Analyzing The Effects Of Gender, Education And profession On An Individual's Decision To Call Emergency Medical Services In Time Of Need In The Eastern Region Of Saudi Arabia*

Owais Zahid

April 19, 2024

In this study, we evaluate the effect of one's gender, education level, and whether their profession is related to the healthcare industry on their decision to call for Emergency Medical Services (EMS) in the Eastern Region of Saudi Arabia. We find that individuals working in healthcare-related professions are more likely to call EMS, whereas individuals are less likely to do so the higher their education level is. Furthermore, there is no clear evidence that the gender of an individual influences their decision to call EMS.

1 Foreword

Having lived in Saudi Arabia my entire life before coming to Canada for university, this topic is close to my heart and brings back a lot of memories of what I consider to be home. I hope you enjoy reading this paper just as much as I did writing it.

2 Introduction

Being aware of any and all forms of medical assistance granted by the government is important in an emergency however individuals may often see fit to take matters into their own hands and head to the hospital instead of relying on these services, perhaps due to wait time, quality of service, etc. This phenomenon is what we will be analyzing today within the Eastern region of Saudi Arabia. We thus aim to investigate whether an individual chooses to call for

*Code and data are available at:https://github.com/FFFiend/KSA_EMS_Population_Eval

Emergency Medical Services (EMS) within the region, given their gender, education level and whether they're in a healthcare-related profession. The estimand we are trying to estimate is the individual's decision on calling for EMS in an emergency which is a binary quantity. The data used in this paper is sourced from Harvard Dataverse, and the dataset, called " The Awareness of Public about the Emergency Medical Services in the Eastern Region of Saudi Arabia" (Alanazy 2024) provides statistics on respondents to an EMS related survey such as their gender, age, education, etc.

The dataset will be explored in full detail in the **Data** section, followed by the presentation of a logistic regression model to justify and predict whether an individual will call EMS, given their aforementioned characteristics. Finally, we discuss and analyze the results obtained upon fitting the model as well as the implications and potential weaknesses of our approach, and possible enhancements.

It is worth noting that an investigation of this kind is necessary as it indicates first and foremost whether emergency healthcare services for residents are in need of improvement. It may also indicate whether more awareness of EMS is needed at one or more education levels, which is why we chose the former as one of the predictor variables. Simply taking matters into your own hands and rushing to the hospital in case of an emergency may be the best course of action, often primarily due to ambulance wait times (speaking from personal experience), however this indicates to the government that perhaps the emergency medical services sector needs to be revamped and is insightful when considering budget allocation for the coming fiscal year.

We shall be using R (R Core Team 2023), as well as the following packages throughout the paper: Knitr (Xie 2023), Arrow (Richardson et al. 2024), ggplot2 (Wickham 2016), modelsummary (Arel-Bundock 2022), rstanarm (Goodrich B 2024)

3 Data

3.1 Data processing and Exploration

The original dataset has 435 entries and consists of the following columns: **No, RENAMED_CONSENT_COL, Gender, Age, Education, health_professions, person_unconscious, Know_EMS, Called_EMS, Emergency Number, EMS_Provide, dispatcher, EMS_Arrive, ambulance_behind, EMS_Scene, patients_field, EMS_coverage, MEDEVAC?, trust_EMS, enter_house, non_urgent, Tawakkalna, current_EMS**. A preview of the original table was not included due to the sheer number of columns. Also, the contents of **RENAMED_CONSENT_COL** were in Arabic which proved to be an inconvenience during data processing.

