Sta 440 Case 3

Srika Gopal, Olivia Fu, Steph Reinke

2025-10-01

1. Background

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

Finding the optimal locations for ambulances is vital because in the medical world EMS arriving seconds earlier can be the difference between life and death.

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

2. Data and Model Fit

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.
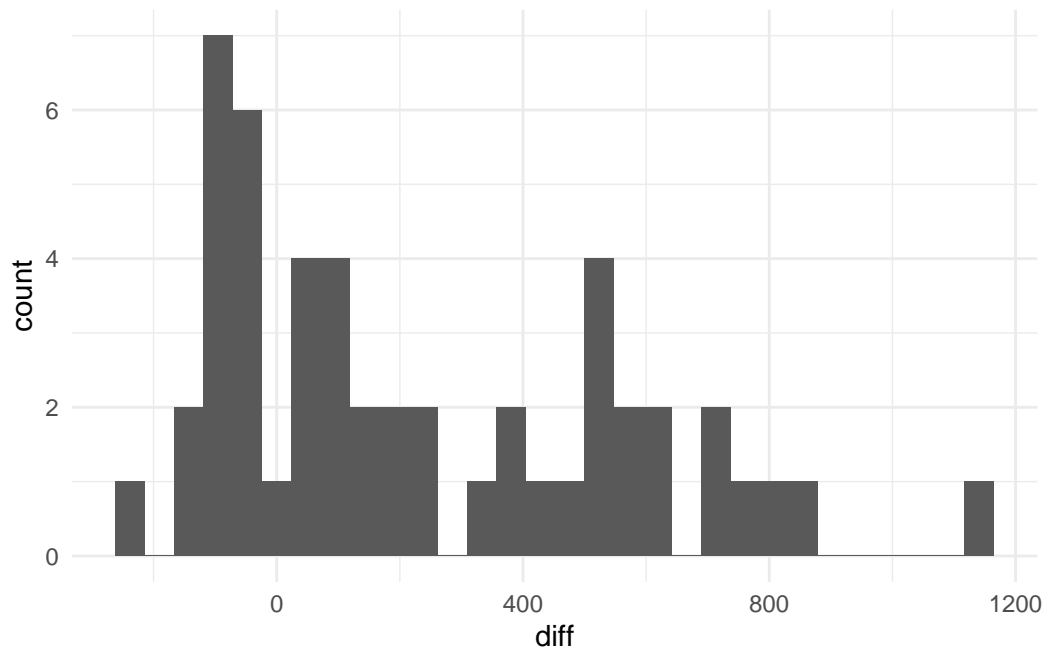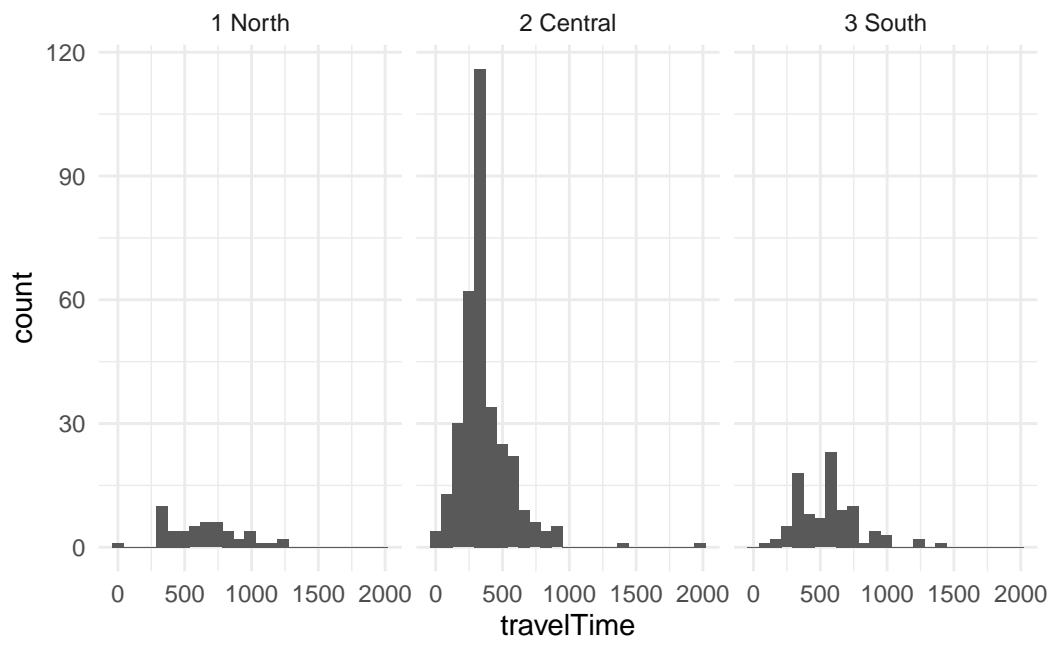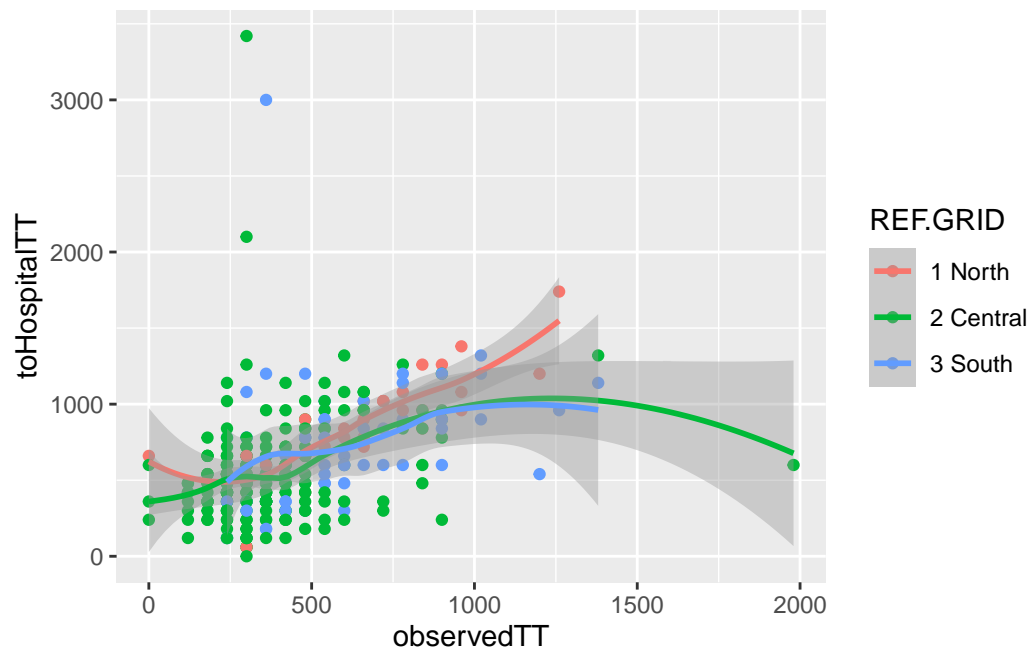