmo em regimes leves, relatam níveis elevados de esforço, sinalizando possíveis condições individuais de saúde física ou mental que merecem atenção e acompanhamento pelo setor responsável.

5. CONSIDERAÇÕES FINAIS

Este estudo propôs e validou uma metodologia quantitativa para o dimensionamento da força de trabalho na SEPPATRI, integrando modelagem preditiva de séries temporais e análise de agrupamento. Os resultados alcançados atendem plenamente aos objetivos propostos:

1. Análise exploratória: identificou-se variação significativa na demanda por perícias conforme o dia da semana (segundas-feiras com maior média: 5,25 ocorrências) e sazonalidade mensal (picos em janeiro, fevereiro e novembro). A série histórica exibiu tendência crescente linear, superdispersão e não estacionariedade, justificando o uso de modelos avançados.
2. Modelagem preditiva: O modelo GAMLSS com distribuição Binomial Negativa, incorporando *lag* de 1 dia e efeitos de dia da semana, demonstrou alta eficácia (todos os coeficientes estatisticamente significativos, $p < 2e\text{-}16$). A previsão para 2 de maio de 2025 foi de 4,47 ocorrências (IC 95%: 0–13,4), alinhando-se à variabilidade histórica.
3. Definição de carga de trabalho ideal: a análise de *k-means* em dados simulados da escala NASA-TLX revelou três regimes de plantão: Grupo 1 (plantão leve), Grupo 2 (plantão moderado) e Grupo 3 (plantão intenso).
4. Indicador de alocação: Propõe-se a métrica dada pela Equação 4 da Seção 3.3.4.
5. Relevância operacional: o modelo substitui critérios históricos por uma abordagem *data-driven*, reduzindo riscos de sobrecarga e subutilização da força de trabalho.
6. Aplicabilidade: a metodologia é replicável em outras seções periciais, com adaptações mínimas aos dados locais.

Quanto às perspectivas futuras, estudos subsequentes poderiam validar a relação carga de trabalho x ocorrências com dados reais da NASA-TLX.

Em síntese, esta pesquisa oferece um marco metodológico para a gestão científica de recursos humanos na segurança pública, fortalecendo a parceria entre a academia (UEG) e instituições (SSP-GO).

REFERÊNCIAS

BITENCOURT, Cezar Roberto. *Tratado de direito penal: parte especial*. 17. ed. São Paulo: Saraiva, 2012.

BORG, Gunnar A. V. Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, v. 14, p. 377–381, 1982.
CERQUEIRA, Daniel. *Caminhos para o estudo da criminalidade no Brasil*. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, 2017.

COHEN, Lawrence E.; FELSON, Marcus. Social change and crime rate trends: a routine activity approach. *American Sociological Review*, v. 44, n. 4, p. 588-608, 1979.

COSTA, J. C. O. R.; SILVA, M. M. A clustering‑based approach for identifying groups of municipalities to support the direction of public security policies. *Pesquisa Operacional*, v. 42, 2022.

DIAS, R. P.; PEREIRA, A.; LANGARO, F.; CORREA, R. N.; SOUZA, N. de; LACERDA, L. L. V. de. Riscos psicossociais e estresse ocupacional, parceiros numa relação presumida com Burnout: um estudo de estressores que envolvem as atividades dos Peritos Criminais. *Revista Brasileira de Criminalística*, [s. l.], v. 2, n. 1, p. 42–50, 2013.

GIL, Antonio Carlos. *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Atlas, 2002.

GIL, Antonio Carlos. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2008.

GOVERNO DE GOIÁS. *Estatísticas da segurança pública*. Goiânia: Governo de Goiás, 2025. Disponível em: https://goias.gov.br/seguranca/estatisticas/. Acesso em: 16 maio 2025.

GROSS, Donald; HARRIS, Carl M. *Fundamentals of queueing theory*. 3. ed. New York: Wiley, 1998.

GÜNTHER, Hartmut. Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão? *Psicologia: Teoria e Pesquisa*, Brasília, v. 22, n. 2, p. 201–210, maio/ago. 2006.

HAIR, J. F. et al. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009.

HART, S. G.; STAVELAND, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: HANCOCK, P. A.; MESHKATI, N. (ed.). *Human mental workload*. Amsterdam: North-Holland, 1988. p. 139–183.

KASSAMBARA, Alboukadel. *Practical guide to cluster analysis in R: unsupervised machine learning*. 1. ed. [S.l.]: STHDA, 2017.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas, 2003.

MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, v. 1, p. 281–297, 1967.

OLIVEIRA, A. G.; BIANCHINI, D.; ABBADE, M. L. F. Métricas para dimensionar recursos humanos nos Centros de Operações de Rede. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS, 25., 2007, Belo Horizonte. *Anais...* Belo Horizonte: SBC, 2007. p. 1-12.