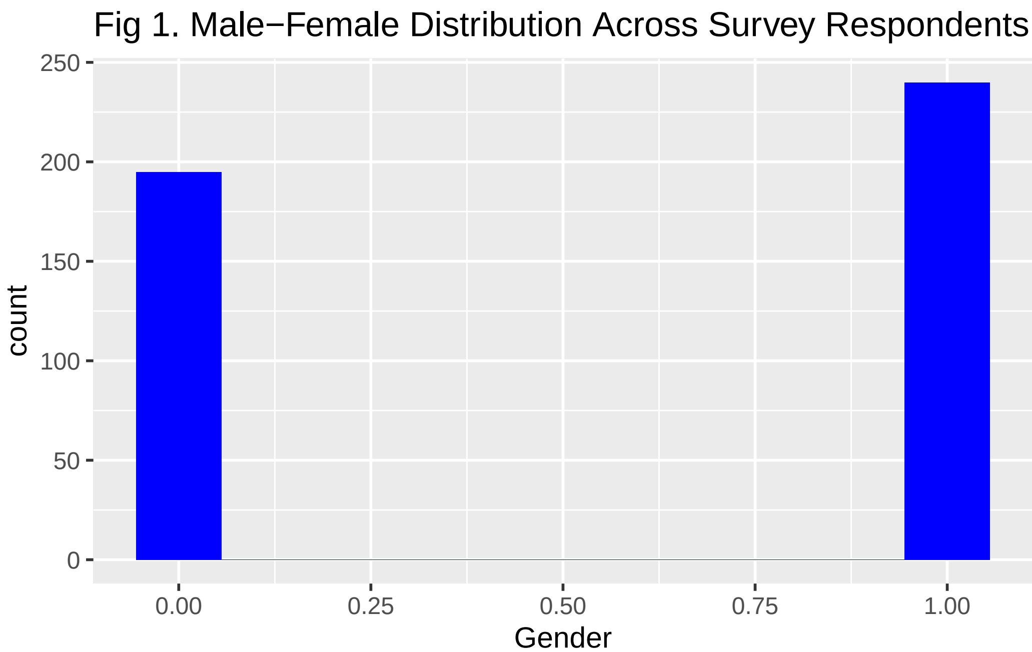
Thus, the only columns we are concerned with for our investigation are **Gender, Education, health_professions and Called_EMS** which are previewed below in our processed data table: "' ::: {.cell caption='false'} ::: {.cell-output-display} Table: Table 1. EMS Population Evaluation Data

Gender	Education	health_professions	Called_EMS
1	3	1	0
0	3	0	0
0	3	0	0
0	3	1	0
1	3	1	0
1	3	1	0
1	3	1	1
0	3	1	0
1	3	1	1
1	1	0	0

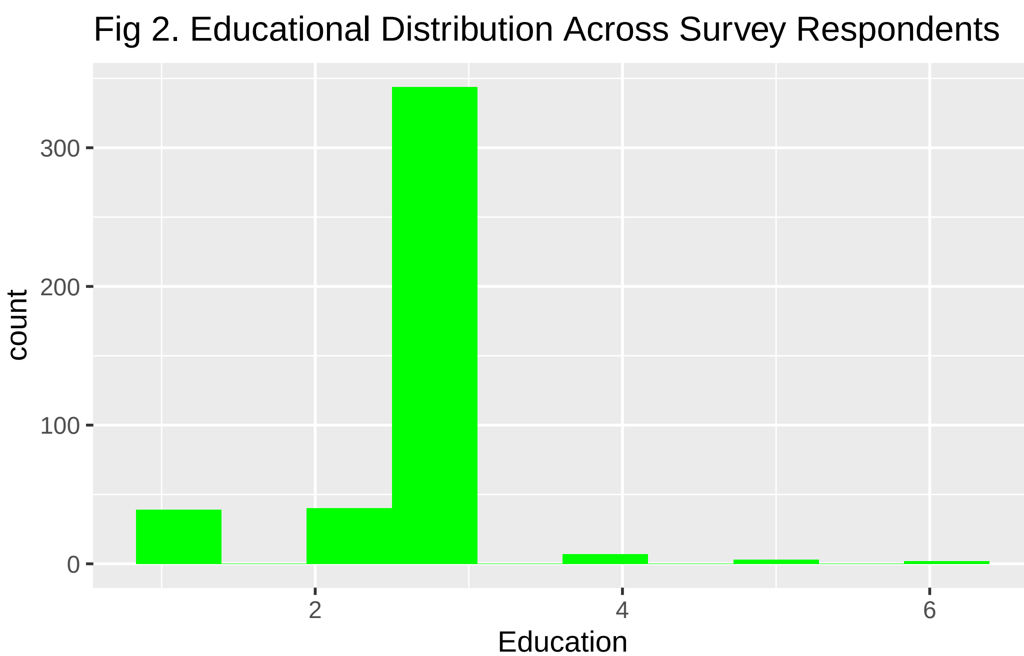
::: ::: We shall now elaborate on the numeric ranges found in each column and what values they correspond to starting with **Gender** where we have values 1 and 0 mapping to Male and Female respectively. Each entry in the **Education** column contains 1 of 6 possible values with 1,2,3,4,5,6 mapping to High Schooler, Diploma, Bachelor's, Master, Doctorates and Other respectively. Then we have another binary quantity in the form of the **health_professions** column where 1 signifies that a person has a healthcare-related job, and 0 signifies the opposite. Finally, our binary response variable **Called_EMS** maps 1 to yes and 0 to no.

