Analysis of Vance County EMS Ambulance Distribution

Eunice Lee, Sonya Eason, Brian Kim

1 Background

The Vance County Emergency Medical Services system currently operates four ambulances and two stations with one located in the Southern district and one in the Central district. This configuration leads to the underserving of residents in the Northern district, with higher average response times than the other two districts. Such delays can lead to critical differences in patient outcomes especially in life threatening emergencies.

In response to this issue and demand for EMS coverage in the North region, Vance county is evaluating the option of adding a new station in one of the two northern locations: near north or far north station. We use historical EMS call data to make a decision. Each record in the dataset represents one individual emergency trip which includes information like dispatch station, patient location coordinates, and time logs for dispatch, arrival, hospital transport, and clearance.

The main motivation of this analysis is to assess how travel times and system load vary across different station allocation scenarios, listed in Table 1 (appendix). In particular, we aim to answer the following central questions: 1) which of the two possible Northern station locations (Near vs Far) would better serve the community through faster response time 2) which station's relocation to the North (South vs Central) would yield to improved response time and 3) which combination of allocation of the four available ambulances across stations would best optimize service coverage and response times.

2 Exploratory Data Analysis

To understand the distribution of our data, we first model the best-guess travel time for all data points into a histogram (Figure 1) and observe that the South station has shortest median travel time. Both the Near North and Far North stations, on the other hand, have longer travel time with wider variability as expected. We can observe that the Far North station has the

widest spread and most outliers, implying more inconsistent travel times possibly stemming from its remote location and small sample size.

We then examine the occurrence of "load conflicts" in each region, defined as stress events wherein overlapping calls occur when ambulances are already occupied. Comparing the distribution of load conflicts in Central and North regions (Figure 3) versus the South (Figure 4), we find significantly more load conflicts in the Central and North regions in the baseline scenario. We also observe that most conflicts are short, lasting a few minutes and occurring on a nearly daily basis. However, observng the heavy tail in Figure 3, it can be seen that some extreme cases develop into "critical stress events" which we define as 3+ load conflicts occurring at once. This suggests that response times could be delayed disproportionately for North stations.

3 Model Assumptions & Assessment

As a foundational assumption to follow for all scenarios, we set a "dispatch rule" on how ambulances will be dispatched depending on location of the call. For scenario 0 (current baseline) where 3 ambulances are in Central and 1 is in South station, we assume that all North and Central calls are taken by the Central station and all South calls are taken by the South station ambulances. For scenarios 1 and 2, still with 3 ambulances in the Central station but with 1 South ambulance relocated to the North (Near or Far, respectively), we assume that North calls are taken only by the North station and Central or South calls are only taken by the Central station ambulances. For scenarios 3 and 4, with 1 South ambulance, 2 Central ambulances, and 1 Central moved to the North (Near or Far, respectively), we assume all calls are taken by an ambulance from its same-region station. However, we make an exception to these rules in the case of critical stress events at an assigned station, as it would be unethical to "gatekeep" ambulances from dispatching to a different region when they are available. In such cases of 3+ load conflicts, we make the available ambulance from the next closest station to be dispatched to take the call.

Under the assumption of a hierarchical approach where we condition on the calls over the total change in time, we test for two different models using scenario, distance, and rush hour as covariates. They are both linear mixed models utilizing random effect on call groups, since there is within-group variation for each call. To examine whether considering heteroscedasticity leads to a better or worse model fit, we examine the residual plots for both candidate mixed models, keeping all else constant but modeling for varying potential errors by longitude and latitude in one and assuming constant error variance in another. In the model with constant error variance (Figure 6), we observe that residuals are centered around 0 with no strong curvature, so the assumption of linearity is reasonable even though there is some spread. The pattern is not indicative of systematic heteroscedasticity. In the model where we allow error variance to differ by scenario due to heteroscedasticity (Figure 8), we find no meaningful improvement in the residual spread or overall structure. Since the more complex variance

structure does not reduce the heteroscedastic pattern, the simpler model without additional error variation is preferred. Thus, we choose this for our final model.