PLASSA, W.; CUNHA, M. S. Sensação de insegurança pública no Brasil: análise estrutural das vulnerabilidades e do efeito da vitimização direta. *Economic Analysis of Law Review*, Brasília, v. 7, n. 1, p. 266-290, jan./jun. 2016.

PRASAD, A. Importance of forensic science in investigation. *Journal of Police and Criminal Psychology*, v. 38, p. 766-773, 2023.

R CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2024.

ROCHA, Thiago Almeida da et al**.** Dimensionamento da força de trabalho no serviço público: como incorporar competências? *R. Adm. FACES Journal*, Belo Horizonte, v. 22, n. 1, p. 110–133, jan./mar. 2023.

RODRIGUES, Cláudio Vilela; SILVA, Márcia Terra da; TRUZZI, Oswaldo Mário Serra**.** Perícia criminal: uma abordagem de serviços. *Gestão & Produção*, São Carlos, v. 25, n. 4, p. 791–804, out./dez. 2018.

RODRIGUES, F. G. Médicos em Minas Gerais: projeções para o período 2010-2020. Dissertação (Mestrado em Economia). Universidade Federal de Minas Gerais, 2008.Disponível

ROSA, B. A.; FILHO, E. M. S. Dimensionamento de recursos humanos em uma empresa fabricante de materiais sanitários: uma abordagem via programação inteira. In: *Simpósio de Pesquisa Operacional e Logística da Marinha*, 11., 2008.

STASINOPOULOS, D. M.; RIGBY, R. A. Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, [s. l.], v. 23, n. 7, p. 1–46, 2007.

VIANNA, C. M. M. et al. Modelos econométricos de estimativa da força de trabalho: uma revisão integrativa da literatura. *Physis: Revista de Saúde Coletiva*, Rio de Janeiro, v. 23, n. 3, p. 925-950, 2013

WORLD HEALTH ORGANIZATION (WHO). *Workload indicators of staffing need (WISN): user's manual*. Geneva: WHO, 2010.

APÊNDICES

ANEXO A - ESCALA NASA-TLX

1. **Demanda Mental (Mental Demand)**

   "Quanta atividade mental e perceptiva foi necessária (pensar, decidir, calcular, lembrar, olhar, buscar etc.)? A tarefa foi fácil ou difícil do ponto de vista mental?"
   Escala: 0 (muito baixa) a 100 (muito alta)

2. **Demanda Física (Physical Demand)**

   "Quanto esforço físico foi necessário (empurrar, puxar, levantar, carregar, movimentar-se etc.)? A tarefa foi fácil ou exigente fisicamente?"
   Escala: 0 (muito baixa) a 100 (muito alta)

3. **Demanda Temporal (Temporal Demand)**

   "Quanto você sentiu que precisava trabalhar rápido ou sob pressão de tempo?"
   Escala: 0 (muito baixa) a 100 (muito alta)

4. **Desempenho (Performance)**

   "Quão bem você acha que executou a tarefa? Você sente que teve sucesso em alcançar os objetivos propostos?"
   Escala: 0 (muito ruim) a 100 (muito bom)
   (Obs: esta é a única dimensão com interpretação inversa em algumas versões — é importante esclarecer ao respondente.)

   Esforço (Effort)

"Quanta energia física e mental foi necessária para alcançar o nível de desempenho que você obteve?"

Escala: 0 (nenhum esforço) a 100 (esforço extremo)

5. **Nível de Frustração (Frustration Level)**

"Quão irritado, estressado, desanimado ou satisfeito você se sentiu durante a execução da tarefa?"

Escala: 0 (nenhuma frustração) a 100 (extrema frustração)

ANEXO B - SCRIPT R UTILIZADO NA ANÁLISE DOS DADOS

```
---
title: "simulacao_gamlss_comparacao_dias"
output: html_document
---
## Carregar pacotes necessários
```{r}
library(gamlss)
library(lubridate)
library(dplyr)
library(tidyverse)
library(scales)
library(ggplot2)
library(MASS) # para ajuste de distribuições
library(dplyr)
```
```

```
library(fitdistrplus)
library(zoo)
library(ggfortify)
library(gridExtra)
library(fda.usc)
library(patchwork)
library(readxl)
```

```{r}
rm(list = ls())
graphics.off()
```

Leitura dos dados brutos

```{r}
dado_bruto_1 = read_excel("../data/dado_bruto_1.xlsx")
dado_bruto_2 = read_excel("../data/dado_bruto_2.xlsx")

dado_bruto<- rbind(dado_bruto_1, dado_bruto_2)
print(head(dado_bruto))
```

Limpando o dado

```{r}
dado_limpo<-dado_bruto%>%
 mutate(data_solicitacao = ymd(DATA_SOLICITACAO)
 )%>%
 group_by(data_solicitacao)%>%
 summarise(numero_solicitacoes = n())%>%
 arrange(data_solicitacao)#%>%
 #filter(data_solicitacao<"2024-01-01")
write_csv2(dado_limpo,"../data/dado_limpo.csv")
```

