

# My title\*

My subtitle if needed

Denise Chang

December 3, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

## 1 Introduction

“In an emergency, seconds count.” (City of Toronto 2024). Emergency medical services (EMS) are essential to public health by providing care during life-threatening situations. However, Toronto’s paramedic services are struggling to meet demand. Toronto Auditor General’s Office reports that there were over 1,200 episodes in 2023 where no ambulances were available. (Toronto Auditor General’s Office 2024). To echo this report, this paper examines EMS demand using Toronto Paramedic Services’ incident data from Open Data Toronto.

In this analysis, I am interested in EMS demand against time-based factors in Toronto from 2017 to 2022. I explore trends in paramedic services, such as dispatch time, incident type, and number of units dispatched per incident to identify periods of high-volume demand and low availability of resources. I find that ... [ADD A FEW RESULTS HERE (or in a different paragraph??)]

The remainder of this paper is organized as follows. Section 2 discusses the data source and an overview of the studied variables. Section 3 constructs a model that predicts a shortage in paramedic resources based on time factors. Section 4 presents the results and findings of the exploration of the data. Section 5 discusses implications, limitations, and suggestions for future research. [TELEGRAPH APPENDIX HERE]

---

\*Code and data are available at: [https://github.com/DeniseChang9/Emergency\\_service\\_demands.git](https://github.com/DeniseChang9/Emergency_service_demands.git).

## 2 Data

### 2.1 Overview

The dataset used for this analysis is titled “Paramedic Services Incident Data” and is published by Toronto Paramedic Services (Toronto Paramedic Services 2023). For this paper, the dataset is retrieved from the City of Toronto Open Data Portal.

The statistical programming language R (R Core Team 2023) is used to process, manage and visualize the data. Specifically, statistical libraries such as `opendatatoronto` (Gelfand 2022), `openxlsx` (Schauberger and Walker 2024) and `janitor` (Firke 2023) are used to simulate, download and clean the raw data. Libraries like `arrow` (Richardson et al. 2024) and `readxl` (Wickham and Bryan 2023) were used to save and read datasets. Other libraries like `knitr` (Xie 2021), `here` (Müller 2020) are used to load and to render tables. The library `tidyverse` (Wickham et al. 2019) is useful throughout the entire data manipulation process.

The initial dataset features data on paramedic dispatch time, type of incident, priority level of each incident, number of paramedic units arrived at scene and forward sortation area of the incident. These features are annually refreshed on Open Data Toronto by Toronto Paramedic Services.

The data used for this paper was retrieved on November 25, 2024 and was last refreshed on October 5, 2023.

### 2.2 Measurements

The incident data is saved in different files according to the year of incident. The variables and measurement method of the variables are preserved throughout the recorded years.

Table 1: First Ten Rows of the Paramedic Services Incident Data from 2017

Dispatch Time	Incident Type	Priority Number	Units Arrived	Location
2017-01-01 00:01:13	Medical	1	1	M4L
2017-01-01 00:03:27	Medical	1	1	M3M
2017-01-01 00:01:41	Medical	5	1	M5B
2017-01-01 00:05:53	Medical	4	1	M5V
2017-01-01 00:03:55	Medical	3	1	M1P
2017-01-01 00:08:05	Medical	4	1	M5G
2017-01-01 00:11:10	Emergency Transfer	1	1	M2J
2017-01-01 00:06:20	Medical	4	1	M1B
2017-01-01 00:05:31	Medical	3	1	M2R
2017-01-01 00:07:16	Medical	5	1	M5S

Table 1 is a sample of the 10 first entries in the incident data in 2017. Each row in the dataset represents a unique incident. “Dispatch Time” is the precise time when the first paramedic unit was assigned to an incident. This is measured in year, month, date, hour, minutes, and seconds. “Incident Type” is the category assigned to the incident by the dispatcher based on the information provided by the 9.1.1. caller(s). The possible categories of incident are medical emergencies, emergency transfers, fire-related incidents and motor vehicle accidents. “Priority number” represents the urgency of an incident with 1 being the most urgent and 5 being the least. The priority number is measured by the Medical Priority Dispatch System (MPDS) based on the information provided by the 9.1.1. caller(s). “Units Arrived” is the total number of paramedics that arrived on the scene of incident. This is measured by counting the number of different paramedic units who were dispatched and responded to the incident. Units part of this count do not have to be simultaneously present, and a unit who leaves the scene and comes back afterwards is considered 1 count. “Location” is the general location of the incident based on Postal Code Forward Sortation Areas. This is determined as the first three characters of the postal code of the incident location.

