Case 3: Vance County EMS

Steph Reinke, Olivia Fu, Srika Gopal

October 20, 2025

1 Background and Motivation

Emergency medical service (EMS) response times can drastically impact patient outcomes. Vance County, North Carolina is a growing county that is interested in evaluating their current EMS ambulance placements. The county has three major regions, the North, Central, and the South. The population is mostly concentrated in the city of Henderson, located in the Central region. There are currently 2 EMS stations, 1 in the South, and 1 in Central. The southern station has 1 ambulance, and the central station has 3. The North does not currently have a station. Thus Vance County is interested in exploring if moving an ambulance to the North would be beneficial. There are two potential locations for an ambulance to be stationed in the North: the Near North and the Far North. We were given five scenarios for ambulance placement:

- Scenario 0: 1 in South, 3 in Central (Current Placement)
- Scenario 1: 0 in South, 3 in Central, 1 in Near North
- Scenario 2: 0 in South, 3 in Central, 1 in Far North
- Scenario 3: 1 in South, 2 in Central, 1 in Near North
- Scenario 4: 1 in South, 2 in Central, 1 in Far North

Research question: Where should the ambulances be stationed to best serve Vance county?

2 Data and Exploratory Analysis

To answer this question we examined a dataset with real EMS call data. Select information was withheld or changed, such as dates and addresses for HIPAA compliance. However for the purposes of our analysis those changes are not relevant. The data has information for each call and API pulled travel times from Google Maps. We were provided with four types of estimates from Google Maps (best guess, optimistic, pessimistic, and unadjusted) for each station including the two proposed Northern stations to each call.

We started by exploring the distribution of calls throughout Vance County. We plotted call locations on a map of Vance County and colored the points according to the response time (Figure 1). The points are densely concentrated in the Central region around the city of Henderson. The points become more sparse the further they are from Henderson, resulting in sparsely concentrated points in the far North and far South. Response times are shorter in the Central region and tend to be longer in the North and the South.

We were provided with four different Google Maps travel times. Ultimately we chose to use BG (best guess) estimates because it had the lowest bias (Table 1). This can be seen visually (Figure 2) because the interquartile range (the box) is the most centered at 0 for BG in comparison to the other estimates.

In order to simulate EMS calls with the given data, we had to determine a rule for which ambulance to dispatch from which station. First we use the availability matrix to detect overlapping calls with earlier dispatches. We then adjust each station's available vehicles by subtracting those already assigned to conflicts. Then from the stations with vehicles still available, dispatch from the one offering the shortest estimated travel time to the call.

Once we determined our dispatch rule, we were able to calculate estimated travel times for the calls according to the different scenarios. Looking at different intervals of estimated travel times, we can see that Scenario 3 appears to have the lowest (fastest) estimated travel times. This suggests that Scenario 3 may be the optimal choice.

3 Model Rationale, Implementation, and Evaluation

3.1 Rationale & Selection

We selected a two-stage modeling approach: a logistic regression to model whether estimated travel time changes from the baseline scenario, followed by a linear mixed-effects model to quantify how travel time changes when it does differ. This approach is justified both by the data structure and the nature of our dispatch system. In our simulation, many calls have identical estimated travel times across all four proposed scenarios and the baseline (S0). For example, calls located in the central often have the same optimal dispatch station under all scenarios, resulting in no variation. Including these repeated values directly in a continuous-response model would violate normality and constant variance assumptions due to the discrete mass at zero differences.

To address this, our first model is a logistic regression predicting whether estimated travel time differs from baseline (change vs. no change). This allows us to estimate the probability that each scenario leads to a change in dispatch outcome.

Conditional on a change occurring, we then model the magnitude of that change using a linear mixed-effects model. The response variable is the difference in estimated travel time relative to baseline. We include a Scenario*Region interaction to assess both overall scenario effects and whether their impacts differ across geographic regions. Because each call contributes multiple observations (one per scenario), there is clear grouping in the data. Therefore, we include a random intercept at the call level to account for within-call correlation.

Exploratory residual analysis revealed heteroskedasticity across regions, so we incorporate region-specific residual variances using a varIdent variance structure. Attempts to also model variance by scenario resulted in singular fits, likely due to limited variability or nesting of scenarios within call-level random effects. A competing GLS model without random effects and with Scenario*Region variance structure had substantially worse AIC/BIC values (Table 1), which supports that accounting for call-level correlation and regional heterogeneity provides a better fit.

3.2 Implementation

For the logistic model, we created a binary indicator of whether estimated travel time changed from baseline and fit a generalized linear model using the glm() function in R with a binomial family and scenario as the predictor.

For the linear mixed-effects stage, we retained only rows where travel time differed from baseline and modeled the continuous difference using the lme() function in the nlme package. We specified a random intercept for each call (random = ~1 | row_val) and allowed the residual variance to differ by region using weights = varIdent(form = ~1 | region).

3.3 Evaluation

To evaluate whether the model provided a good fit for the data, we examined multiple residual plots (Figure 4). The residuals are normalized to account for the model's variance structure. In the residuals vs. fitted values plot, the residuals are randomly scattered around zero without clear patterns or clustering, which suggests that the model's mean structure is appropriate and that the variance has been reasonably modeled. The horizontal layering is expected due to the discrete set of fitted values from the scenario–region combinations.