Ler dados

```{r}
dado <- read_csv2("../data/dado_limpo.csv",show_col_types = FALSE)%>%
 filter(data_solicitacao>= ymd("2020-01-01") & data_solicitacao<= ymd("2025-05-01") )
head(dado)
```

Gráfico da contagem

```{r}
# Criar a variável de soma acumulada
dado <- dado %>%
 arrange(data_solicitacao) %>%
 mutate(soma_acumulada = cumsum(numero_solicitacoes))
# Gráfico 1: Número de solicitações por data
```

```
g1 <- ggplot(dado, aes(x = data_solicitacao, y = numero_solicitacoes)) +
 geom_line(color = "red", linewidth = 0.5) +
 scale_y_continuous(breaks = pretty_breaks(n = 10)) +
 labs(
 title = "Número de Solicitações por Data",
 x = "Data da Solicitação",
 y = "Número de \n Solicitações"
 ) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
 axis.text.x = element_text(angle = 45, hjust = 1),
 panel.grid.minor = element_blank()
 ) +
 scale_x_date(
 date_breaks = "6 month",
 date_labels = "%b/%Y"
 )

# Gráfico 2: Soma acumulada
g2 <- ggplot(dado, aes(x = data_solicitacao, y = soma_acumulada)) +
 geom_line(color = "blue", linewidth = 0.8) +
 scale_y_continuous(breaks = pretty_breaks(n = 10)) +
 labs(
 title = "Soma Acumulada de Solicitações",
```

```
 x = "Data da Solicitação",
 y = "Total Acumulado"
 ) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
 axis.text.x = element_text(angle = 45, hjust = 1),
 panel.grid.minor = element_blank()
 ) +
 scale_x_date(
 date_breaks = "6 month",
 date_labels = "%b/%Y"
 )

# Juntar os dois gráficos verticalmente
g1 / g2 # usando patchwork
```

Box-plots por dia de semana

```{r}
dias_corretos <- c("Segunda-feira", "Terça-feira", "Quarta-feira",
  "Quinta-feira", "Sexta-feira", "Sábado", "Domingo")
```

```
dado_dia_semana <- dado %>%
 mutate(
 dia_semana = wday(data_solicitacao, label = TRUE, abbr = FALSE, week_start = 1),
 dia_semana = factor(dia_semana,
   levels = levels(dia_semana),
   labels = dias_corretos)
 )

ggplot(dado_dia_semana, aes(x = dia_semana, y = numero_solicitacoes, fill = dia_semana)) +
 geom_boxplot(color = "black", outlier.color = "gray30", outlier.shape = 16) +
 scale_y_continuous(breaks = pretty_breaks(n = 10)) +
 scale_fill_brewer(palette = "Set3") +
 labs(
 title = "Distribuições Para Cada Dia da Semana",
 x = "Dia da Semana",
 y = "Número de Solicitações por Plantão",
 fill = "Dia"
 ) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
 axis.text.x = element_text(angle = 45, hjust = 1),
 panel.grid.minor = element_blank(),
 legend.position = "none"
 )
```

```
```

```{r}
#3. Contagem média por dia da semana
resumo_dados<-dado_dia_semana %>%
 group_by(dia_semana) %>%
 summarise(media = mean(numero_solicitacoes),
  desvio = sd(numero_solicitacoes),
  n = n())

resumo_dados
```

```{r}
dado_mes <- dado %>%
 mutate(
 mes = month(data_solicitacao, label = TRUE, abbr = FALSE, locale = "pt_BR.UTF-8")
 )

ggplot(dado_mes, aes(x = mes, y = numero_solicitacoes, fill = mes)) +
 geom_boxplot(color = "black", outlier.color = "gray30", outlier.shape = 16) +
 scale_y_continuous(breaks = pretty_breaks(n = 10)) +
 scale_fill_brewer(palette = "Set3") +
 labs(
 title = "Distribuição da Contagem por Mês",
```

```
 x = "Mês",
 y = "Número de Solicitações",
 fill = "Mês"
 ) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
 axis.text.x = element_text(angle = 45, hjust = 1),
 panel.grid.minor = element_blank(),
 legend.position = "none"
 )
```

```
```{r}
resumo_dados<-dado_mes %>%
 group_by(mes) %>%
 summarise(media = mean(numero_solicitacoes),
  desvio = sd(numero_solicitacoes),
  n = n())

resumo_dados
```
```

Compração entre dias da semana

```{r}
pairwise.wilcox.test(dado_dia_semana$numero_solicitacoes, dado_dia_semana$dia_semana,
  p.adjust.method = "bonferroni")
```

Comparação entre meses

```{r}
pairwise.wilcox.test(dado_mes$numero_solicitacoes, dado_mes$mes,
  p.adjust.method = "bonferroni")
```