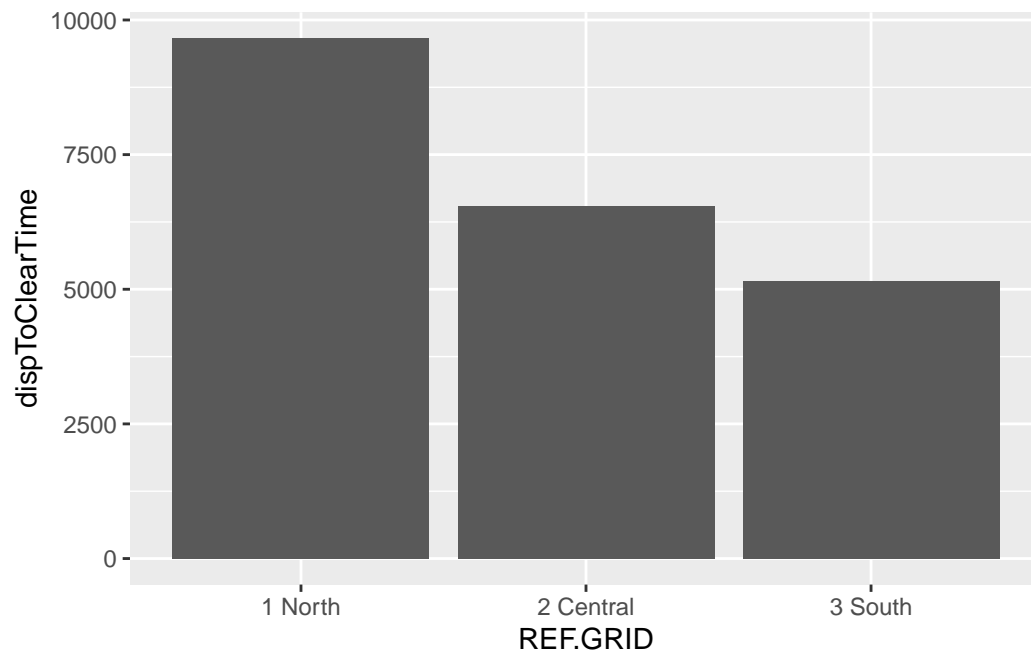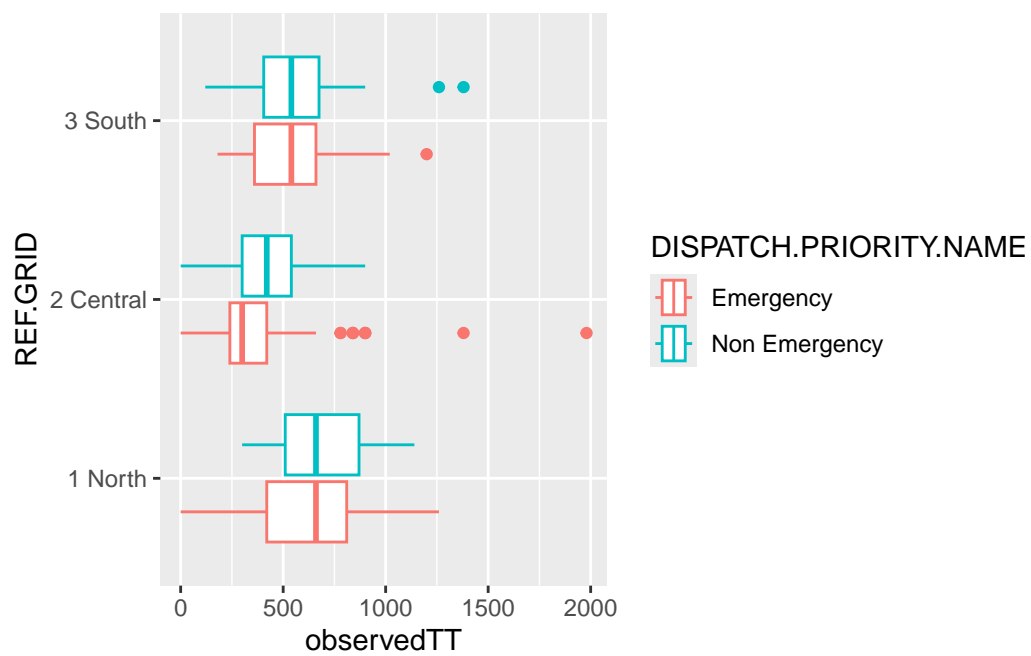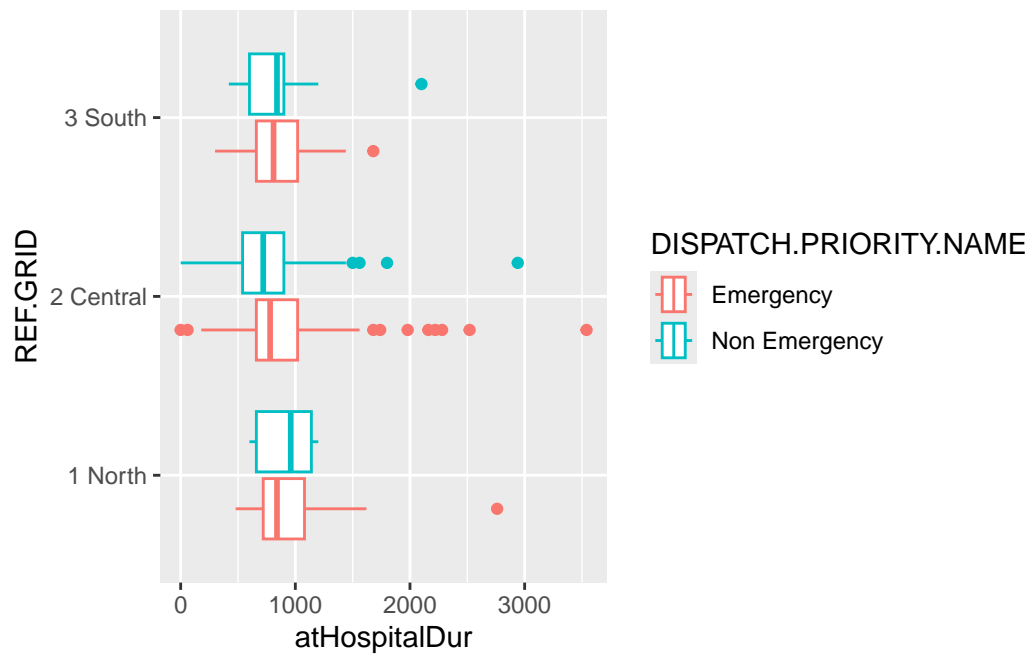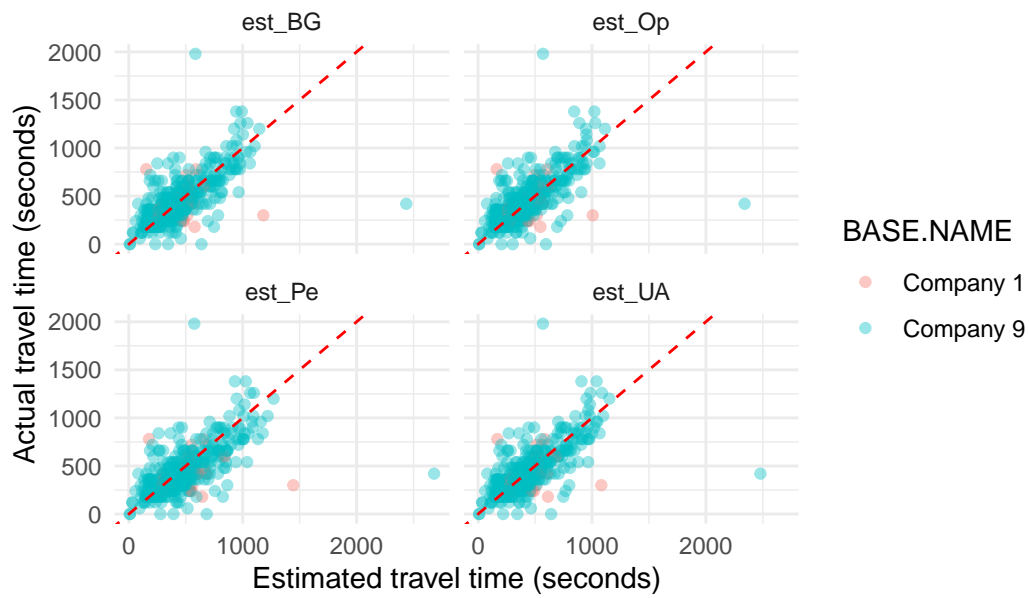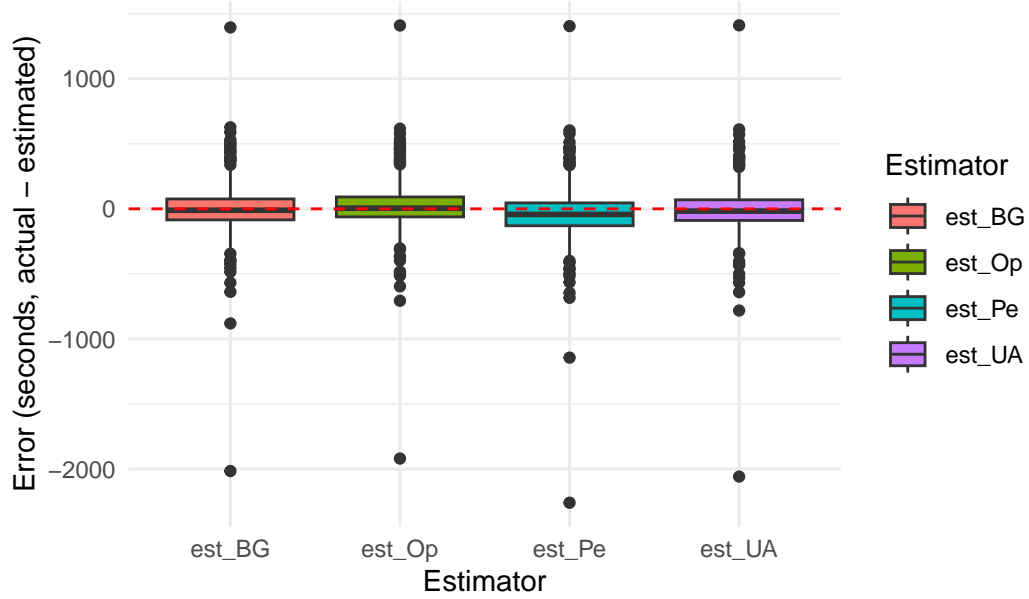
```
# A tibble: 4 x 4
  Estimator   MAE  RMSE   Bias
  <chr>     <dbl> <dbl>  <dbl>
1 est_BG     116.  192.  -1.74
2 est_Op     112.  186.  16.3
3 est_Pe     134.  214. -41.7
4 est_UA     116.  192.  -9.14
```
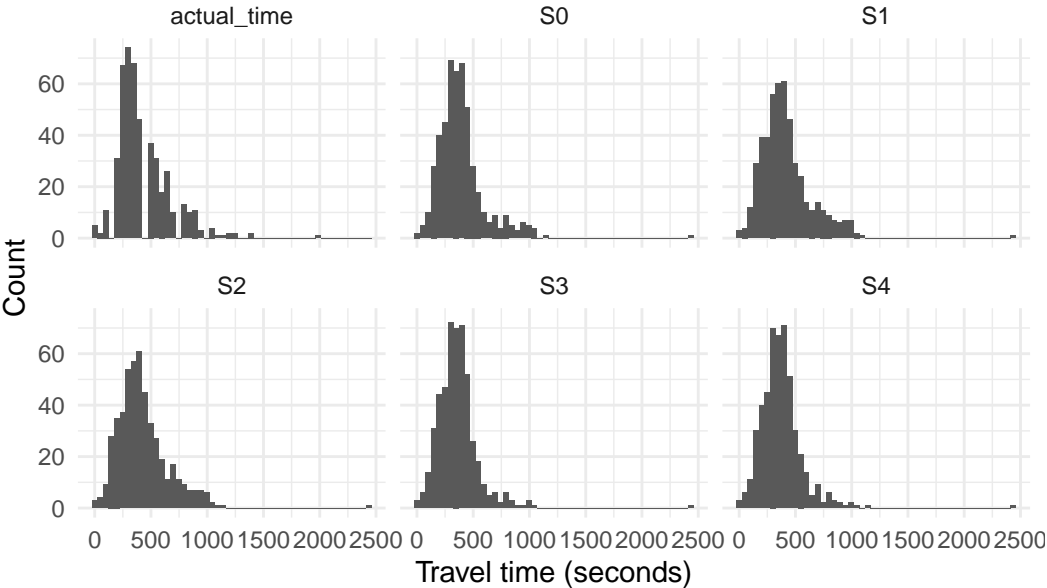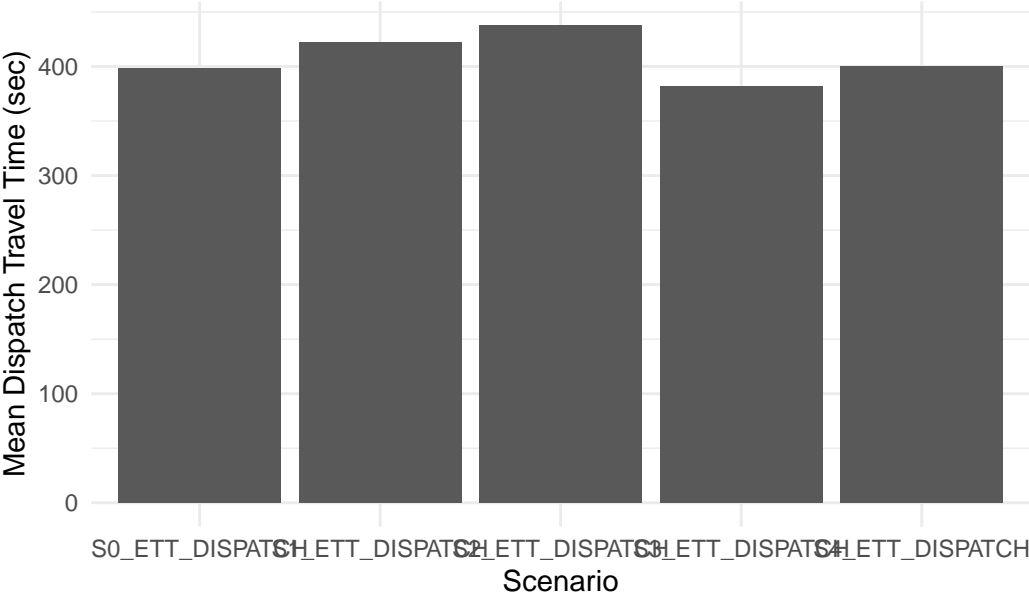
Actual vs. Estimated Travel Times
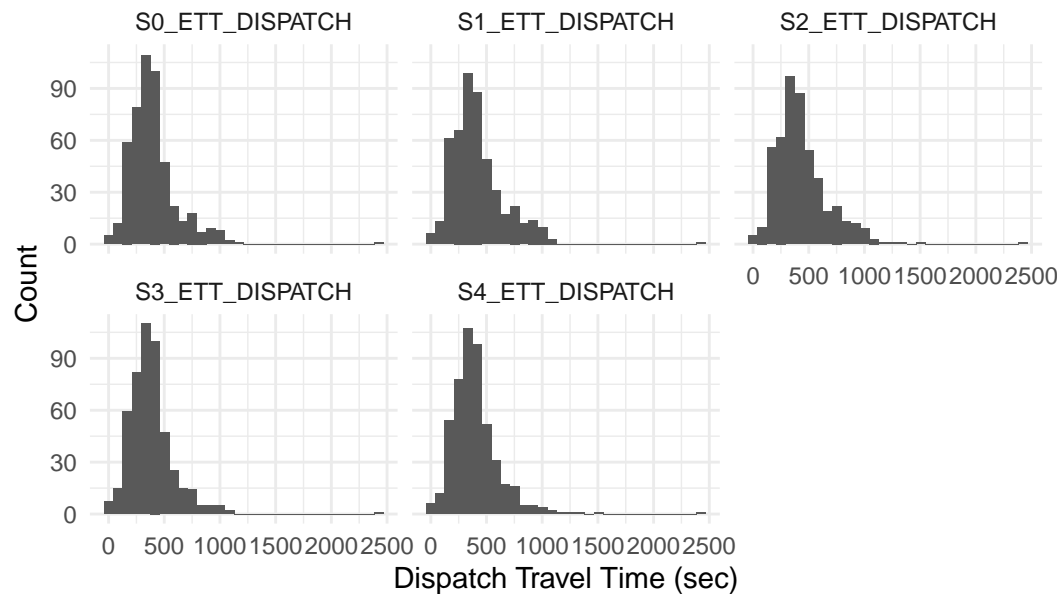