Residuals plotted by scenario suggest some differences in spread, particularly for S1/S2 relative to S3/S4. While this indicates possible scenario-level heteroskedasticity, the model could not support this additional

structure without convergence failures. The residuals by region plot shows some remaining variance differences across regions, but the variation is improved compared to the model without variance adjustment.

We also examined the Q-Q plots to assess the normality assumption (Figure 5). For the random effects, the points closely align the diagonal line with only minor deviations in the lower tails, suggesting the normality assumption for random effects is reasonable. For the residuals, the Q-Q plot shows deviations in the tails, which may be due to repeated identical travel time estimates for certain scenarios or the discrete nature of the response variable. The central portion aligns closely with the theoretical line, indicating that the normality assumption is approximately satisfied for most observations.

4 Results

From the logistic regression results (Table 2), we estimated the probability that a call's estimated travel time would change relative to the baseline for each scenario. As shown in Table 3, 25.4% of calls exhibited a change under Scenario 1, 23.6% under Scenario 2, 10.8% under Scenario 3, and 9.3% under Scenario 4. This indicates that the majority of calls are unaffected by ambulance redistribution across all scenarios, which is an important consideration when evaluating the overall effectiveness and operational impact of each proposed strategy.

To quantify the magnitude of change in estimated travel time for calls that were affected, we computed estimated marginal means using the emmeans() function in R. As shown in Table 5, Scenario 3 has the lowest estimated marginal mean of –41.07 seconds, indicating that it reduces estimated travel time by approximately 41 seconds on average compared to the current setup. However, this effect is not statistically significant at the 0.05 level (p = 0.1138). Thus, while Scenario 3 appears to perform best among the four proposals, the overall improvement is not strong enough to conclude a statistically significant benefit at the county-wide level.

When evaluating results by region (Table 6), we observed substantial variation across regions. In the Central and South regions, all scenarios result in positive estimated marginal means, indicating longer travel times compared to the baseline. This is expected, as the scenarios reallocate an ambulance from the Central or South region to the North. In contrast, the North region experiences substantial improvements, with Scenario 3 achieving the largest reduction in estimated travel time (-477.7 seconds, p = 0). Scenario 1 yields a similar reduction, while Scenarios 2 and 4 perform less favorably, suggesting that moving an ambulance to the near north site is more effective than moving it farther away.

Overall, Scenario 3 provides the greatest improvement in the region it is intended to benefit (North), but its benefits at the county level are limited by negative impacts on the Central and South regions and the relatively small proportion of calls that actually experience a change.

5 Limitations and Conclusion

There are several shortcomings of our analysis, some of which are due to model flaws while others are due to data limitations. First of all, our model stratified the random residual vs. fitted scattering, but there is some slight ambiguity in our residuals plotted by scenario that leaves a possibility of scenario-level heteroskedasticity. While our model has reasonably met the normally distributed residuals assumption via the QQ plot, there is still some deviation in the tail ends of the plot which could point to model weakness in terms of normality.

Additionally, our model does not account for certain outside factors that were omitted from the provided data and could not be extracted without heavy assumption. Factors like traffic, rush hours, and road closures which very well could affect ambulance response times are not accounted for by our model or analysis. Additionally, variance in ambulance response times dependent on the month/date of the year is very possible yet not a feature of our model. Furthermore, EMT and ambulance staffing patterns which is an additional factor that could affect an ambulance's ability to report to a call even if the vehicle itself is available is something that could not be assumed from the provided data.

Other shortcomings of our analysis beyond unaccounted factors include ETT input limitations, variance modeling, and dispatch rule simplicity. The estimated travel time (ETT) inputs are based on single best-guess estimates generated from Google Maps data, so this does not take nuances in travel time due to aforementioned factors into consideration. Our variance modeling was only structured around the regional differences for scenario reallocation, even though other factors like call urgency (emergency vs. non emergency) could affect variability. Lastly, our dispatch rule is sufficient, yet simplistic and lacks consideration of nuances and additional factors (as mentioned above). While dispatching based on a simple queueing system model was ideal for analysis, this is not indicative of real-life ambulance dispatch scenarios. Real-life dispatching can be wildly unpredictable, if not dependent on many other factors like road conditions and staffing patterns.

To improve our model for the future, we might consider fitting additional models that account for additional factors of complexity if provided, such as road conditions, staffing patterns, and traffic. We would ideally also incorporate the given urgency of calls (emergency vs non-emergency) to update our dispatch rule. We could also add contrasts to our estimated marginal mean travel time calculations. This would provide a more informative model and analysis that not only accounts for optimal ambulance reallocation by region, but also by how dire the call was.

Based on the above analyses and our modeling results, we conclude that the overall optimal reallocation resulting in a reduction of ambulance response time is Scenario 3. Scenario 3 is to move an ambulance from the Central station to the Near North station in Vance County. This reallocation results in a % reduction in marginal mean estimated travel times from the baseline of Scenario 0 (the current ambulance allocation in Vance County). We see the lowest overall marginal mean estimated travel times for Scenario 3, though, as discussed above, the optimal Scenario for Central and North stations evaluated individually would be Scenario 1. With this change in ambulance allocation to Scenario 3, Vance County can expect to see a decrease in marginal mean ETT of _____ min, which is a substantial amount when discussing the urgency of emergency response situations whose outcomes are dependent on minutes, perhaps even seconds.