Análise de Tendência

```{r}
y <- dado$numero_solicitacoes
tempo<-dado$data_solicitacao
janela <- 30
dados_rolling <- tibble(
 tempo = tempo,
 media = rollapply(y, width = janela, FUN = mean, align = "right", fill = NA),
 variancia = rollapply(y, width = janela, FUN = var, align = "right", fill = NA),
 desvio_padrao = rollapply(y, width = janela, FUN = sd, align = "right", fill = NA)
) %>%
 na.omit()
```

```
# Gráfico
ggplot(dados_rolling, aes(x = tempo)) +
 geom_line(aes(y = variancia, color = "Variância")) +
 geom_line(aes(y = media, color = "Média")) +
 geom_line(aes(y = desvio_padrao, color = "Desvio-Padrão")) +
 scale_color_manual(values = c("Variância" = "darkgray", "Média" = "magenta",
"Desvio-Padrão" = "limegreen")) +
 labs(
 title = "Desvio-Padrão, Média e Variância Móveis (janela de 30 dias)",
 x = "Tempo",
 y = "Valor",
 color = "Indicador"
 ) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
 legend.position = "bottom"
 )+scale_x_date(date_breaks = "1 year", date_labels = "%Y")
```

Testando a presença de tendência para a média

```{r}
options(contrasts = c("contr.treatment", "contr.poly")) # tratamento para fatores nominais
```

```
dado_tendencia<-dado%>%
 mutate(tempo = 1:nrow(dado))
# Modelo nulo (sem tendência)
modelo_nulo <- gamlss(numero_solicitacoes ~ 1, family = NBI(), data = dado_tendencia)
summary(modelo_nulo)
# Modelo com tendência
modelo_tend <- gamlss(numero_solicitacoes ~ tempo, family = NBI(), data = dado_tendencia)
summary(modelo_tend)
# Comparar via AIC
AIC(modelo_nulo, modelo_tend)
```

```{r}
dado_dia_semana$dia_semana <- factor(dado_dia_semana$dia_semana, ordered = FALSE)
modelo_dia_semana <- gamlss(numero_solicitacoes ~ dia_semana, family = NBI(), data =
dado_dia_semana)
summary(modelo_dia_semana)
# Modelo com tendência
modelo_nulo <- gamlss(numero_solicitacoes ~ 1, family = NBI(), data = dado_dia_semana)
summary(modelo_tend)
# Comparar via AIC
AIC(modelo_nulo, modelo_dia_semana)
```

Testando a presença de tendência para a variância

```{r}
dado_tendencia<-dado%>%
 mutate(tempo = 1:nrow(dado))
# Modelo nulo (sem tendência)
modelo_nulo <- gamlss(numero_solicitacoes ~ 1, sigma.formula = ~1, family = NBI(), data = dado_tendencia)
summary(modelo_nulo)
# Modelo com tendência
modelo_tend <- gamlss(numero_solicitacoes ~ 1, sigma.formula = ~tempo, family = NBI(), data = dado_tendencia)
summary(modelo_tend)

# Comparar via AIC
AIC(modelo_nulo, modelo_tend)
```

```{r}
modelo_dia_semana <- gamlss(numero_solicitacoes ~1, sigma.formula = ~ dia_semana, family = NBI(), data = dado_dia_semana)
summary(modelo_dia_semana)
# Modelo com tendência
modelo_nulo <- gamlss(numero_solicitacoes ~ 1, sigma.formula = ~1, family = NBI(), data = dado_dia_semana)
summary(modelo_tend)
# Comparar via AIC
AIC(modelo_nulo, modelo_dia_semana)
```

Testando sozonalidade

```{r}
dado_sazonal <- dado %>%
 mutate(
 tempo = 1:nrow(dado),
 mes = factor(month(data_solicitacao, label = TRUE, abbr = FALSE), ordered = FALSE),
 dia_semana = factor(wday(data_solicitacao, label = TRUE, abbr = FALSE, week_start = 1), ordered = FALSE)
 )

# 2. Modelo nulo (média constante)
modelo_mu_base <- gamlss(numero_solicitacoes ~ 1,
    sigma.formula = ~ 1,
    family = NBI(),
    data = dado_sazonal)

# 3. Modelo com sazonalidade mensal na média
modelo_mu_mes <- gamlss(numero_solicitacoes ~ mes,
   sigma.formula = ~ 1,
   family = NBI(),
   data = dado_sazonal)

# 4. Modelo com sazonalidade semanal (dia da semana) na média
modelo_mu_semana <- gamlss(numero_solicitacoes ~ poly(as.numeric(dia_semana), degree = 1)
```

```
,
   sigma.formula = ~ 1,
   family = NBI(),
   data = dado_sazonal)
modelo_mu_semana
summary(modelo_mu_semana)
# 5. Modelo com ambos os efeitos (mensal e semanal) na média
modelo_mu_completo <- gamlss(numero_solicitacoes ~ mes + dia_semana,
   sigma.formula = ~ 1,
   family = NBI(),
   data = dado_sazonal)

summary(modelo_mu_completo)

# 6. Comparar os modelos via AIC
AIC(modelo_mu_base, modelo_mu_mes, modelo_mu_semana, modelo_mu_completo)

# 7. Visualizar os efeitos estimados da média (μ)
# Efeitos dos meses
plot(modelo_mu_mes, which = 1)

# Efeitos dos dias da semana
plot(modelo_mu_semana, which = 1)

# Efeitos combinados
```

```
plot(modelo_mu_completo, which = 1)
summary(modelo_mu_semana)
```