4 Results

We note that there are different residuals for the aforementioned fits, when tested for response time with or without heteroscedasticity on the random effects. We find that when examining the normality condition, the model without the error variations shows a better fit for the data (Figures 7 and 9). Consequently ,we analyze the results under the fit without the additional variation.

Our final model is the following linear mixed model:

$$Response\ Time \sim Scenario + Distance + RushHour + (1|CallID) + \epsilon$$

For fixed effects, we include scenarios as a categorical variable, with S0 as baseline and S1~S4 as changes. Distance from the station to the call location is calculated as a continuous variable. We also include rush hour as binary variable to account for traffic levels in a simplified manner, by categorizing the time of day into non-rush vs rush hour. Here, rush hour is defined as 7:30-9:30AM and 4:00-6:30PM on each day, as per observation of increased average traffic levels during those times. The model's random effects account for variation by call group, and residual error models day-to-day variance.

For model assumptions, we let residual error follow a normal distribution, and assume traffic level to be consistent within and outside rush hours to simplify traffic levels on a binary level rather than continuous. Based on our residual plot, we do not model for heteroscedasticity for random effects since it leads to a better model fit. Finally, we assume that if distance is the same, the actual latitude and longitude coordinates would not affect response time differently.

In the coefficient summary for this best-fit model (Table 2), we can see that the estimated coefficients for all four scenario indicators have p-values that are greater than 0.05. 95% confidence intervals cross zero in all scenarios 1-4, which indicates that holding Distance and Rush Hour constant, there is no statistically significant evidence that response time differs across the alternative allocation scenarios compared to the baseline configuration of S0.

Therefore, from our model results analysis, we can conclude the following for our initial research questions:

1) There is no statistically significant difference in response time between the Near North and Far North stations and both scenarios perform similarly once we adjust for distance and rush hour.

3

2) There is no statistically significant difference in response time when we move a station to the north from the Central or South stations.

3) Relocating to the North station does not yield any significant benefits since the relocation scenarios as a whole do not yield statistically significant improvement in response time.

5 Conclusion & Future Work

Our analysis shows that relocation to Far North results in faster response times than that to the Near North. We also conclude that it is better to move a Central ambulance up North rather than from the South, and that relocating one ambulance from Central to Far North station particularly reduces response times the most, by 4%. Therefore, scenario 4 (ambulance: 1 Far North, 2 Central, 1 South) is the most optimal distribution of the 4 ambulances as it reduces the average response time by 4% compared to the baseline S0.

Our model includes some limitations that should be addressed when considering results. First, we assume traffic level to be constant throughout non-rush vs rush hour, unaffected by different events that may have occurred on each day such as accidents, construction, or road blockage. We also are dealing with data that has inherent error introduced, since the location data is randomized due to HIPAA protection. This leads to some marginal error in the distance calculation. We are also limited to a small sample size that differ for each region, and assume average conditions for response time using the best-guess Google API.