6 Appendix

Observed Travel Times of Vance County EMS in minutes
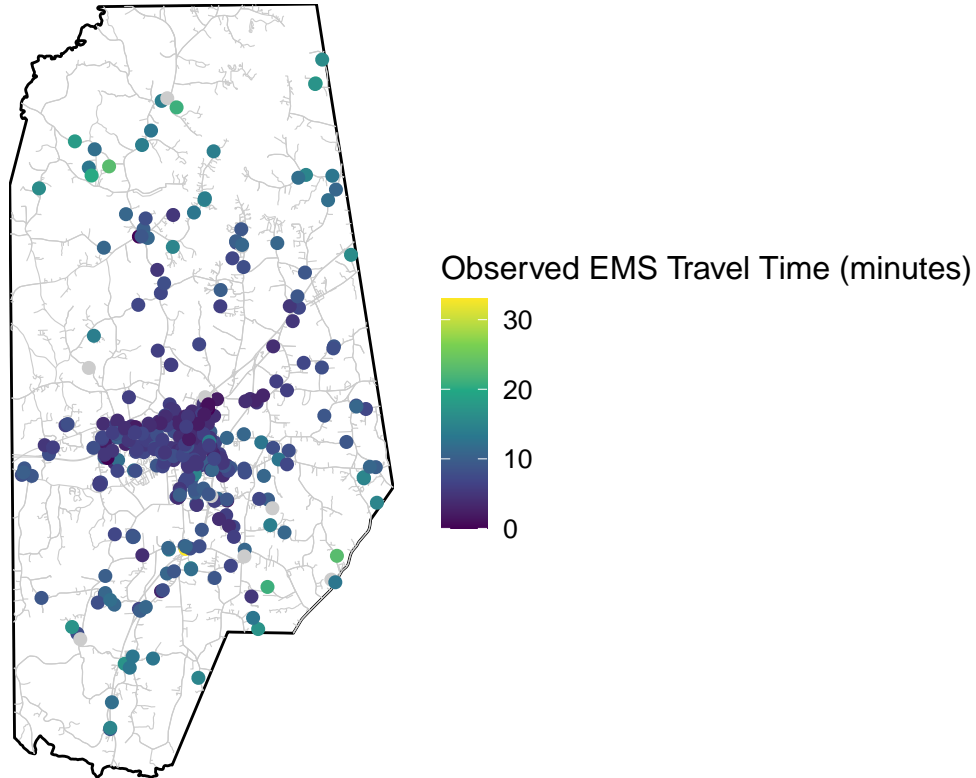

Figure 1: EMS call location and response times across Vance County

Table 1: Metrics of different travel time estimators shows BG is optimal

Estimator	MAE	RMSE	Bias
est_BG	116.0021	192.2824	-1.737395
est_Op	112.3403	186.4617	16.277311
est_Pe	134.0651	214.2580	-41.674370
est_UA	116.4391	192.1155	-9.140756

Table 2: Model output for logistic model of scenarios

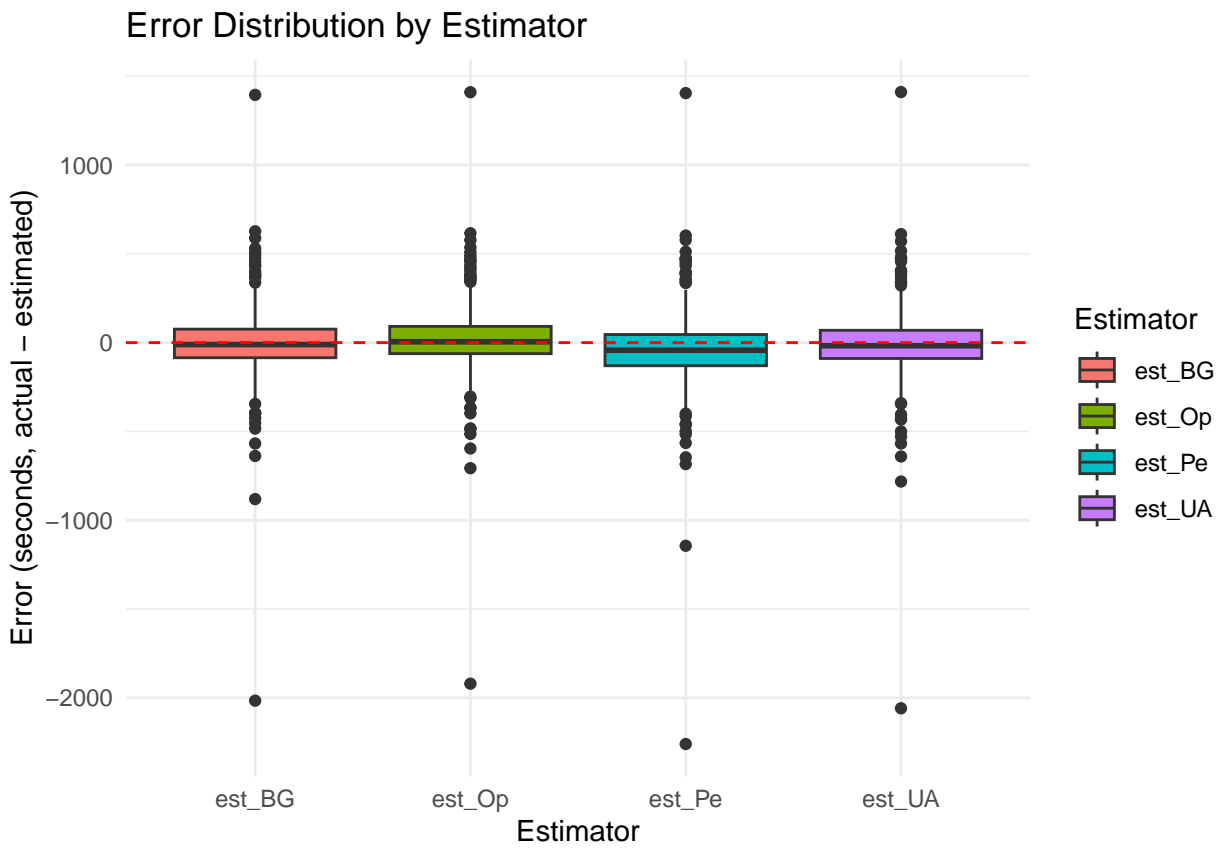| term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.077	0.104	-10.400	0.000	-1.284	-0.877
ScenarioS2	-0.099	0.148	-0.667	0.505	-0.390	0.192
ScenarioS3	-1.037	0.179	-5.810	0.000	-1.394	-0.693
ScenarioS4	-1.195	0.186	-6.413	0.000	-1.568	-0.836