Error Distribution by Estimator

Shortest Travel Time Distributions by Scenario

Average Dispatch ETT per Scenario

Histogram of Dispatch ETT by Scenario


```
# A tibble: 20 x 3
   Scenario         DelayGroup      N
   <chr>            <fct>       <int>
 1 S0_ETT_DISPATCH  0-6 min      244
 2 S0_ETT_DISPATCH  7-9 min      166
 3 S0_ETT_DISPATCH  10-15 min     63
 4 S0_ETT_DISPATCH  >15 min       19
 5 S1_ETT_DISPATCH  0-6 min      228
 6 S1_ETT_DISPATCH  7-9 min      154
 7 S1_ETT_DISPATCH  10-15 min     84
 8 S1_ETT_DISPATCH  >15 min       26
 9 S2_ETT_DISPATCH  0-6 min      212
10 S2_ETT_DISPATCH  7-9 min      158
11 S2_ETT_DISPATCH  10-15 min     95
12 S2_ETT_DISPATCH  >15 min       27
13 S3_ETT_DISPATCH  0-6 min      253
14 S3_ETT_DISPATCH  7-9 min      166
15 S3_ETT_DISPATCH  10-15 min     61
16 S3_ETT_DISPATCH  >15 min       12
17 S4_ETT_DISPATCH  0-6 min      236
18 S4_ETT_DISPATCH  7-9 min      169
19 S4_ETT_DISPATCH  10-15 min     74
20 S4_ETT_DISPATCH  >15 min       13
```