3.2 Trends Across Predictor Variables

We now examine each predictor variable to gain a better understanding of the dataset.

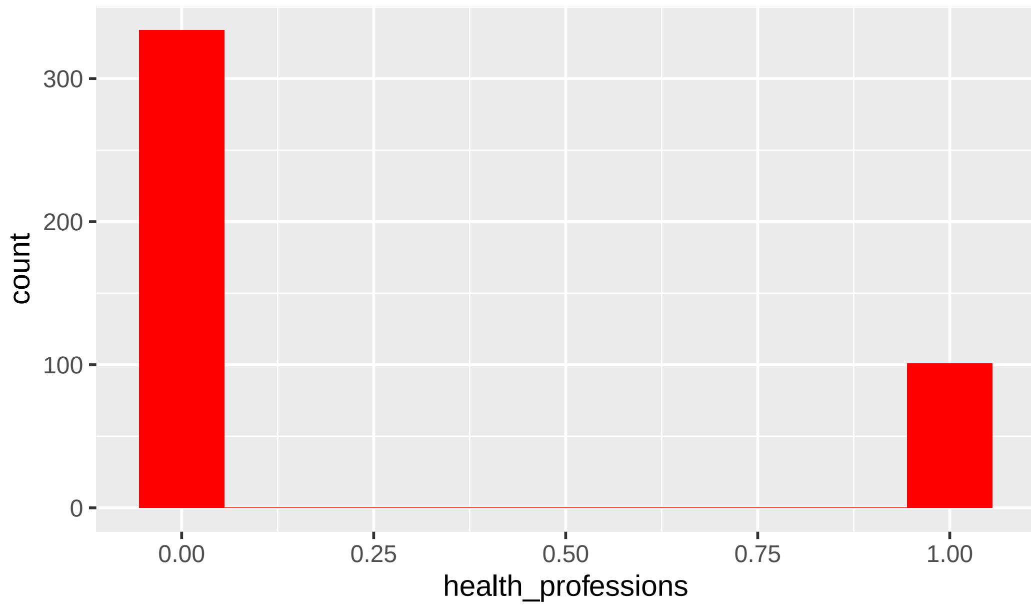


We note from Fig 1. that there are more male respondents compared to female, however not by much as we have a M-F 240-195 split.



Bachelor's degree holders comprised the highest percentage of survey respondents, at exactly 344 out of 435 individuals followed by both High School and Diploma education holders at 39 and 40 respondents respectively as seen in Fig 2. Lastly, there were 7 respondents with a Masters degree, only 3 respondents with a Doctorates, and 2 respondents chose to answer Other.

Fig 3. Health−Related Profession Distribution



Finally, we see in Fig 3. that 334 respondents did NOT work in a healthcare-related field, and 101 respondents held healthcare-related jobs.

Now that the variables of interest have been isolated and our response variable is well defined, it is worth noting that the data used for this investigation was obtained as part of a separate study titled "The Awareness of Public about the Emergency Medical Services in the Eastern Region of Saudi Arabia" and evaluates knowledge and trust in Emergency Medical Services (EMS) in Eastern Saudi Arabia, identifying factors contributing to low awareness. A cross-sectional study from September 2022 to September 2023 surveyed participants aged 18-60 from various backgrounds. Data was gathered using a validated questionnaire covering demographics, hypothetical scenarios, EMS knowledge, and trust in EMS. Evidently, the value mappings above were obtained by looking up the corresponding values within the survey answer key.