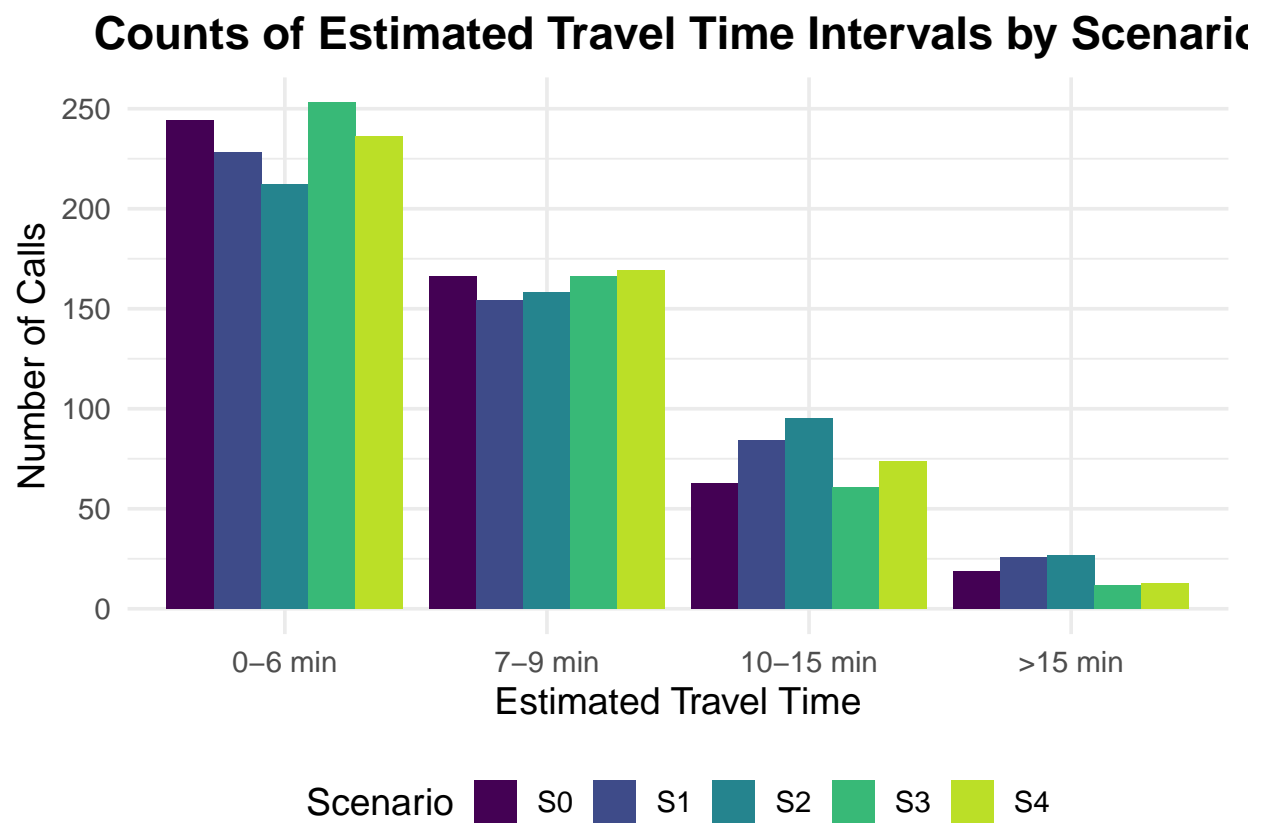Figure 2: Comparison of different travel time estimators

Figure 3: Estimated travel times intervals by scenario

Table 4: Estimated marginal means by scenario

Scenario	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
S1	-21.09875	23.21680	140	-66.99961	24.802115	-0.9087706	0.3650330
S2	64.64958	24.13432	140	16.93472	112.364433	2.6787398	0.0082750
S3	-41.06652	25.80494	140	-92.08427	9.951237	-1.5914207	0.1137704
S4	110.67488	26.90801	140	57.47630	163.873455	4.1130835	0.0000663