Counts of Dispatch Delays by Scenario
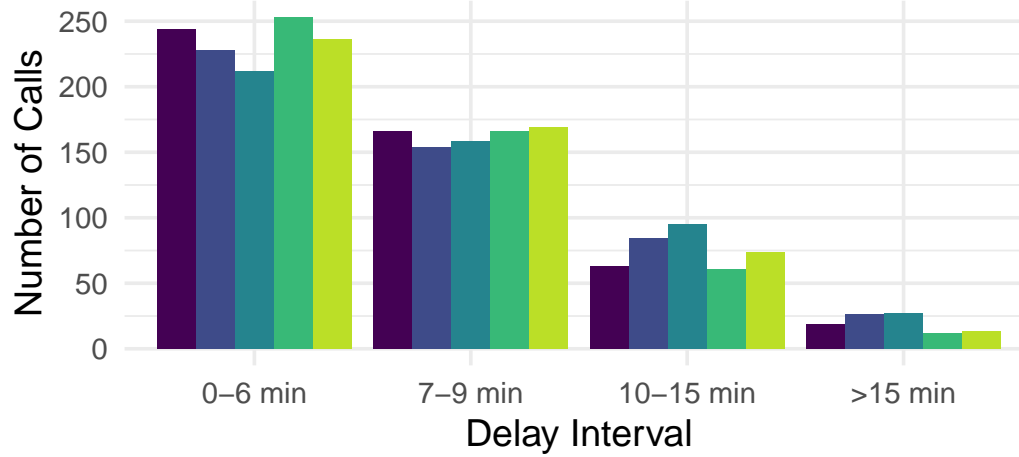
Dispatch Delay Counts by Scenario