## 2.3 Variables of Interest

In the analysis dataset, most variables were constructed using the variables in the raw dataset. From the original dataset, “Dispatch Time” was broken down into four separate variables in the analysis dataset: “Year”, “Month”, “Day of the Week” and “Hour of the Day”. The aim for this construction is to isolate the effect of each temporal factor on the paramedic incident request paramedic response. While holding the time variables and the “Incident Type” variable constant, two more variables were constructed from the “Units Arrived” variable from the raw dataset: “Average Units Arrived” and “Number of Incidents”.

Table 2: Ten First Rows of the Analysis Data on Paramedic Services Incident Data from 2017 to 2022

Year	Month	Day of the Week	Hour of the Day	Incident Type	Average Units Arrived	Number of Incidents
2017	Jan	Sun	0	emergency transfer	1.3	6
2017	Jan	Sun	0	fire	2.7	3
2017	Jan	Sun	0	medical	1.1	179
2017	Jan	Sun	0	motor vehicle accident	1.7	6
2017	Jan	Sun	1	emergency transfer	1.0	2
2017	Jan	Sun	1	fire	4.0	1
2017	Jan	Sun	1	medical	1.0	182

Year	Month	Day of the Week	Hour of the Day	Incident Type	Average Units Arrived	Number of Incidents
2017	Jan	Sun	1	motor vehicle accident	1.7	3
2017	Jan	Sun	2	emergency transfer	1.0	4
2017	Jan	Sun	2	medical	1.0	143

Table 2 is a sample of the analysis dataset used in this paper. The variables of interest of the analysis dataset are:

- Number of incidents: This variable represents the total count of emergency incidents recorded for each unique combination of time and incident-type factors. This is the outcome variable in the model, since it reflects the demand for paramedic services.
- Average number of units arrived: This variable reflects the average number of paramedic units dispatched and arrived at the scene for a given combination of incident type, time, and date. This is another outcome variable, since it provides insight into resource allocation.
- Incident Type: This categorical variable describes the nature of each emergency, such as medical emergencies, fire-related incidents, motor vehicle accidents, or emergency transfers. It captures the diversity of incidents and is useful for exploring differences in resource needs.
- Year: The year the incident occurred, capturing long-term trends in incident frequency.
- Month: The month of occurrence, which highlights seasonal trends.
- Day of Week: The day on which the incident took place, accounting for patterns related to weekdays versus weekends.
- Hour: The hour of the day, capturing diurnal patterns in emergency service demand.

### 3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

### 3.1 Model set-up

We run the model in R (R Core Team 2023) using the `rstanarm` package (Goodrich et al. 2022). Default priors from `rstanarm` are applied. The estimating equation is as follows:

Define  $y_i$  as the number of incidents observed for a specific combination of factors. Then,  $\mu_i$  is the expected number of incidents based on temporal and incident-specific predictors.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \alpha + \beta_{\text{year}} \cdot \text{year}_i + \sum_{k=1}^{11} \beta_{\text{month},k} \cdot \text{month}_{ik} + \sum_{j=1}^6 \beta_{\text{day},j} \cdot \text{day}_{ij} \quad (2)$$

$$+ \beta_{\text{hour}} \cdot \text{hour}_i + \sum_{t=1}^3 \beta_{\text{incident},t} \cdot \text{incident}_{it} + \beta_{\text{units}} \cdot \text{avg\_units}_i \quad (3)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\beta \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\sigma \sim \text{Exponential}(1) \quad (6)$$

In this model, we define the variables as follows:

- $y_i$  is the number of incidents observed for a specific combination of factors
- $\mu_i$  is the expected number of incidents based on predictors
- $\alpha$  is the intercept term, representing the baseline number of incidents.
- $\beta_{\text{year}}$  is the coefficient for the year variable.
- $\beta_{\text{month}}$  is the set of coefficients for the month of the year.
- $\beta_{\text{day}}$  is the set of coefficients for the day of the week.
- $\beta_{\text{hour}}$  is the coefficient for the hour of the day.
- $\beta_{\text{incident}}$  is the set of coefficients for the incident types.
- $\beta_{\text{units}}$  is the coefficient for the average number of units arriving on scene
- $\sigma$  is the standard deviation of the residuals, representing unexplained variability in incident counts.