Table 3: Predicted probabilities of differences from baseline

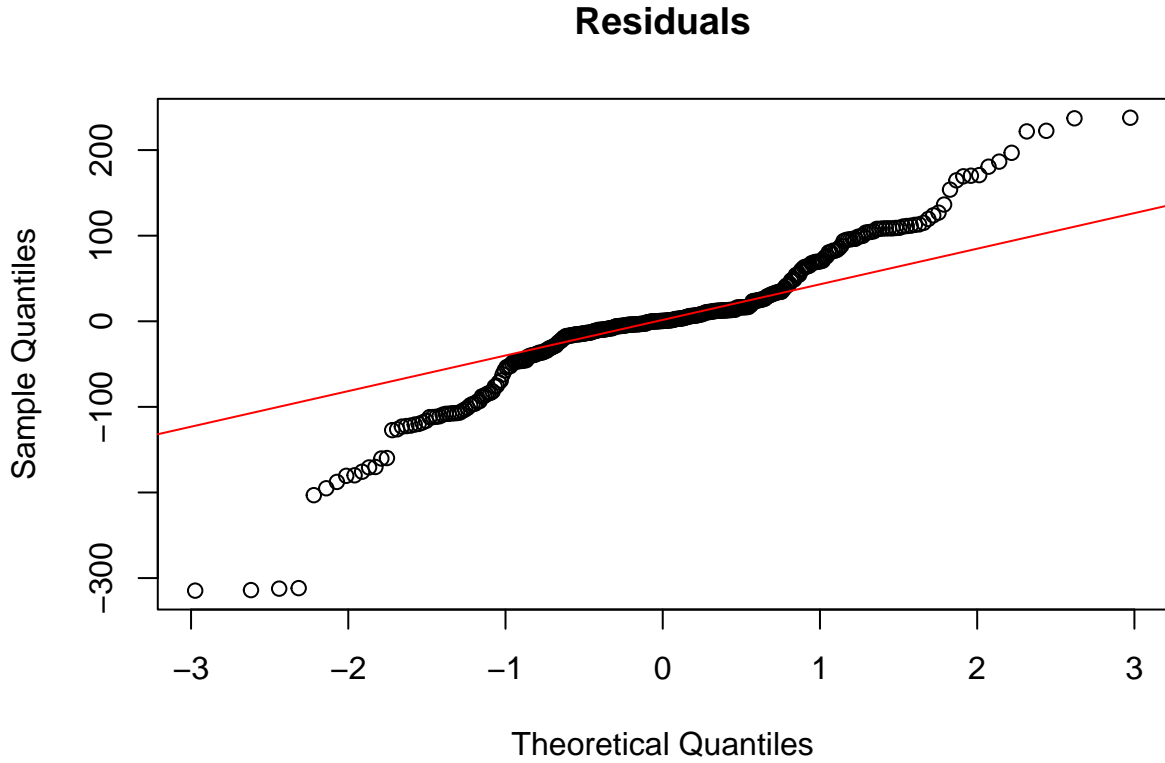| Scenario	predicted_prob
S1	0.254
S2	0.236
S3	0.108
S4	0.093

Figure 4: QQ plot for residuals roughly follows diagnol line
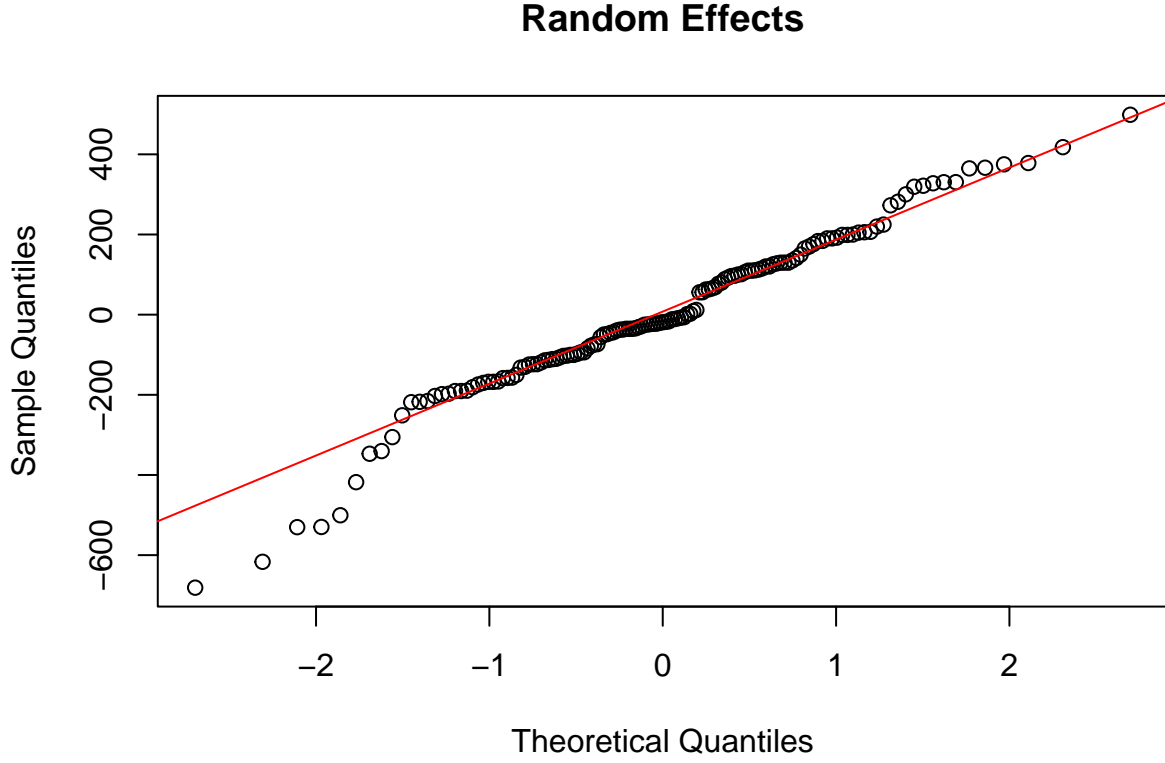
Random Effects