Emergency/Non–Emergency EMS Calls by Region
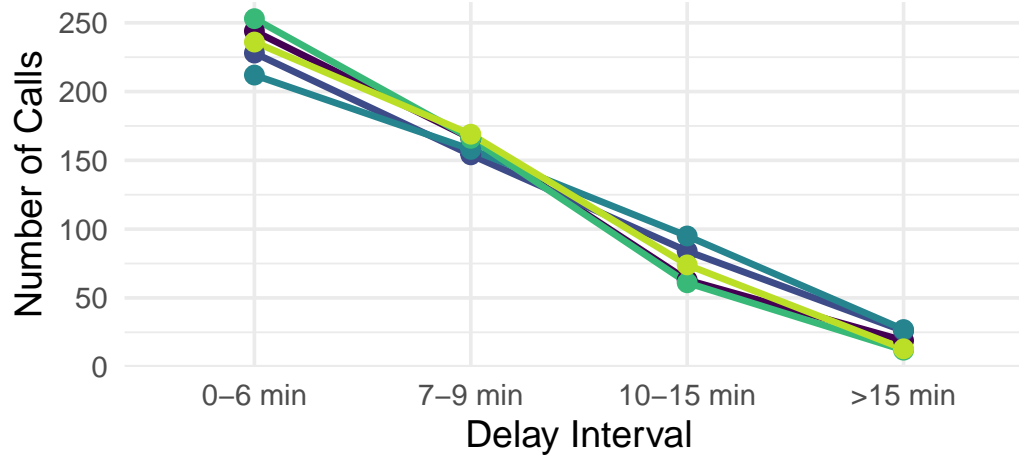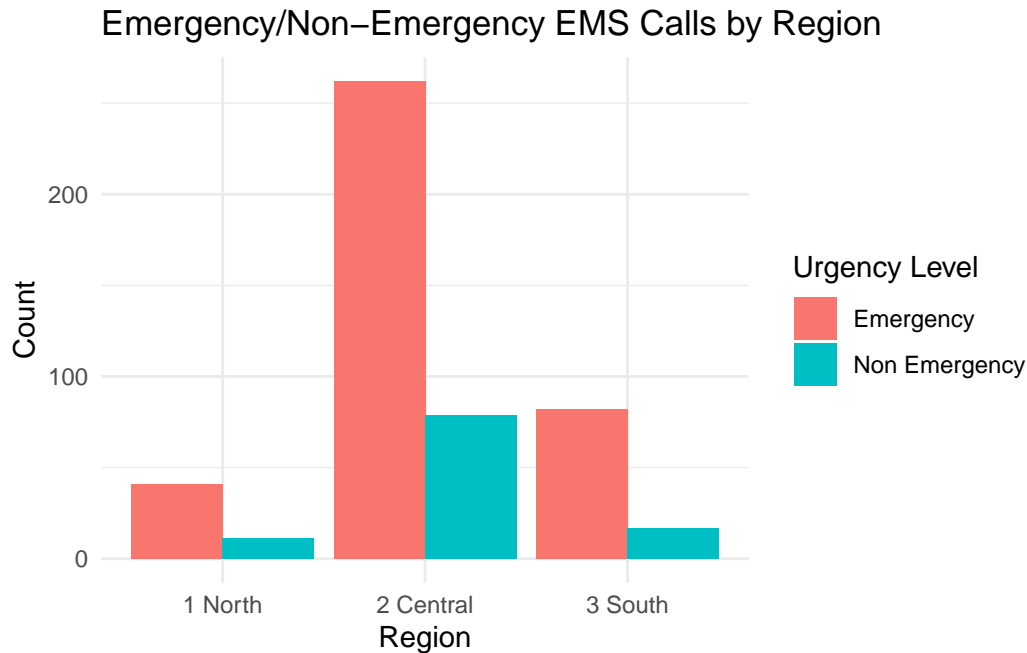

```
Linear mixed model fit by REML ['lmerMod']
Formula: obsTTmodified ~ (1 | REF.GRID) + DISPATCH.PRIORITY.NAME
   Data: x

REML criterion at convergence: 2560.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9133 -0.5399 -0.2182  0.3473  7.6984

Random effects:
 Groups   Name        Variance Std.Dev.
 REF.GRID (Intercept)  5.617   2.370
 Residual             12.509   3.537
Number of obs: 476, groups:  REF.GRID, 3

Fixed effects:
                                 Estimate Std. Error t value
(Intercept)                         8.377      1.388   6.037
DISPATCH.PRIORITY.NAMENon Emergency    1.138      0.393   2.896

Correlation of Fixed Effects:
           (Intr)
DISPATCH.PE -0.059

Generalized least squares fit by REML
  Model: obsTTmodified ~ DISPATCH.PRIORITY.NAME
  Data: x
       AIC       BIC     logLik
```

```
   2568.907 2585.552 -1280.454

Correlation Structure: Compound symmetry
 Formula: ~1 | region_category
 Parameter estimate(s):
      Rho
0.3098952

Coefficients:
                                      Value Std.Error  t-value p-value
(Intercept)                        8.377225  1.387607 6.037175   0.000
DISPATCH.PRIORITY.NAMENon Emergency 1.137910  0.392978 2.895608   0.004

 Correlation:
                                     (Intr)
DISPATCH.PRIORITY.NAMENon Emergency -0.059

Standardized residuals:
       Min          Q1         Med          Q3         Max
-2.2348794 -1.0281062 -0.5583537   0.1138833   5.7833057

Residual standard error: 4.257561
Degrees of freedom: 476 total; 474 residual


Call:
glm(formula = change_flag ~ region_category + DISPATCH.PRIORITY.NAME,
    family = binomial, data = model_df)

Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -1.84130    0.07694 -23.933  < 2e-16 ***
region_categoryNorth                 1.86199    0.14458  12.878  < 2e-16 ***
region_categorySouth                 3.06905    0.12819  23.942  < 2e-16 ***
DISPATCH.PRIORITY.NAMENon Emergency -0.46880    0.13877  -3.378 0.000729 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2965.4  on 2459  degrees of freedom
Residual deviance: 2188.6  on 2456  degrees of freedom
AIC: 2196.6