### 3.1.1 Model justification

For this analysis, a Bayesian Generalized Linear Model (GLM) was chosen to predict incident call counts as a function of temporal factors, incident type, and resource allocation. This approach provides the flexibility to model log-transformed continuous outcomes and includes prior beliefs about the relationships between predictors and the outcome. The Bayesian framework quantifies uncertainty in parameter estimates and predictions, which is useful in our context where variability and uncertainty are prevalent. This model is appropriate given the observed overdispersion in the count data.

In Bayesian modeling, priors were defined to reflect plausible values. A Normal(0, 10) prior was used for the intercept ( $\alpha$ ) to allow a wide range of baseline incident counts across predictor combinations. Predictor coefficients ( $\beta$ ) were assigned a Normal(0, 2.5) prior to accommodate moderate effect sizes without being too restrictive. The residual standard deviation ( $\sigma$ ) was given an Exponential(1) prior to emphasize smaller residual errors. These priors avoid constraints that are too narrow, reducing bias in the estimates.

This model relies on several assumptions. First, it assumes a linear relationship between predictors and the log-transformed outcome variable, which simplifies interpretation. The model also assumes residuals follow a Gaussian distribution, consistent with the log-transformation of the outcome variable. Despite its strengths, the model has limitations. It does not explicitly model nonlinear relationships or interactions, like those between temporal and incident type variables, which could capture more complex dynamics. Finally, the Gaussian residual assumption may be inappropriate in certain cases of extreme overdispersion.

Several alternative models were considered. Poisson regression, a natural choice for count data, was not chosen due to overdispersion in the outcome variable. Negative binomial regression was also considered, but does not incorporate prior beliefs or uncertainties as naturally as Bayesian methods.

Ultimately, the Bayesian Gaussian GLM was selected since it can handle continuous log-transformed data and include prior beliefs while still being digestible. Through Bayesian inference, the model quantifies uncertainty and provides a probabilistic framework ideal for emergency response planning. This approach ensures actionable predictions for resource allocation decisions.

## 4 Results

Our results are summarized in [tbl-model](#).

The Bayesian GLM model, based on 38,494 observations and 24 predictors, has a posterior predictive mean (mean\_PPD) of 44.4, which aligns with observed incident counts. This shows that the model captures variability.

Table 3: ?(caption)

## Model Info:

```

function:      stan_glm
family:        gaussian [identity]
formula:       count ~ year + month + day_of_week + hour + incident_type + avg_units_arrived
algorithm:     sampling
sample:        4000 (posterior sample size)
priors:        see help('prior_summary')
observations:  38494
predictors:    24

```

## Estimates:

	mean	sd	10%	50%	90%
(Intercept)	0.2	137.7	-174.2	1.8	174.3
year	0.0	0.1	-0.1	0.0	0.1
month.L	2.3	0.4	1.8	2.3	2.8
month.Q	-1.3	0.4	-1.8	-1.3	-0.8
month.C	-2.0	0.4	-2.5	-2.0	-1.5
month^4	3.5	0.4	3.0	3.5	4.0
month^5	-0.1	0.4	-0.6	-0.1	0.4
month^6	1.3	0.4	0.8	1.3	1.8
month^7	-1.6	0.4	-2.1	-1.6	-1.0
month^8	1.8	0.4	1.3	1.8	2.3
month^9	-0.1	0.4	-0.6	-0.1	0.4
month^10	0.9	0.4	0.4	0.9	1.4
month^11	-0.4	0.4	-0.9	-0.4	0.1
day_of_week.L	0.9	0.3	0.5	0.9	1.2
day_of_week.Q	-0.7	0.3	-1.1	-0.7	-0.3
day_of_week.C	-0.4	0.3	-0.8	-0.4	0.0
day_of_week^4	-0.8	0.3	-1.2	-0.8	-0.4
day_of_week^5	-0.1	0.3	-0.5	-0.1	0.3
day_of_week^6	-0.2	0.3	-0.6	-0.2	0.2
hour	1.2	0.0	1.2	1.2	1.2
incident_typefire	-6.6	0.5	-7.3	-6.6	-6.0
incident_typemedical	126.4	0.3	126.0	126.4	126.8
incident_typemotor vehicle accident	0.2	0.3	-0.2	0.2	0.6
avg_units_arrived	1.3	0.2	1.1	1.3	1.6
sigma	22.5	0.1	22.4	22.5	22.6