Auto correlação

```{r}
# Série de contagem
y <- dado$numero_solicitacoes
serie_ts <- ts(y, frequency = 1)

# Gráfico ACF
acf_plot <- autoplot(acf(serie_ts, plot = FALSE),
  conf.int.fill = "plum",
  conf.int.value = 0.95,
  conf.int.type = "white") +
 labs(
 title = "Função de Autocorrelação (ACF)",
 x = "Defasagem (lag)",
 y = "ACF"
 ) +
 ylim(c(-0.25, 0.25)) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
```

```
 axis.text.x = element_text(angle = 0),
 panel.grid.minor = element_blank(),
 panel.grid.major = element_line(color = "gray85")
 ) +
 geom_hline(yintercept = 0, linetype = "dashed", color = "black")

# Gráfico PACF
pacf_plot <- autoplot(pacf(serie_ts, plot = FALSE),
   conf.int.fill = "cyan",
   conf.int.value = 0.95,
   conf.int.type = "white") +
 labs(
 title = "Função de Autocorrelação Parcial (PACF)",
 x = "Defasagem (lag)",
 y = "PACF"
 ) +
 ylim(c(-0.25, 0.25)) +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
 axis.text.x = element_text(angle = 0),
 panel.grid.minor = element_blank(),
 panel.grid.major = element_line(color = "gray85")
 ) +
 geom_hline(yintercept = 0, linetype = "dashed", color = "black")
```

```
# Exibir os dois gráficos na vertical
grid.arrange(acf_plot, pacf_plot, ncol = 1)
```

Modelo Final

```{r}
# Carregar pacotes
library(dplyr)
library(gamlss)

# Criar defasagens de 1 a 5
dado_lags <- dado %>%
 mutate(
 lag1 = lag(numero_solicitacoes, 1),
 lag2 = lag(numero_solicitacoes, 2),
 lag3 = lag(numero_solicitacoes, 3),
 lag4 = lag(numero_solicitacoes, 4),
 lag5 = lag(numero_solicitacoes, 5)
 ) %>%
 # Remover linhas com NA causados pelas defasagens
 filter(!is.na(lag1), !is.na(lag2), !is.na(lag3), !is.na(lag4), !is.na(lag5))%>%
 mutate(

 mes = factor(month(data_solicitacao, label = TRUE, abbr = FALSE)),
```

```
 dia_semana = factor(wday(data_solicitacao, label = TRUE, abbr = FALSE, week_start = 1),
ordered = F)
 )

# Ajustar os modelos com diferentes lags

mod1 <- gamlss(
 formula = numero_solicitacoes ~ lag1 + dia_semana,
 sigma.fo = ~1,
 family = NBI(),
 data = dado_lags
)

mod1
summary(mod1)
mod2 <- gamlss(
 formula = numero_solicitacoes ~ lag1 + lag2 + dia_semana,
 sigma.fo = ~1,
 family = NBI(),
 data = dado_lags
)

summary(mod2)

mod3 <- gamlss(
```

```
 formula = numero_solicitacoes ~ lag1 + lag2 + lag3 + dia_semana,
 sigma.fo = ~1,
 family = NBI(),
 data = dado_lags
)

mod4 <- gamlss(
 formula = numero_solicitacoes ~ lag1 + lag2 + lag3 + lag4 + dia_semana,
 sigma.fo = ~1,
 family = NBI(),
 data = dado_lags
)

mod5 <- gamlss(
 formula = numero_solicitacoes ~ lag1 + lag2 + lag3 + lag4 + lag5 + dia_semana,
 sigma.fo = ~1,
 family = NBI(),
 data = dado_lags
)

# Comparar os modelos por AIC
aic_comparacao <- AIC(mod1, mod2, mod3, mod4, mod5)
print(aic_comparacao)