Number of Fisher Scoring iterations: 4

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: log(ETT_sec) ~ Scenario + region_category + DISPATCH.PRIORITY.NAME +
```

```
    (1 | row_val)
  Data: model_df

REML criterion at convergence: 2109.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.4455 -0.1945  0.0088  0.1643  6.1492

Random effects:
 Groups   Name        Variance Std.Dev.
 row_val  (Intercept) 0.23670  0.4865
 Residual             0.07798  0.2793
Number of obs: 2460, groups:  row_val, 492

Fixed effects:
                                  Estimate Std. Error         df t value
(Intercept)                        5.70728    0.03208  631.72338 177.925
ScenarioS1                         0.04018    0.01780 1964.00002   2.257
ScenarioS2                         0.08253    0.01780 1964.00002   4.636
ScenarioS3                        -0.04064    0.01780 1964.00002  -2.283
ScenarioS4                         0.00551    0.01780 1964.00002   0.309
region_categoryNorth               0.37144    0.07479  487.99996   4.967
region_categorySouth               0.44788    0.05744  487.99996   7.798
DISPATCH.PRIORITY.NAMENon Emergency 0.04280   0.05498  487.99996   0.778
                                  Pr(>|t|)
(Intercept)                        < 2e-16 ***
ScenarioS1                          0.0241 *
ScenarioS2                        3.79e-06 ***
ScenarioS3                          0.0226 *
ScenarioS4                          0.7570
region_categoryNorth             9.43e-07 ***
region_categorySouth             3.85e-14 ***
DISPATCH.PRIORITY.NAMENon Emergency 0.4367
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) ScnrS1 ScnrS2 ScnrS3 ScnrS4 rgn_cN rgn_cS
ScenarioS1 -0.278
ScenarioS2 -0.278  0.500
ScenarioS3 -0.278  0.500  0.500
ScenarioS4 -0.278  0.500  0.500  0.500
rgn_ctgryNr -0.314  0.000  0.000  0.000  0.000
rgn_ctgrySt -0.424  0.000  0.000  0.000  0.000  0.173
DISPATCH.PE -0.397  0.000  0.000  0.000  0.000  0.015  0.057
```
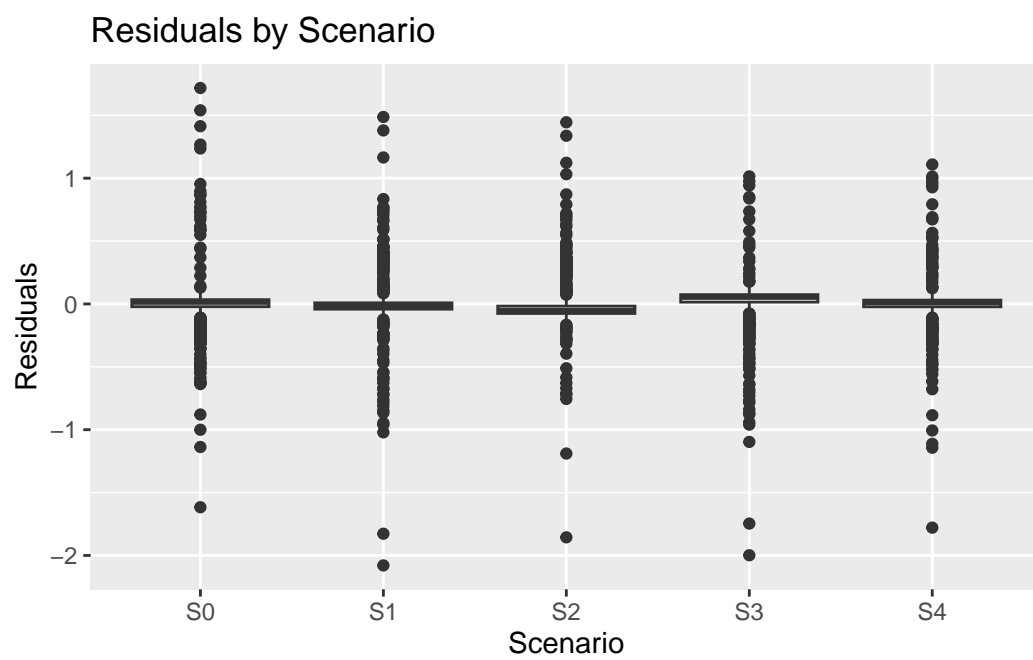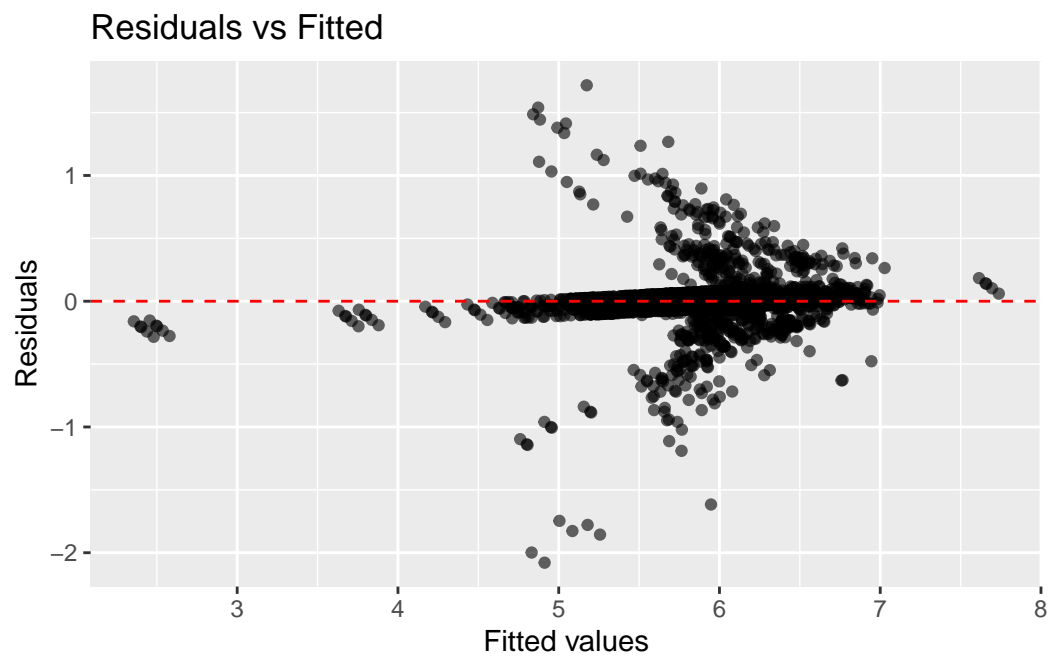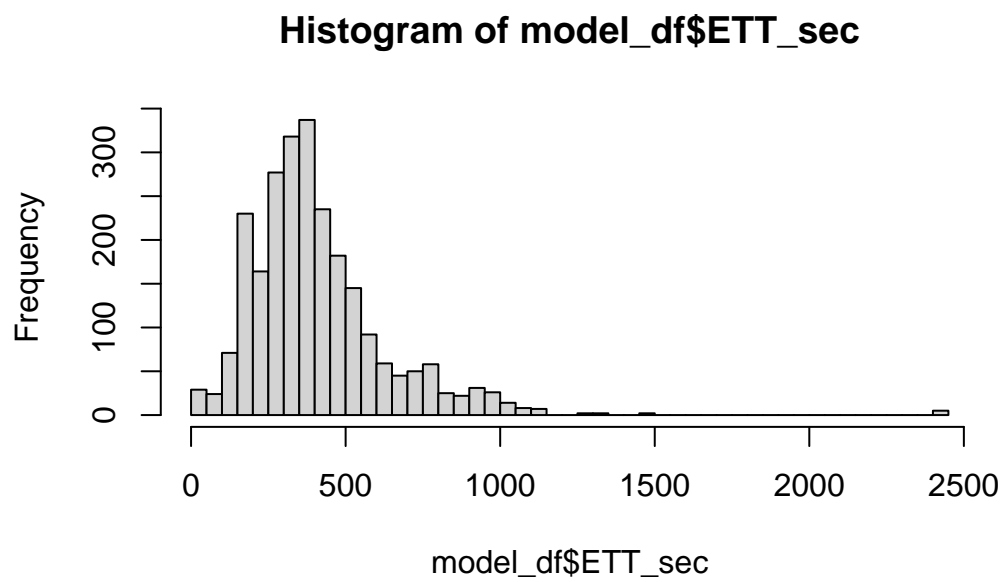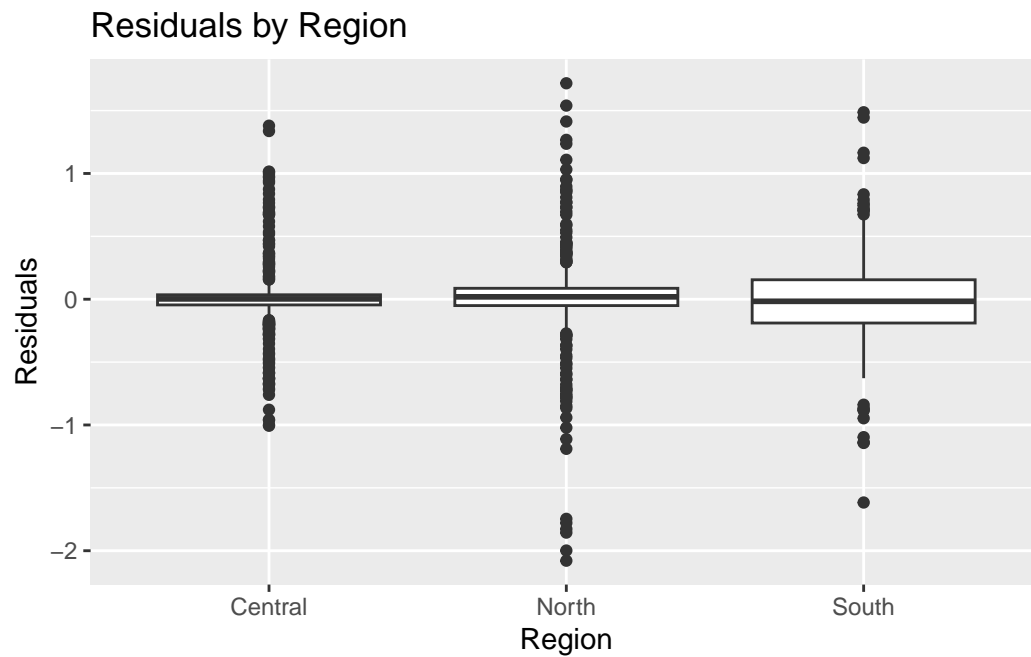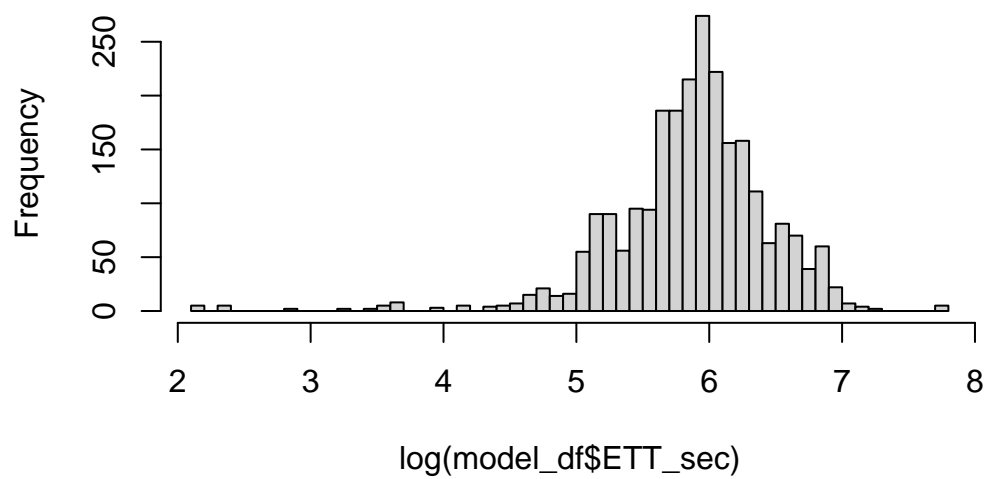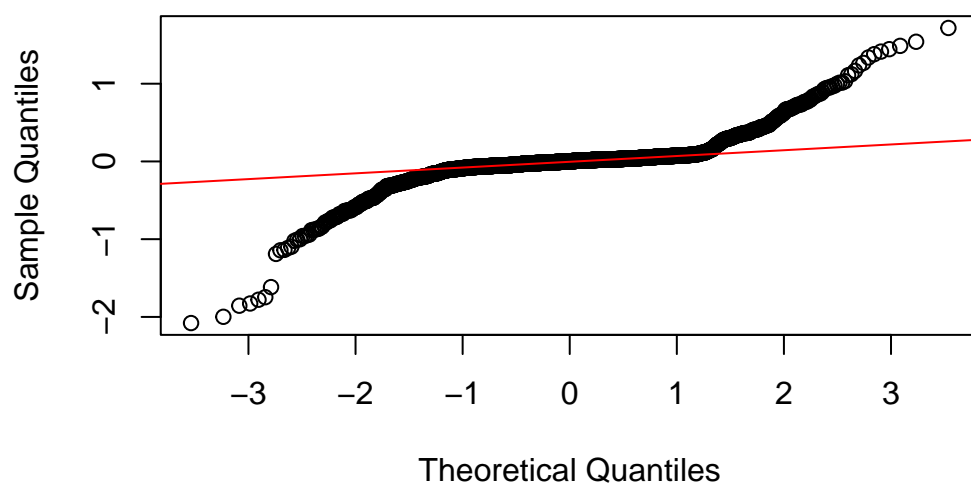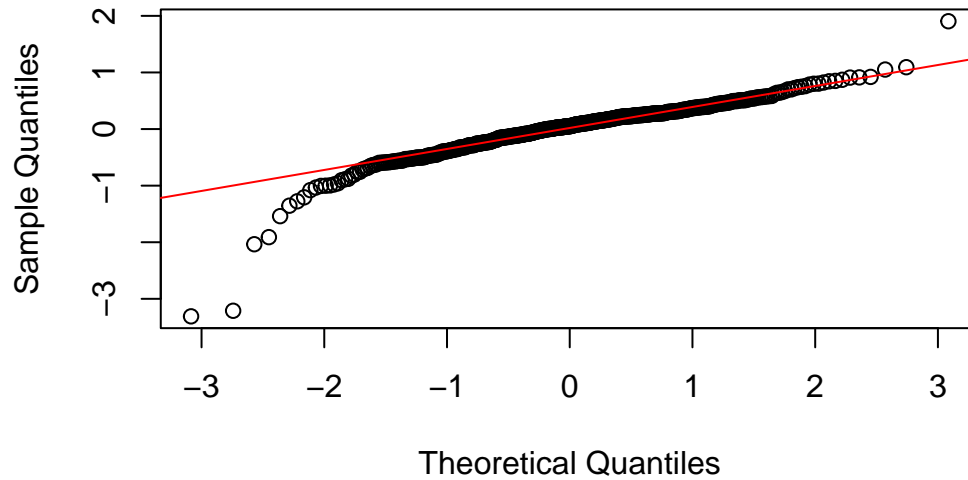
Residuals vs Fitted