## Fit Diagnostics:

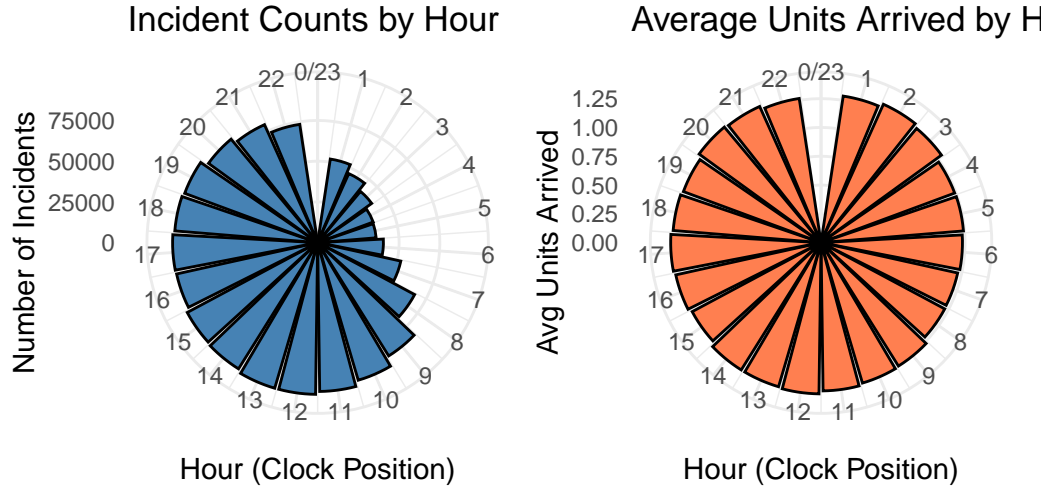
	mean	sd	10%	50%	90%
mean_PPD	44.4	0.2	44.2	44.4	44.6

The mean\_ppd is the sample average posterior predictive distribution of the outcome variable

## MCMC diagnostics

	mcse	Rhat	n_eff
(Intercept)	1.4	1.0	9511
year	0.0	1.0	9514
month.L	0.0	1.0	6858
month.Q	0.0	1.0	9371
month.C	0.0	1.0	9799

Table 4: ?(caption)



Temporal factors play a role in determining incident volume. The coefficient for hour (mean = 1.2) shows predictable variations in demand throughout the day, with highly precise estimates (sd = 0.0). Monthly and daily trends are captured through polynomial terms, reflecting the nonlinear nature of incident occurrences across different times of the year and week.

The analysis suggests the dominance of medical incidents in driving overall incident counts. The coefficient for medical calls (mean = 126.4) is much higher than the coefficient for fire-related incidents (mean = -6.6) and the coefficient for motor vehicle accidents (mean = 0.2). This shows that medical emergencies is the primary source of paramedic service demand, requiring special attention in operational planning.

The positive relationship between the average number of units arriving and incident counts (mean = 1.3) suggests that resource allocation is another determining factor. There is a correlation where increased resource availability may correspond to higher incident volumes. This reflects areas of concentrated demand and hints at resource allocation strategies.



#### **4.1 Hour of the Day on Paramedic Incident Demand and Reponse**

### **5 Discussion**

#### **5.1 First discussion point**

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

#### **5.2 Second discussion point**

Please don't use these as sub-heading labels - change them to be what your point actually is.

#### **5.3 Third discussion point**

#### **5.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

# Appendix

## A Additional data details

## B Model details

### B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected  
by, the data

Figure 1: `?(caption)`

### B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC  
algorithm

Figure 2: `?(caption)`

## References

- City of Toronto. 2024. “Public Safety Alerts - City of Toronto.” 2024. <https://www.toronto.ca/community-people/public-safety-alerts/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragos Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Schauberger, Philipp, and Alexander Walker. 2024. *Openxlsx: Read, Write and Edit Xlsx Files*. <https://CRAN.R-project.org/package=openxlsx>.
- Toronto Auditor General’s Office. 2024. “Toronto Paramedic Services: Rising Response Times Caused by Staffing Challenges and Pressures in the Healthcare System.” 2024. <https://www.torontoauditor.ca/report/toronto-paramedic-services-rising-response-times-caused-by-staffing-challenges-and-pressures-in-the-healthcare-system/>.
- Toronto Paramedic Services. 2023. “Paramedic Services Incident Data.” 2023. <https://open.toronto.ca/dataset/paramedic-services-incident-data/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, and Jennifer Bryan. 2023. *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>.
- Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.