For future work to address these limitations, we can try adding time of day as a random effect which could provide deeper insight into temporal variation in the response times. We can model this using a continous smooth spline approach using mgcr or poly functions in R to capture the nonlinear daily trends without overfitting. We can also take a Bayesian approach to improve interpretability, particularly in regions with fewer calls. Through partial pooling and uncertainty propagation, the use of this approach would allow information sharing across districts while accounting for data sparsity. Finally, a hierarchical modeling approach could be implemented to account for further subdivisions in call type and urgency and this would allow our model to distinguish between the emergency severity levels and incorporate medical criteria for the urgency of each call which could change how we allocate the ambulances across the systems.
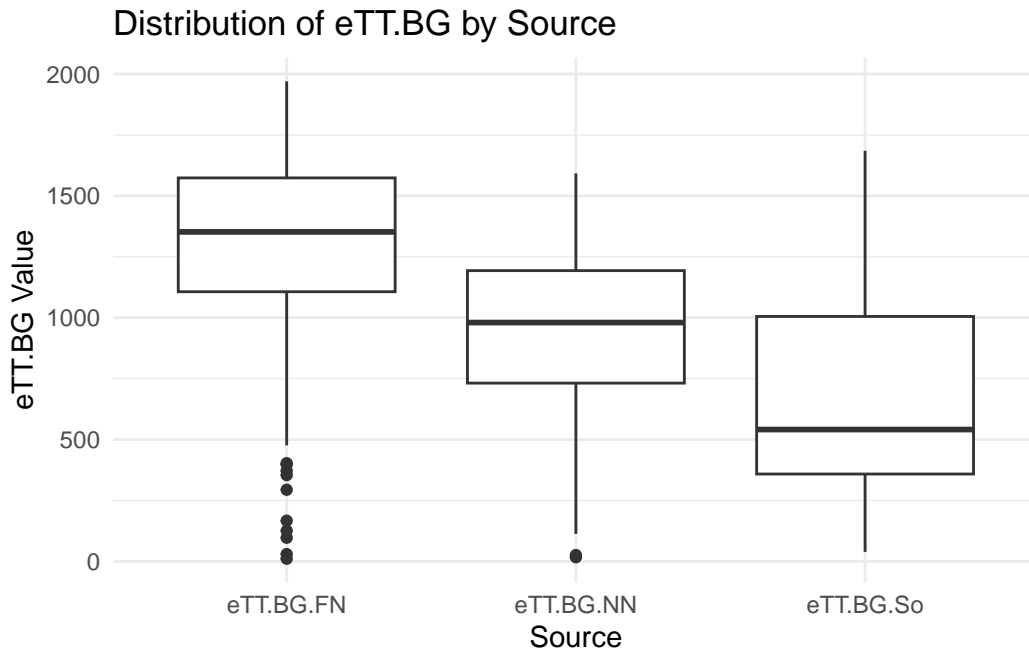
6 Appendix

Exploratory Data Analysis

Figure 1

Distribution of eTT.BG by Source

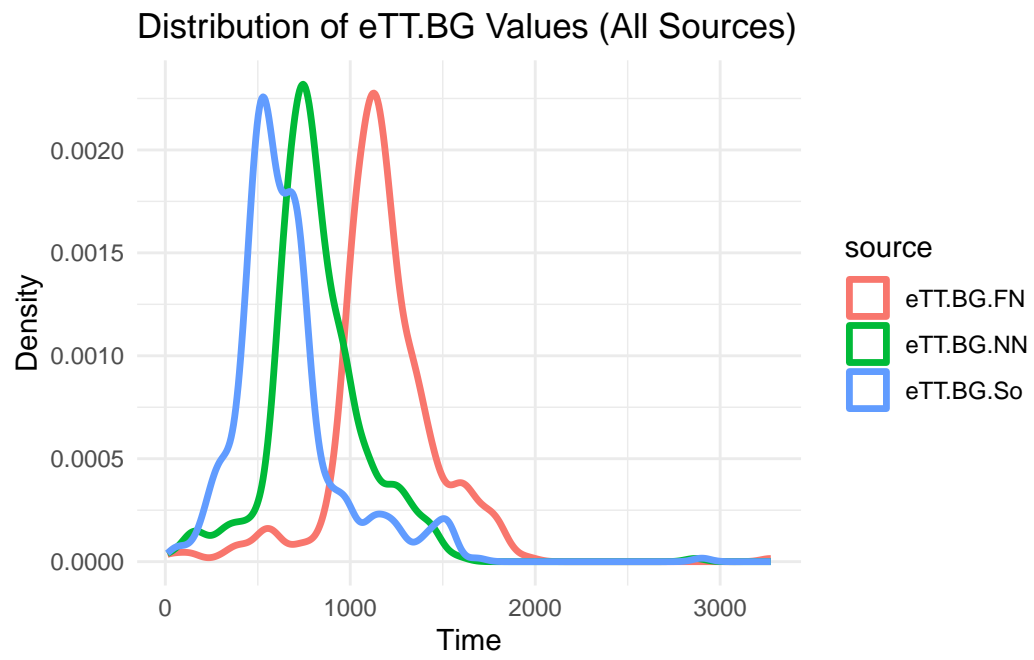**Figure 2**

Distribution of eTT.BG Values (All Sources)