4 Model

Since our estimand is a binary value, i.e whether or not a person calls EMS given their gender, education and professional-classification, we estimate the following model:

$$
\begin{aligned}
y_i|\pi_i, &\sim \text{Bern}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i + \beta_3 \times \text{health-profession}_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(1, 2.5) \\
\beta_2 &\sim \text{Normal}(3, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5)
\end{aligned}
$$

Elaborating on the model set up: The outcome variable y_i follows a Bernoulli distribution parameterized by π_i.

$\text{logit}(\pi_i)$ represents the logistic regression model equation. It models the log odds of π_i , which is the probability of y_i being 1 (i.e., calling EMS), as a linear combination of predictor variables (gender, education, and health_profession), each multiplied by their respective coefficients (β_i).

Each of the predictor variable coefficients β_i follow a Normal distribution with standard deviation 2.5 although with varying mean values. Recall from our exploration of trends across predictor variables that for gender: most were male, education: most had a bachelors and most under the health-profession column were NOT working jobs related to the healthcare industry. Thus, we can set the means for the priors of the aforementioned coefficients as 1, 3 and 0 respectively.

4.1 Model Justification

Logistic regression is a suitable model for investigating whether a person calls Emergency Medical Services (EMS) given their gender, education, and profession, for several reasons:

Binary Outcome: Logistic regression is appropriate when the outcome variable is binary, such as whether a person calls EMS (yes/no). Since the outcome in this case is binary, logistic regression is a natural choice.

Interpretability: Logistic regression provides interpretable coefficients that allow us to understand the relationship between the predictor variables and the probability of the outcome. In this case, we can interpret the coefficients to understand how gender, education, and profession relate to the likelihood of calling EMS.

Model Flexibility: Logistic regression allows for the inclusion of multiple predictor variables, making it suitable for analyzing the influence of gender, education, and profession simultaneously on the probability of calling EMS.

Non-linearity Handling: Logistic regression can capture nonlinear relationships between predictor variables and the log odds of the outcome, providing flexibility in modeling complex relationships.

Assumption of Independence: Logistic regression does not assume that the predictor variables are independent of each other, making it suitable for situations where predictors may be correlated, such as gender, education, and profession.

Well-Established Method: Logistic regression is a well-established statistical method with extensive theoretical and practical support. It is widely used in various fields, including healthcare research, for modeling binary outcomes.

Probability Estimation: Logistic regression provides estimates of probabilities rather than just predictions, allowing us to quantify the likelihood of calling EMS based on gender, education, and profession.

Robustness: Logistic regression tends to perform well even with relatively small sample sizes, making it suitable for analyzing datasets commonly encountered in healthcare research.

Overall, logistic regression is a robust and interpretable method for investigating the relationship between predictor variables (gender, education, profession) and the likelihood of calling EMS, making it a suitable choice for this situation.

5 Results

Intercept (-0.284): The intercept represents the log odds of the outcome variable (calling EMS) when all predictor variables are zero. Since logistic regression works on the log odds scale, this value indicates the log odds of calling EMS for a hypothetical individual with zero values for Gender, Education, and health_professions. It does not make sense to consider such a hypothetical for our investigation as education only spans non-negative values.

Gender Coefficient (0.143): For every one-unit increase in the Gender variable (e.g., moving from male to female, assuming the variable is coded as 0 and 1), the log odds of calling EMS increase by 0.143, holding Education and health_professions constant. This also means that whether a person calls EMS does not strongly depend on their gender.

Education Coefficient (-0.535): For every one-unit increase in the Education variable (e.g., moving from a lower to a higher education level), the log odds of calling EMS decrease by 0.535, holding Gender and health_professions constant, meaning that an individual is less likely to call EMS the more educated they are.