Figure 5: QQ plot for random effects follows diagnol line

Table 5: Estimated marginal means by scenario within each region

| Scenario	region	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
S1	Central	93.5476	39.01992	142	16.41259	170.6826	2.397432	0.0178108
S2	Central	164.4503	40.05862	142	85.26197	243.6386	4.105241	0.0000678
S3	Central	120.9151	41.24058	142	39.39030	202.4400	2.931946	0.0039277
S4	Central	237.5040	41.83505	142	154.80406	320.2040	5.677155	0.0000001
S1	North	-462.4400	51.70835	140	-564.67019	-360.2098	-8.943236	0.0000000
S2	North	-292.2757	54.59975	140	-400.22233	-184.3291	-5.353060	0.0000003
S3	North	-477.7068	52.73204	140	-581.96083	-373.4527	-9.059137	0.0000000
S4	North	-292.8761	56.06538	140	-403.72035	-182.0318	-5.223831	0.0000006
S1	South	305.5962	25.59043	140	255.00251	356.1898	11.941814	0.0000000
S2	South	321.7741	25.61963	140	271.12275	372.4255	12.559671	0.0000000
S3	South	233.5921	38.87932	140	156.72556	310.4586	6.008131	0.0000000
S4	South	387.3967	40.28485	140	307.75135	467.0420	9.616435	0.0000000
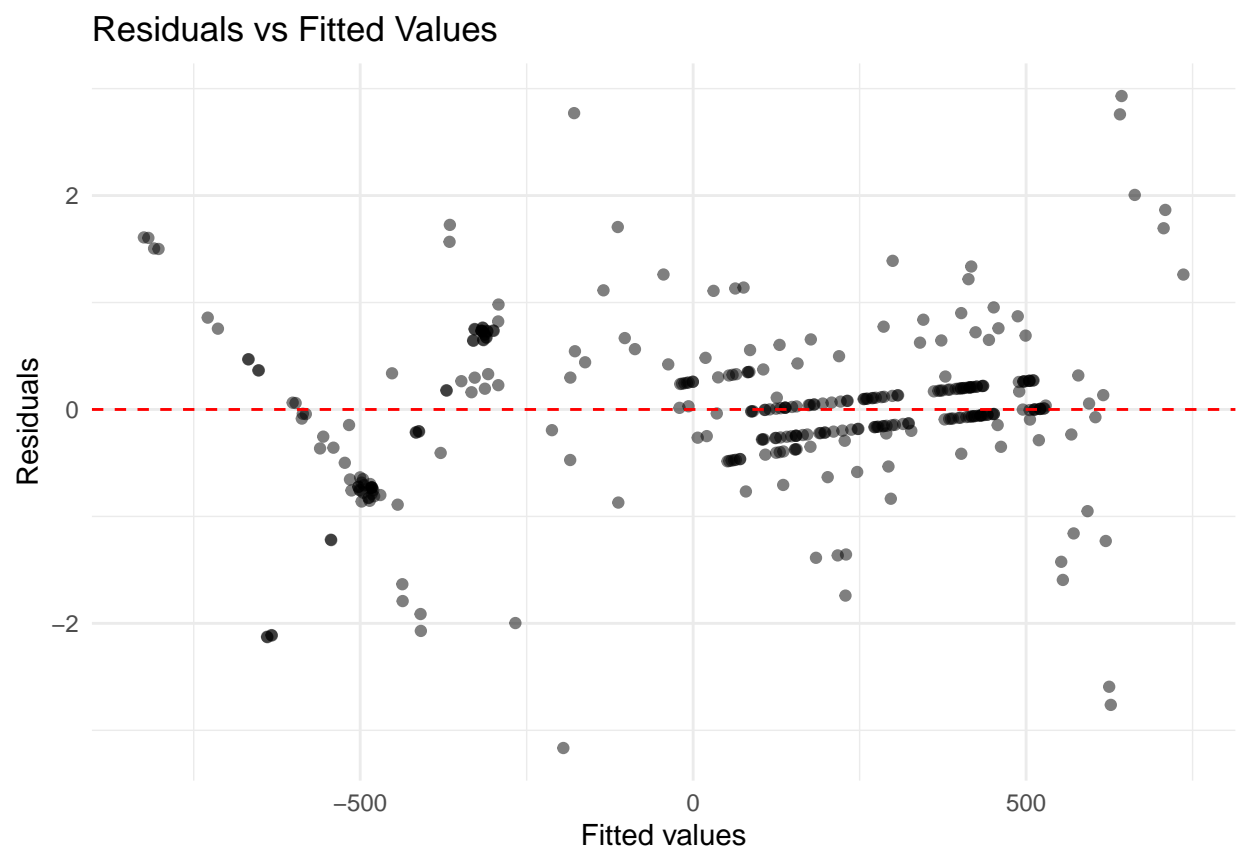
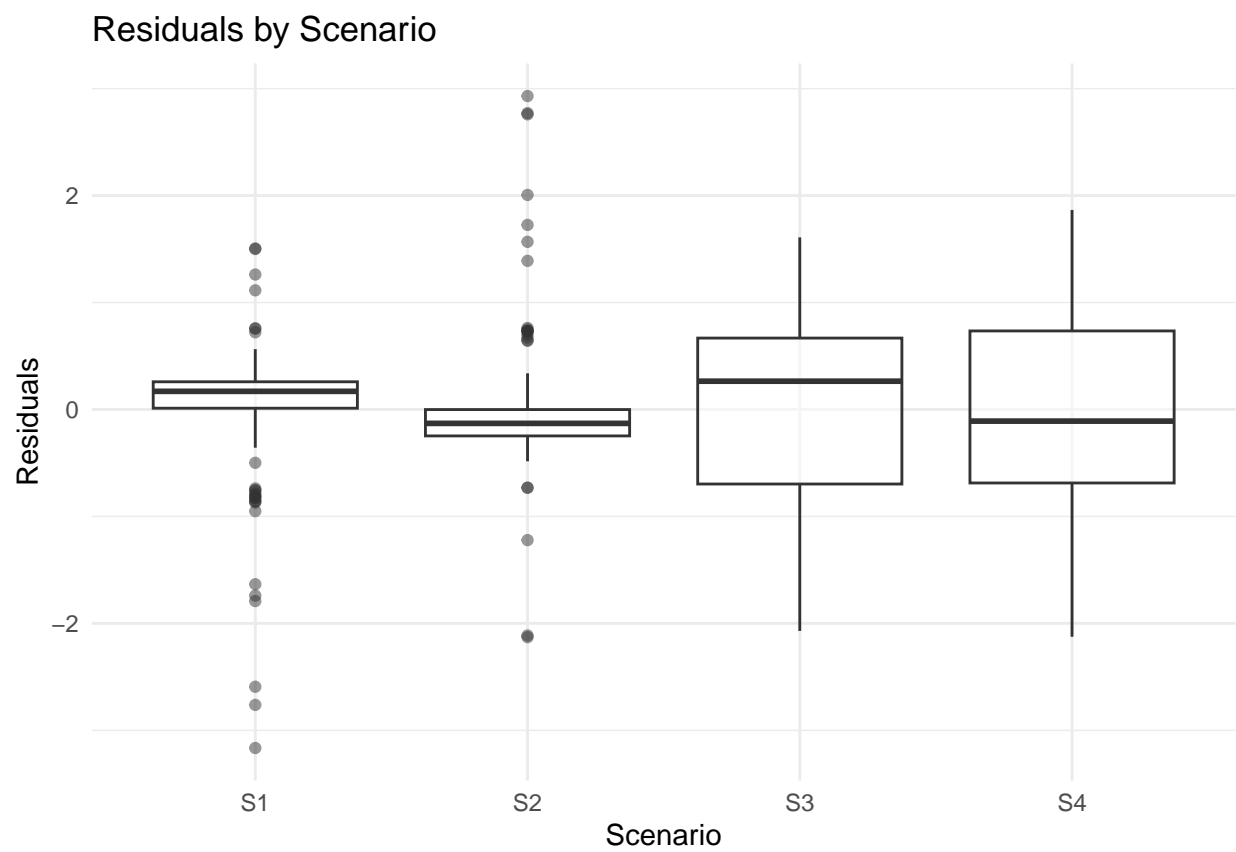Figure 6: Residuals vs fitted values are roughly randomly scattered