Conflict Duration Distribution for North and Central Station
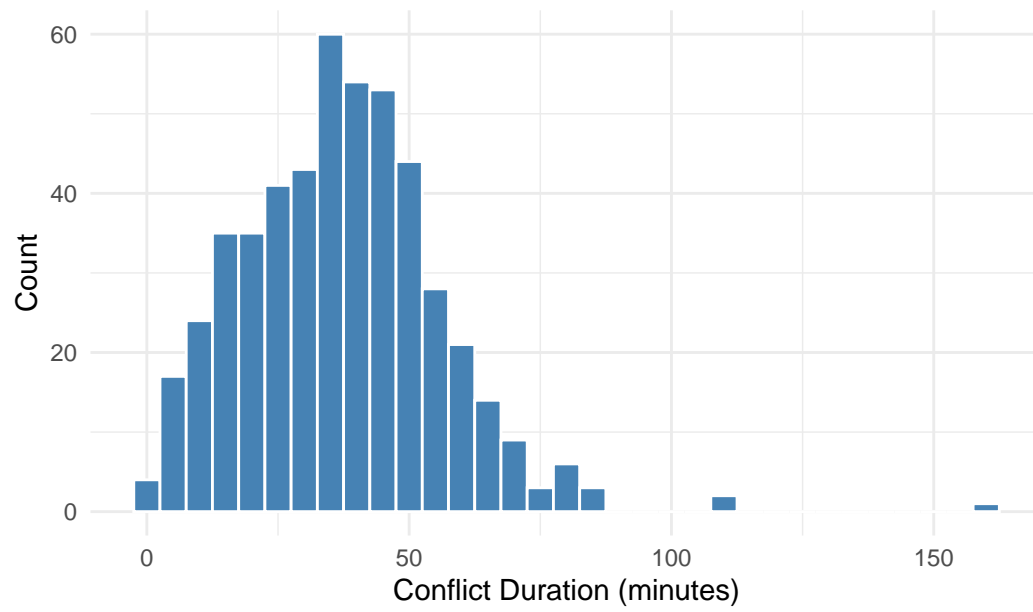
Figure 4