```

Cusum

```{r}
library(ggplot2)
library(dplyr)
library(scales)
library(qcc)
# Ajustar datas para português (isso funciona em sistemas que têm o locale pt_BR disponível)
Sys.setlocale("LC_TIME", "pt_BR.UTF-8") # No Windows, use "Portuguese_Brazil"
# 1. Limpeza dos dados
dados_filtrados <- dado %>%
 filter(!is.na(data_solicitacao), !is.na(numero_solicitacoes))

# 2. Calcular CUSUM
resultado <- qcc::cusum(
 dados_filtrados$numero_solicitacoes,
 decision.interval = 5,
 se.shift = 1.5
)

# 3. Criar data frame com CUSUM positivo e negativo
df_cusum <- data.frame(
 data = dados_filtrados$data_solicitacao,
 pos = resultado$pos,
 neg = resultado$neg
```

```
)

# 4. Limites
limite <- resultado$decision.interval
limite_superior <- limite
limite_inferior <- -limite

# 5. Classificar os pontos por tipo
df_cusum <- df_cusum %>%
 mutate(
 tipo_pos = case_when(
 pos > limite_superior ~ "viol_pos",
 TRUE ~ "normal_pos"
 ),
 tipo_neg = case_when(
 neg < limite_inferior ~ "viol_neg",
 TRUE ~ "normal_neg"
 )
 )

# 6. Gráfico com cores diferentes por tipo
ggplot(df_cusum, aes(x = data)) +
 # Positivos
 geom_point(aes(y = pos, color = tipo_pos), size = 1.5) +
 # Negativos
```

```
geom_point(aes(y = neg, color = tipo_neg), size = 1.5) +
# Limites e linha de base
geom_hline(yintercept = 0, color = "black") +
geom_hline(yintercept = c(limite_superior, limite_inferior),
 color = "red", linetype = "dashed") +
labs(
title = "Gráfico CUSUM com Classificação de Violações",
x = "Data da Solicitação",
y = "Estatística CUSUM",
color = "Tipo de Ponto"
) +
scale_color_manual(
values = c(
"normal_pos" = "blue",
"viol_pos" = "red",
"normal_neg" = "darkgreen",
"viol_neg" = "red"
),
labels = c(
"normal_pos" = "Positivo (normal)",
"viol_pos" = "Positivo (violação)",
"normal_neg" = "Negativo (normal)",
"viol_neg" = "Negativo (violação)"
)
) +
```

```
 scale_x_date(date_breaks = "6 months", date_labels = "%b/%Y") +
 theme_minimal(base_size = 12) +
 theme(
 plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
 axis.text.x = element_text(angle = 45, hjust = 1),
 legend.position = "bottom",
 panel.grid.minor = element_blank()
 )
```