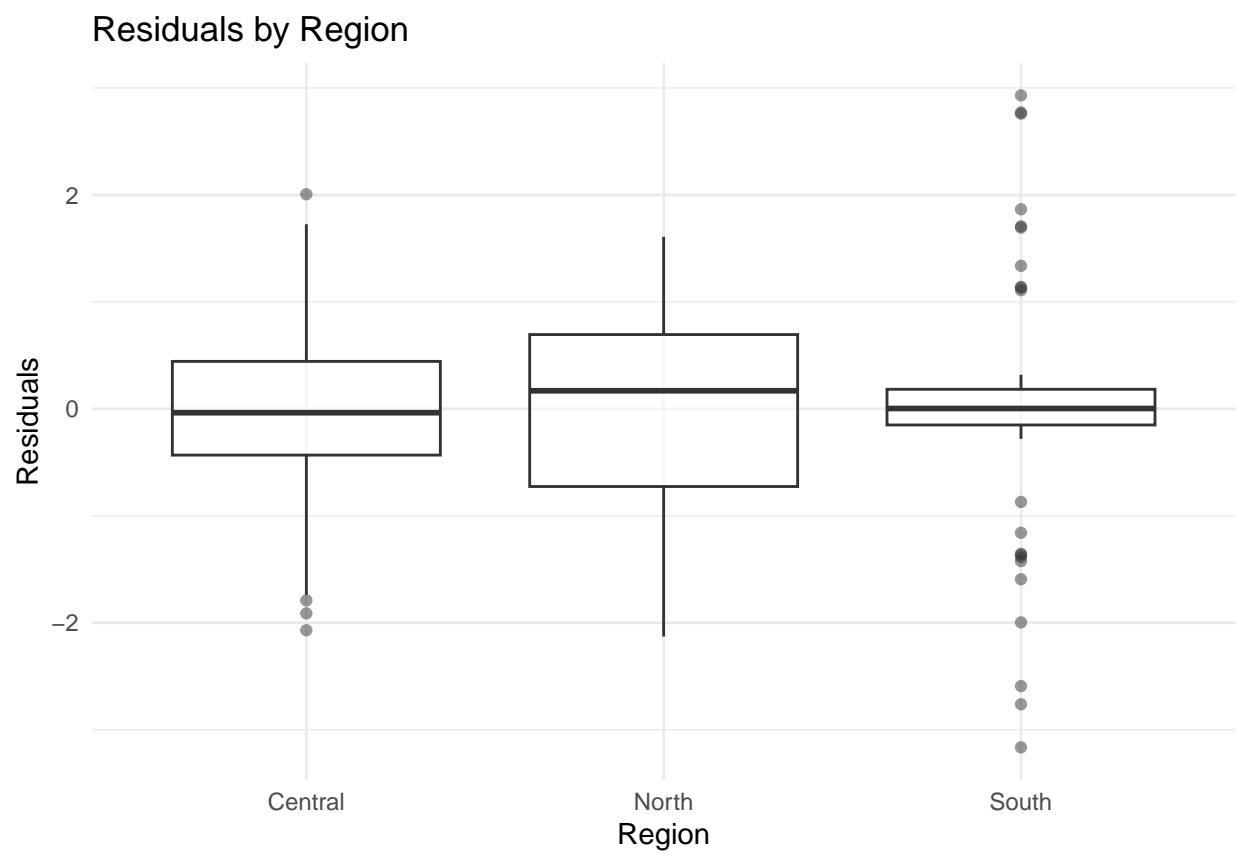Figure 7: Boxplot of residuals by scenario shows variation across scenarios