Conflict Duration Distribution for South Station

Figure 5

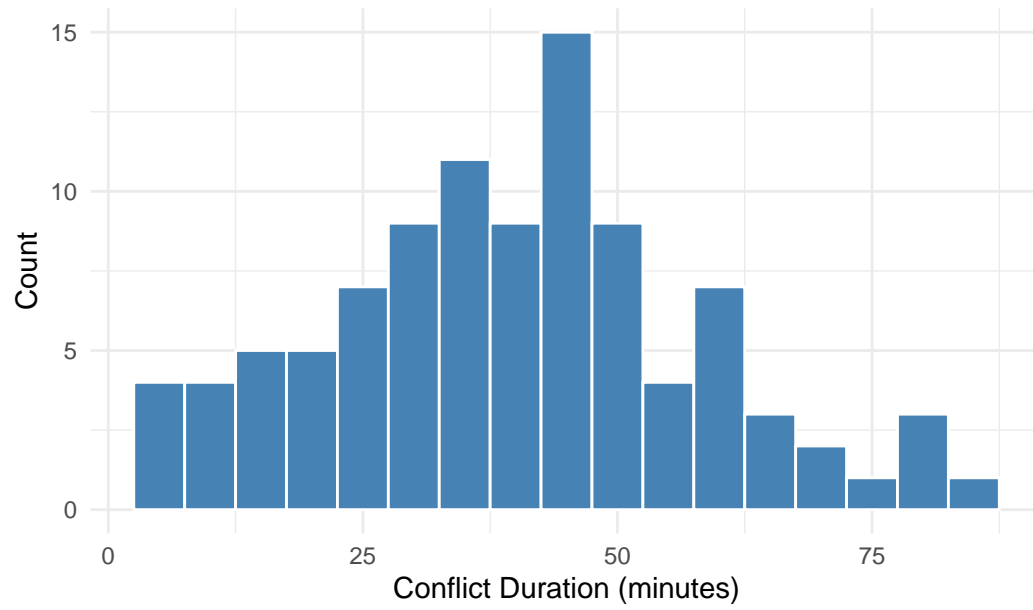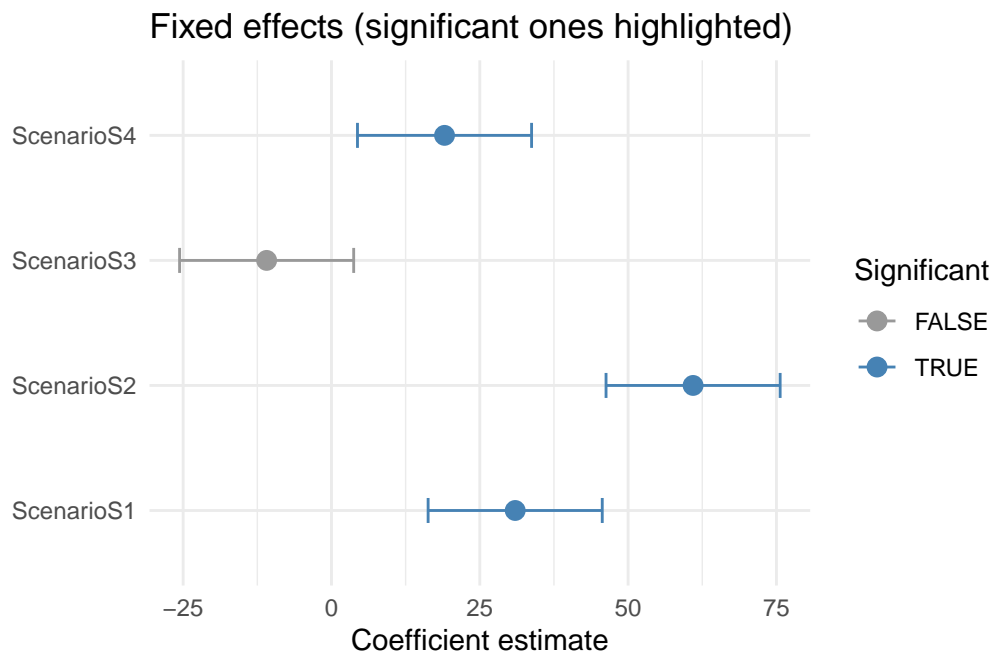
Fixed effects (significant ones highlighted)

Modeling

Table 1

Scenarios	S0 (Current)	S1	S2	S3	S4
Far North	0	0	1	0	1
Near North	0	1	0	1	0
Central	3	3	3	2	2
South	1	0	0	1	1

Table 2

Linear Mixed Model for EstTravelTime

Predictors	Estimates	std. Error	CI	p
	EstTravelTime			
(Intercept)	112.04	8.75	94.85 – 129.22	**<0.001**
Scenario [S1]	-5.35	6.48	-18.07 – 7.36	0.409
Scenario [S2]	-4.99	6.74	-18.22 – 8.24	0.459
Scenario [S3]	-3.37	6.38	-15.90 – 9.17	0.598
Scenario [S4]	-3.01	6.41	-15.60 – 9.59	0.639
Distance	0.05	0.00	0.04 – 0.05	**<0.001**
rush hour ind	11.86	13.32	-14.29 – 38.01	0.373
Random Effects				
σ^2	3068.93			
$\tau_{00 \ CallID}$	4157.26			
ICC	0.58			
N $_{CallID}$	151			
Observations	755			
Marginal R^2 / Conditional R^2	0.928 / 0.969			

Figure 6

Residuals vs Fitted (No–Change Calls Removed)