Predicao

```{r}

dado_final <- dado %>%
 mutate(
 lag1 = lag(numero_solicitacoes, 1)
 ) %>%
 filter(!is.na(lag1)) %>%
 mutate(
 dia_semana = factor(wday(data_solicitacao, label = TRUE, abbr = FALSE, week_start = 1))
 )
# Ajustar o modelo final
modelo_final <- gamlss(
 formula = numero_solicitacoes ~ lag1 + dia_semana,
```

```
 sigma.formula = ~1,
 family = NBI(),
 data = dado_final
)

# Prever para o dia seguinte
ultimo_dia <- tail(dado$data_solicitacao, 1)
lag1_novo <- tail(dado$numero_solicitacoes, 1)
dia_semana_novo <- factor(
 wday(ultimo_dia + 1, label = TRUE, abbr = FALSE, week_start = 1),
 levels = levels(dado_final$dia_semana)
)
novo_dado <- data.frame(
 lag1 = lag1_novo,
 dia_semana = dia_semana_novo
)

# Previsão
media_predita <- predict(modelo_final, newdata = novo_dado, type = "response")
media_predita
mu <- media_predita

sigma <- predict(modelo_final, newdata = novo_dado, what = "sigma", type = "response")
sigma
# Intervalo de confiança (95%)
```

```
# ic_lower <- qNBI(0.025, mu = mu, sigma = sigma)

# ic_lower

# ic_upper <- qNBI(0.975, mu = mu, sigma = sigma)

# ic_upper


ic_lower <- max(0,mu-qnorm(1 - 0.05/2)*sqrt(mu+sigma*mu**2))

ic_upper <- mu+qnorm(1 - 0.05/2)*sqrt(mu+sigma*mu**2)


# Construir dataset para plot

df_plot <- tail(dado, 30)

df_plot$data_solicitacao <- as.Date(df_plot$data_solicitacao)

df_plot


df_prev <- data.frame(

 data_solicitacao = ultimo_dia + 1,

 media = mu,

 ic_inf = ic_lower,

 ic_sup = ic_upper

)


# Construir gráfico com mapeamento para legenda

p <- ggplot() +

 # Série

 geom_line(data = df_plot, aes(x = data_solicitacao, y = numero_solicitacoes, color = "Série"),
linewidth = 1) +
```

```
# Predição
geom_point(data = df_prev, aes(x = data_solicitacao, y = media, color = "Predição"), size = 3) +

# Intervalo de confiança
geom_errorbar(data = df_prev,
 aes(x = data_solicitacao, y = media, ymin = ic_inf, ymax = ic_sup, color = "IC 95%"),
 width = 0.3) +

# Escala de cores manual
scale_color_manual(
name = "", # Título da legenda (vazio)
values = c("Série" = "green", "Predição" = "red", "IC 95%" = "blue")
) +

labs(
title = "Previsão para o Próximo Dia com Intervalo de Confiança (95%)",
x = "Data",
y = "Número de Solicitações"
) +

theme_minimal(base_size = 13) +
theme(
plot.title = element_text(hjust = 0.5, face = "bold"),
legend.position = "bottom",
```

```
 panel.grid.minor = element_blank()
 )+scale_x_date(date_breaks = "1 week", date_labels = "%d/%b/%Y")

# Exibir gráfico
p
```

```{r}
# Número de participantes a simular
n <- 100
# Função para gerar uma única resposta com 8 pontos (2 por faixa)
simular_resposta <- function() {
 faixa1_min <- 0
 faixa1_max <- sample(1:2, 1) + faixa1_min # pequeno intervalo
 pontos_1 <- sample(faixa1_min:faixa1_max, 2, replace = TRUE)
 faixa2_min <- faixa1_max + 1
 faixa2_max <- sample(1:3, 1) + faixa2_min
 pontos_2 <- sample(faixa2_min:faixa2_max, 2, replace = TRUE)
 faixa3_min <- faixa2_max + 1
 faixa3_max <- sample(1:3, 1) + faixa3_min
 pontos_3 <- sample(faixa3_min:faixa3_max, 2, replace = TRUE)
 faixa4_min <- faixa3_max + 1
 faixa4_max <- faixa4_min + sample(3:10, 1)
 pontos_4 <- sample(faixa4_min:faixa4_max, 2, replace = TRUE)
 c(pontos_1, pontos_2, pontos_3, pontos_4)
}
```

```
# Gerar todas as respostas
set.seed(123)
respostas <- t(replicate(n, simular_resposta()))

# Converter para data.frame
colnames(respostas) <- paste0("Ponto_", 1:8)
respostas_df <- as.data.frame(respostas)

# Visualizar os primeiros dados
head(respostas_df)
```

```{r}
n <- 100

resultado_matrix<- matrix(0, nrow = n, ncol = 7)

for (i in 1:n) {
 leve_min <- 0
 leve_max <- sample(1:5, 1)
 leve_max
 moderado_min <- leve_max + 1
 moderado_max <- leve_min + sample(1:5, 1)
 moderado_max
 alto_min <- moderado_max + 1
```

```
 alto_max <- moderado_min + sample(1:5, 1)

 muito_alto<- alto_max + 1


 resultado_matrix[i, ] <- c(leve_min, leve_max, moderado_min, moderado_max, alto_min,
alto_max, muito_alto)


}
```

```{r}
# Suponha que 'df' seja seu data.frame com colunas V1 a V7

# Ajuste se o nome do seu objeto for outro

df <- as.data.frame(resultado_matrix) # substitua aqui se necessário


# Encontrar o máximo valor de V7 (definirá o eixo x final)

x_max <- max(df$V7)

# Criar uma lista para armazenar as séries

series <- list()

# Construir as séries para cada linha

for (i in 1:nrow(df)) {

 y <- rep(0, x_max + 1) # eixo x vai de 0 a x_max

 y[(0:df[i, "V2"]) + 1] <- 1

 y[(df[i, "V3"]:df[i, "V4"]) + 1] <- 2

 y[(df[i, "V5"]:df[i, "V6"]) + 1] <- 3

 y[(df[i, "V7"]:x_max) + 1] <- 4

 series[[i]] <- y
```

```
}

# Transformar em matriz

matriz_series <- do.call(rbind, series)

# Plotar todas as curvas

matplot(t(matriz_series), type = "l", lty = 1, col = rainbow(nrow(matriz_series)),

 xlab = "Número de Ocorrências", ylab = "Nível de Esforço", main = "Séries simuladas por respondente")

legend("topright", legend = paste("Resp", 1:nrow(matriz_series)), col = rainbow(nrow(matriz_series)), lty = 1, cex = 0.6)
```

ESCALA NASA-TLX - simulacao

```
set.seed(123)

num_plantoes <- 5

num_individuos<- 17


resultado<- data.frame(

 individuo = integer(num_individuos*num_plantoes),

 num_ocorrencias = integer(num_individuos*num_plantoes),

 score = numeric(num_individuos*num_plantoes)

)


resultado

for(j in 1:num_individuos)

{
```

```
set.seed(j)

for(i in 1:num_plantoes)
{
num_ocorrencias<- sample(0:15, 1)

num_ocorrencias
demanda_mental<- sample((num_ocorrencias*2):100,1)
demanda_mental<- min(demanda_mental, 100)
demanda_mental

demanda_fisica <- sample((num_ocorrencias):70,1)
demanda_fisica <- min(demanda_fisica, 100)
demanda_fisica

demanda_temporal <- sample(round(num_ocorrencias*1.5):60,1)
demanda_temporal <- min(demanda_temporal, 100)
demanda_temporal

desempenho <- sample(round(num_ocorrencias*2.5):90,1)
desempenho <- min(desempenho, 100)
desempenho

esforco <- sample(round(num_ocorrencias*3):75,1)
esforco <- min(esforco, 100)
```

```
 esforco


 frustracao <- sample(round(num_ocorrencias):(num_ocorrencias+70),1)

 frustracao <- min(frustracao, 100)

 frustracao


 score <- (demanda_mental + demanda_fisica + demanda_temporal + desempenho + esforco +
frustracao) /6


 resultado[(j-1)*num_plantoes+ i, ] <- c(j,num_ocorrencias, score)


 }


}
```

```
```

```` ```{r} ````