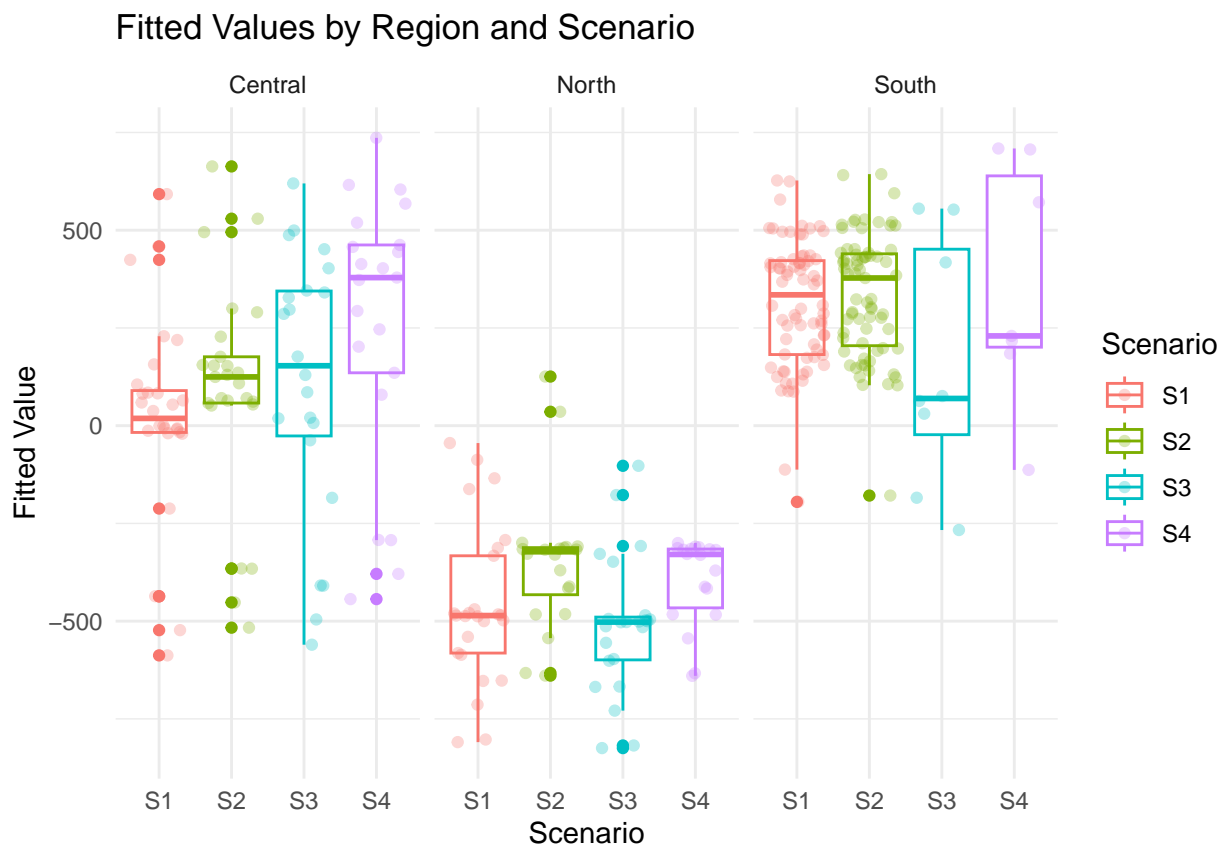Figure 8: Boxplots of residuals by region show variantion across regions

Figure 9: Boxplots of fitted values by region and scenario show scenario variations