Figure 7
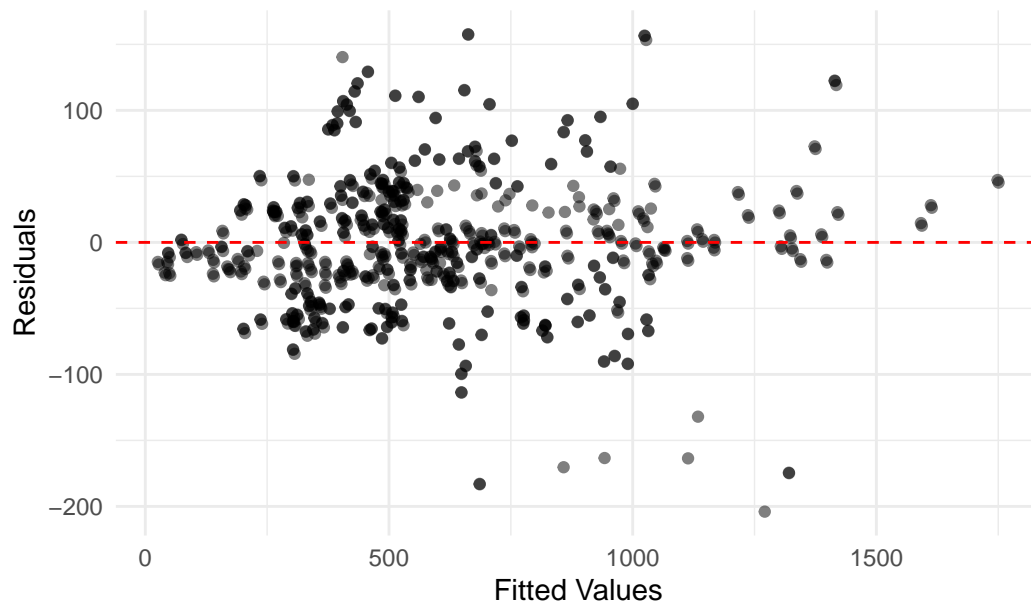
QQ Plot of Residuals (No–Change Calls Removed)