```
# Supondo que o seu data frame se chama `resultado` e já está no ambiente

ggplot(resultado, aes(x = num_ocorrencias, y = score)) +

 geom_point(color = "blue", size = 3, alpha = 0.7) + # pontos

 geom_smooth(method = "loess", se = TRUE, color = "red", linewidth = 1.2) + # curva suavizada

 labs(

 title = "Relação entre Número de Ocorrências e Score",

 x = "Número de Ocorrências",
```
```

```
 y = "Score"
 ) +
 theme_minimal(base_size = 14) +
 theme(
 plot.title = element_text(hjust = 0.5, face = "bold"),
 axis.title = element_text(face = "bold")
 )
```

```
```{r}

set.seed(123)
num_plantoes <- 1000
resultado<- data.frame(
 num_ocorrencias = integer(num_plantoes),
 score = numeric(num_plantoes)

)

 for(i in 1:num_plantoes)
 {
 num_ocorrencias<- sample(0:15, 1)
 num_ocorrencias
 demanda_mental<- sample((num_ocorrencias*3+10):100,1)
 demanda_mental<- min(demanda_mental, 100)
```
```

```
 demanda_mental

 demanda_fisica <- sample((num_ocorrencias):70,1)

 demanda_fisica <- min(demanda_fisica, 100)

 demanda_fisica

 demanda_temporal <- sample(round(num_ocorrencias*3.5+10):60,1)

 demanda_temporal <- min(demanda_temporal, 100)

 demanda_temporal

 desempenho <- sample(round(num_ocorrencias*4.5):90,1)

 desempenho <- min(desempenho, 100)

 desempenho

 esforco <- sample(round(num_ocorrencias*6):75,1)

 esforco <- min(esforco, 100)

 esforco

 frustracao <- sample(round(num_ocorrencias**2):(num_ocorrencias**2+10),1)

 frustracao <- min(frustracao, 100)

 frustracao

 score <- (demanda_mental + demanda_fisica + demanda_temporal + desempenho + esforco +
frustracao) /6

 resultado[i, ] <- c(num_ocorrencias, score)


 }

 ```

```{r}


# Aplicar k-means com 3 grupos
```

```
set.seed(123)
kmeans_result <- kmeans(resultado, centers = 3)
# Adicionar grupos ao data frame
resultado$grupo <- as.factor(kmeans_result$cluster)

# Extrair centróides
centroides <- as.data.frame(kmeans_result$centers)
centroides$grupo <- as.factor(1:3)

# Ajustar modelo linear para o grupo 1 (reta vermelha)
modelo_grupo1 <- lm(score ~ num_ocorrencias, data = resultado %>% filter(grupo == 1))
a <- coef(modelo_grupo1)[1]
b <- coef(modelo_grupo1)[2]
x_intersec <- (50 - a) / b # ponto onde score = 50 na reta do grupo 1

# Gráfico com linhas de regressão e marcação score = 50 para grupo 1
ggplot(resultado, aes(x = num_ocorrencias, y = score, color = grupo)) +
 coord_cartesian(xlim = c(-5, 18), ylim = c(10, 90))+
 geom_point(size = 3, alpha = 0.8) +
 #stat_ellipse(type = "norm", level = 0.95, linewidth = 1, linetype = "dashed") +
 geom_smooth(method = "lm", se = FALSE, linewidth = 1.2) + # retas por grupo
 #geom_point(data = centroides, aes(x = num_ocorrencias, y = score),
 #  shape = 4, size = 5, stroke = 2, color = "black") +
 # Segmento horizontal de (0, 50) até (x_intersec, 50)
 geom_segment(aes(x = -5, y = 50, xend = x_intersec, yend = 50),
```

```
 color = "black", linetype = "dashed",linewidth = 1) +
# Segmento vertical de (x_intersec, 50) até (x_intersec, 0)
geom_segment(aes(x = x_intersec, y = 50, xend = x_intersec, yend = 0),
 color = "black", linetype = "dashed",linewidth = 1) +
labs(
title = "Agrupamento K-means (3) com Regressão e Mrcação \n no Score = 50 da Reta Vermelha",
x = "Número de Ocorrências Atendidas por Plantão e por Perito",
y = "Score",
color = "Grupo"
) +
theme_minimal(base_size = 14) +
theme(
plot.title = element_text(hjust = 0.5, face = "bold"),
legend.position = "bottom"
)
```