Health_professions Coefficient (1.094): For every one-unit increase in the health_professions variable (e.g., a person belonging to a health profession versus not belonging), the log odds of calling EMS increase by 1.094, holding Gender and Education constant, meaning that individuals working in healthcare-related fields are more likely to call EMS.

	(1)
(Intercept)	−0.284
Gender	0.143
Education	−0.535
health_professions	1.094
Num.Obs.	435
R2	0.078
Log.Lik.	−208.438
ELPD	−212.6
ELPD s.e.	11.5
LOOIC	425.3
LOOIC s.e.	23.0
WAIC	425.3
RMSE	0.40

6 Discussion

6.1 Individuals working in non healthcare-related professions are less likely to call EMS

This may be due to a lack of awareness amongst individuals working in said professions, either at the regional level or perhaps the entire country. Saudi Arabia is a relatively young country, and perhaps more attention needs to be drawn to services such as EMS that are available to the public. Recall that one of the columns in the original dataset, **Know_EMS**, is a possible candidate for a predictor that we can extend our model by. Note also that simply adding the former as a fourth predictor variable would only help us test the "lack of awareness" possibility within the eastern region, and that more data from multiple regions is needed to make a claim at a country-wide level.

6.2 An individual is less likely to call for EMS the higher their education level is

Our results show that more educated individuals in the Eastern Saudi Arabian region are less likely to call EMS. This is contrary to findings in a related paper titled "Does education lead to higher generalized trust? The importance of quality of government" (Nicholas Charron 2016) which states that "Strong support is found for a positive link between education and trust at the individual level when quality of institutions is high". This difference in public opinion amongst individuals with higher education may be attributed to cultural differences between Saudi Arabia and the 24 European countries which were surveyed in the aforementioned study. It is however likely that the discrepancy arises from a lack of sufficient data, i.e we fit our model on survey responses from 435 individuals in a specific region of Saudi Arabia, and perhaps increasing the radius to capture public opinion in more regions and/or surveying a larger number of individuals within the Eastern region may provide a similar outcome.

6.3 Weaknesses & Next Steps

6.3.1 Weaknesses

A possible weakness of our analysis is the low number of candidates that responded to the survey in the study titled "The Awareness of Public about the Emergency Medical Services in the Eastern Region of Saudi Arabia" that we based our model on. Secondly, the study focuses only on the Eastern Saudi Arabian region, and it is possible that we obtained a myopic view of public opinion regarding Emergency Medical Services (EMS) in the country. Searching for and pooling similar studies together to obtain a higher respondent count, or creating our own survey and sending it out to separate regions at once is a possible next step that can be discussed.

Secondly, an uninformed selection of standard deviations across priors for our model's predictor coefficients may be a reason why our results differ from the study by Charron and Rothstein. Recall that we selected 2.5 as the standard deviation for all coefficient priors, when we could have simply calculated the standard deviation for each predictor column. The selection of standard deviations was thus uninformed on the dataset, which may have led to skewed results.

6.3.2 Next Steps

As alluded to in the prior two sections, there are a couple things we can do to extend our analysis:

1. Using the **Know_EMS** column as a predictor variable, adding a fourth predictor to our model and conducting analysis on the same dataset to see if higher education is related to an increased likelihood of calling for EMS.
2. Gathering similar studies of the Saudi Arabian population or creating our own survey on Emergency Medical Services and sending it to multiple regions within the country may give us a more accurate view of the population's opinion and decision-making regarding EMS, thereby increasing model performance as more data to train on leads to better fitted results.

References

Alanazy, Ahmed. 2024. "The Awareness of Public about the Emergency Medical Services in the Eastern Region of Saudi Arabia." Harvard Dataverse. https://doi.org/10.7910/DVN/BPVQET.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Goodrich B, Ali I & Brilleman S., Gabry J. 2024. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm.

Nicholas Charron, Bo Rothstein. 2016. "Does Education Lead to Higher Generalized Trust? The Importance of Quality of Government." *International Journal of Educational Development* 50: 59–73. https://doi.org/https://doi.org/10.1016/j.ijedudev.2016.05.009.

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, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://github.com/apache/arrow/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.