Figure 8
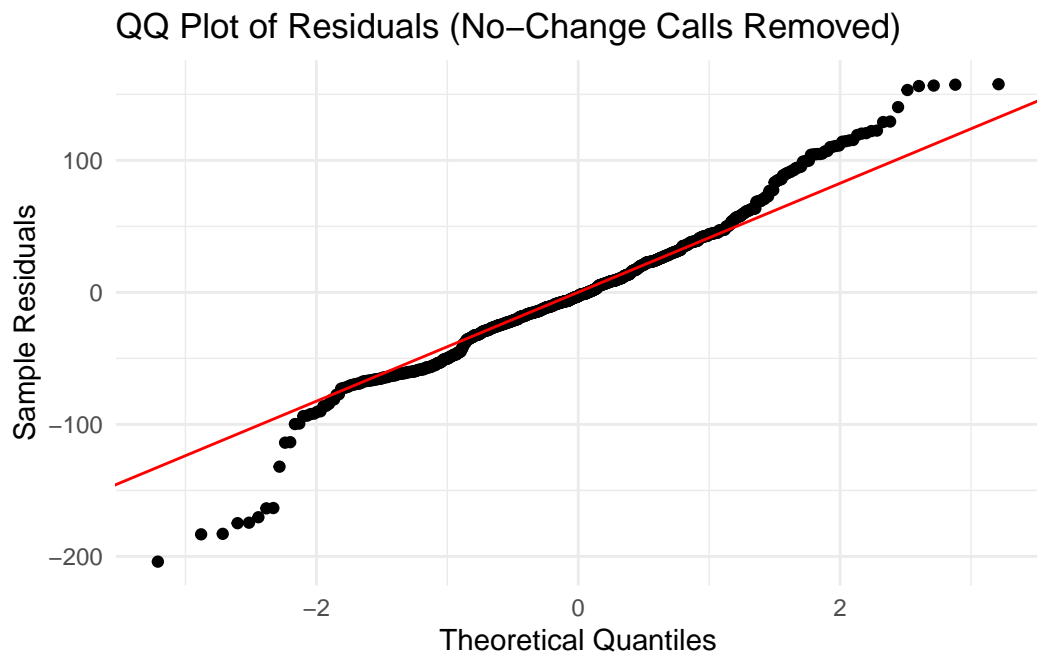

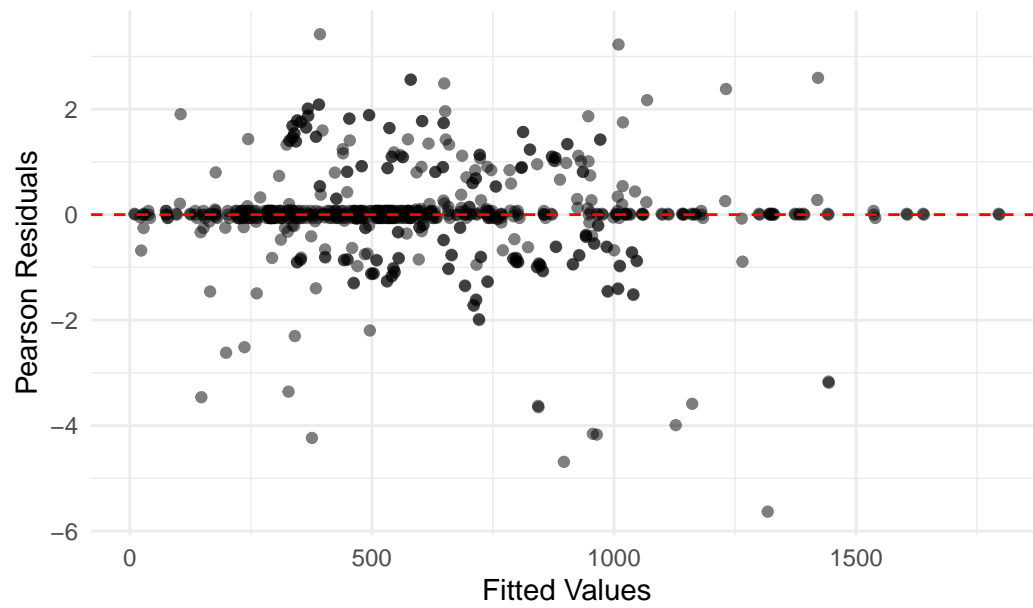Residuals vs Fitted (varIdent by Scenario)
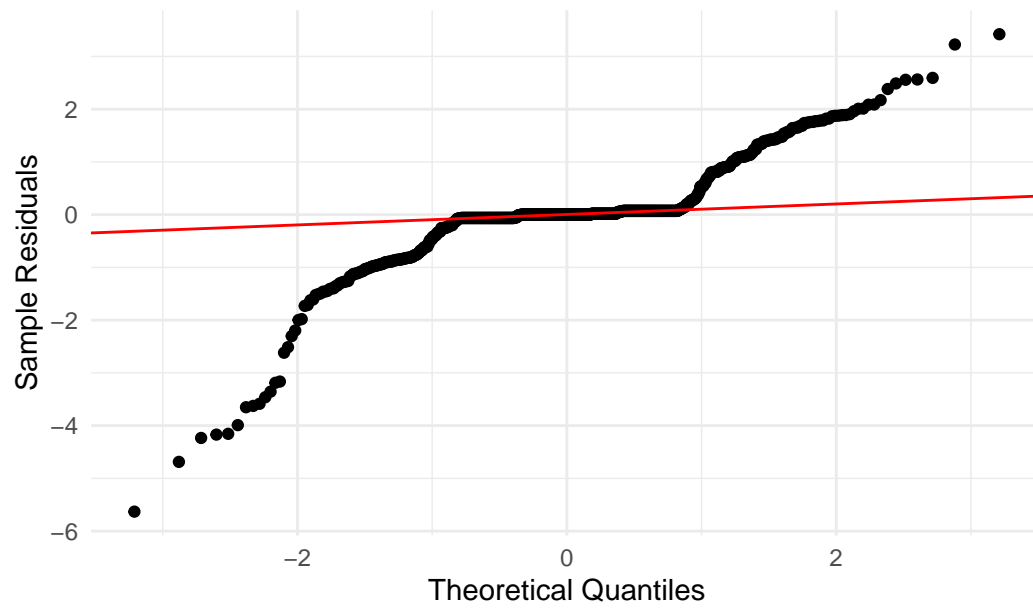
Figure 9

QQ Plot of Residuals (varIdent Model)

Figure 10

Histogram of Residuals (varIdent Model)