Residuals by Scenario

Residuals by Region

Histogram of model_df$ETT_sec

Histogram of log(model_df$ETT_sec)

Residuals

Random Effects


```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: ETT_sec ~ Scenario + region_category + DISPATCH.PRIORITY.NAME +
    (1 | row_val)
   Data: changed_df

REML criterion at convergence: 9613

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4866 -0.5641 -0.0985  0.5039  3.4132

Random effects:
 Groups   Name        Variance Std.Dev.
 row_val  (Intercept) 17449    132.1
 Residual             34793    186.5
Number of obs: 715, groups:  row_val, 143

Fixed effects:
                                 Estimate Std. Error      df t value
(Intercept)                       426.613     28.211 237.481  15.122
ScenarioS1                         79.252     22.059 568.000   3.593
ScenarioS2                        134.245     22.059 568.000   6.086
ScenarioS3                        -56.573     22.059 568.000  -2.565
ScenarioS4                          6.196     22.059 568.000   0.281
region_categoryNorth               -5.336     39.322 139.000  -0.136
region_categorySouth               72.876     29.886 139.000   2.438
DISPATCH.PRIORITY.NAMENon Emergency 46.085    35.611 139.000   1.294
                                 Pr(>|t|)
(Intercept)                      < 2e-16 ***
ScenarioS1                       0.000356 ***
```

```
ScenarioS2                            2.14e-09 ***
ScenarioS3                            0.010585 *
ScenarioS4                            0.778913
region_categoryNorth                  0.892265
region_categorySouth                  0.016012 *
DISPATCH.PRIORITY.NAMENon Emergency   0.197760
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Correlation of Fixed Effects:
           (Intr) ScnrS1 ScnrS2 ScnrS3 ScnrS4 rgn_cN rgn_cS
ScenarioS1  -0.391
ScenarioS2  -0.391  0.500
ScenarioS3  -0.391  0.500  0.500
ScenarioS4  -0.391  0.500  0.500  0.500
rgn_ctgryNr -0.520  0.000  0.000  0.000  0.000
rgn_ctgrySt -0.671  0.000  0.000  0.000  0.000  0.483
DISPATCH.PE -0.205  0.000  0.000  0.000  0.000  0.039 -0.013
```
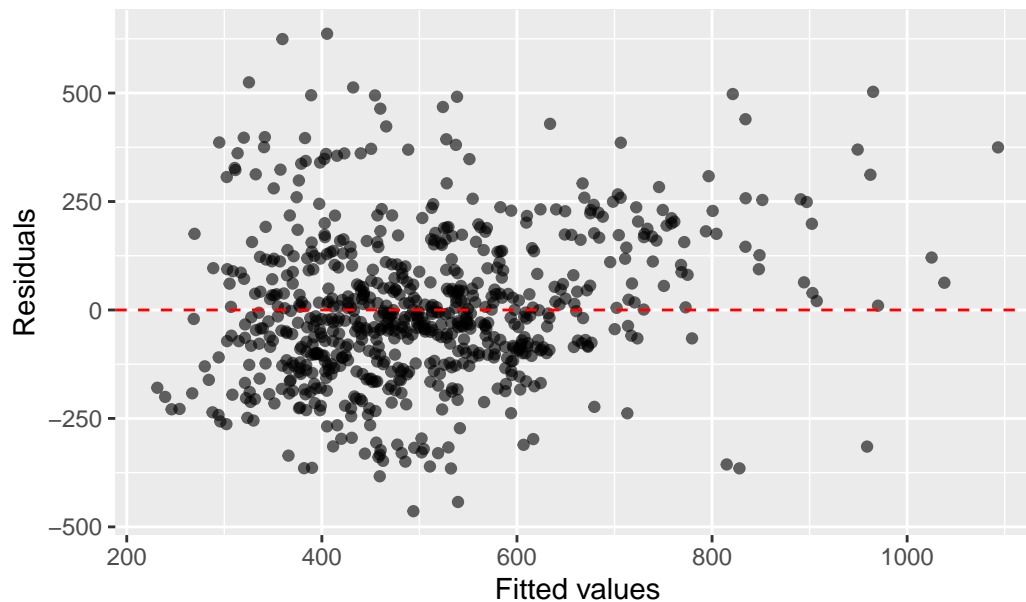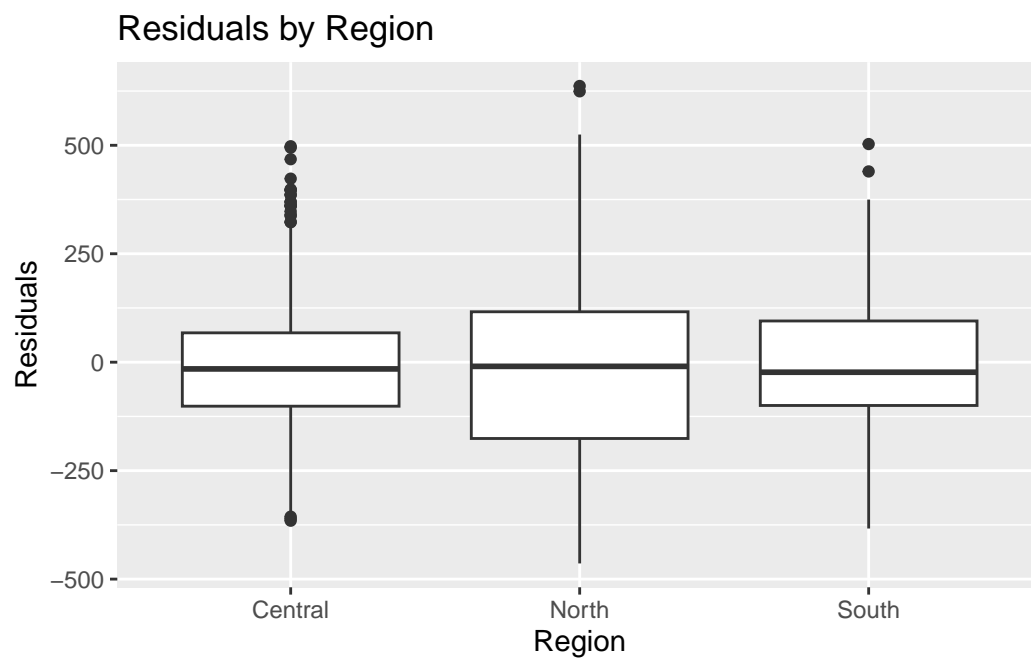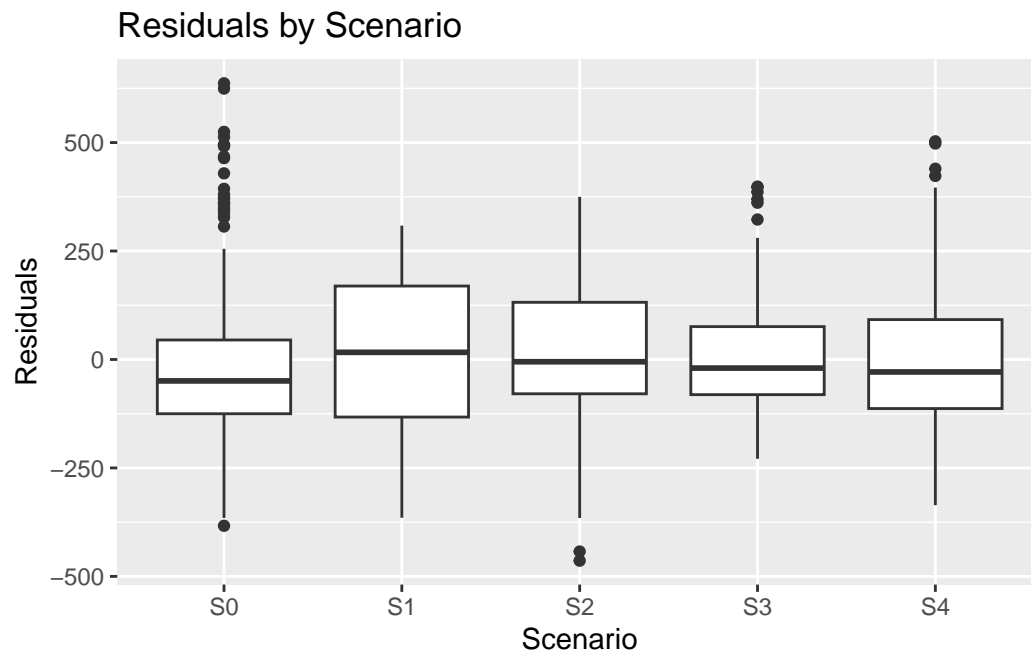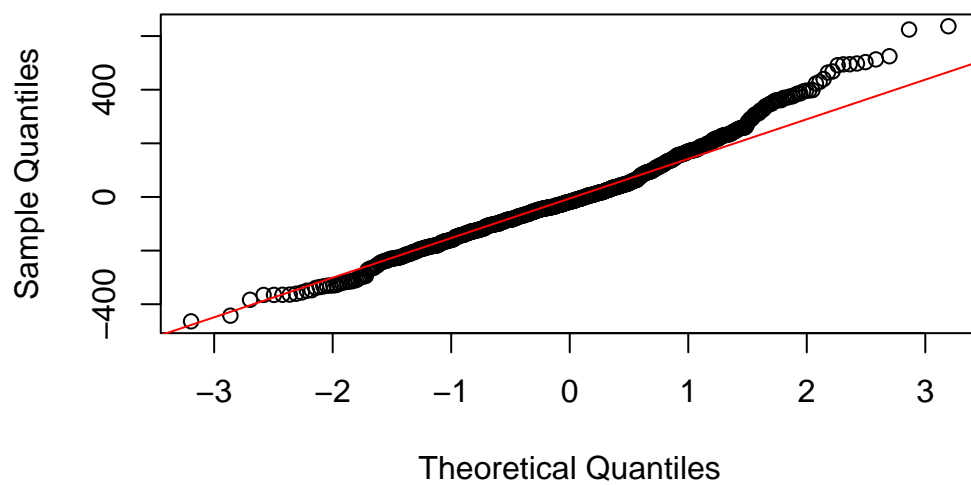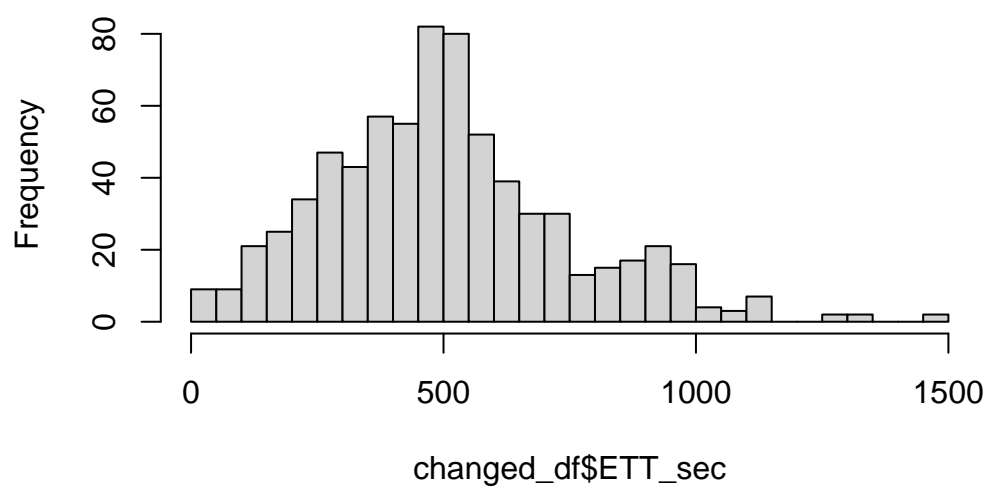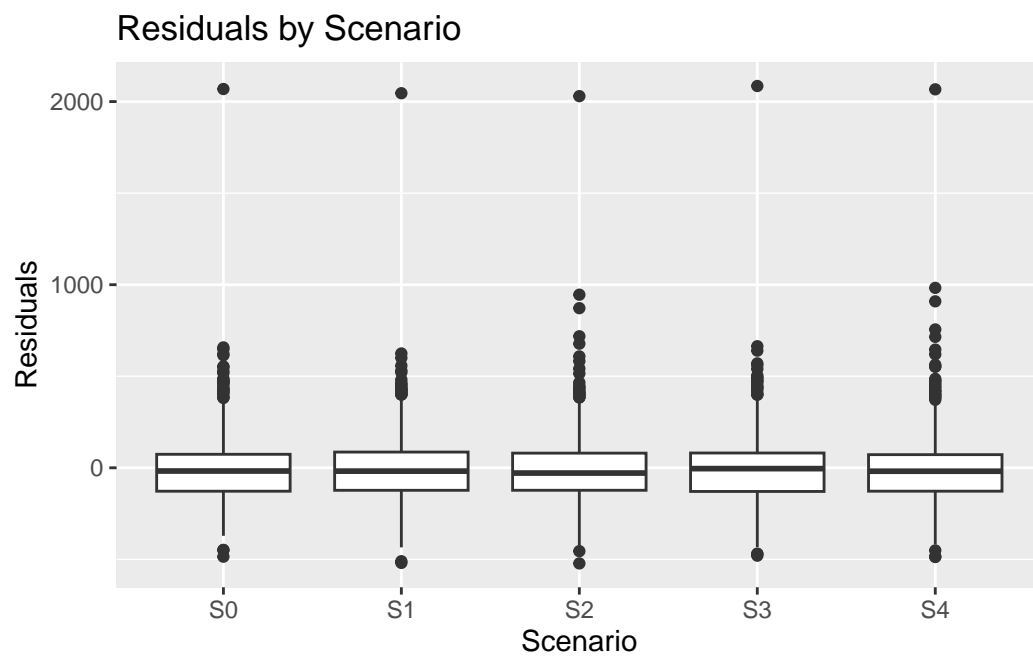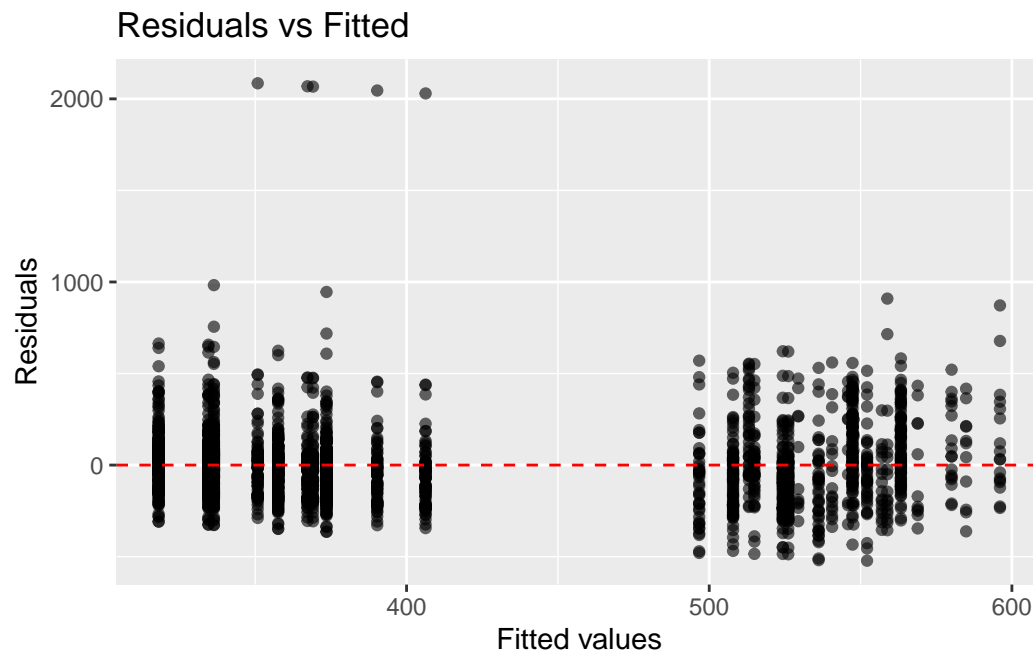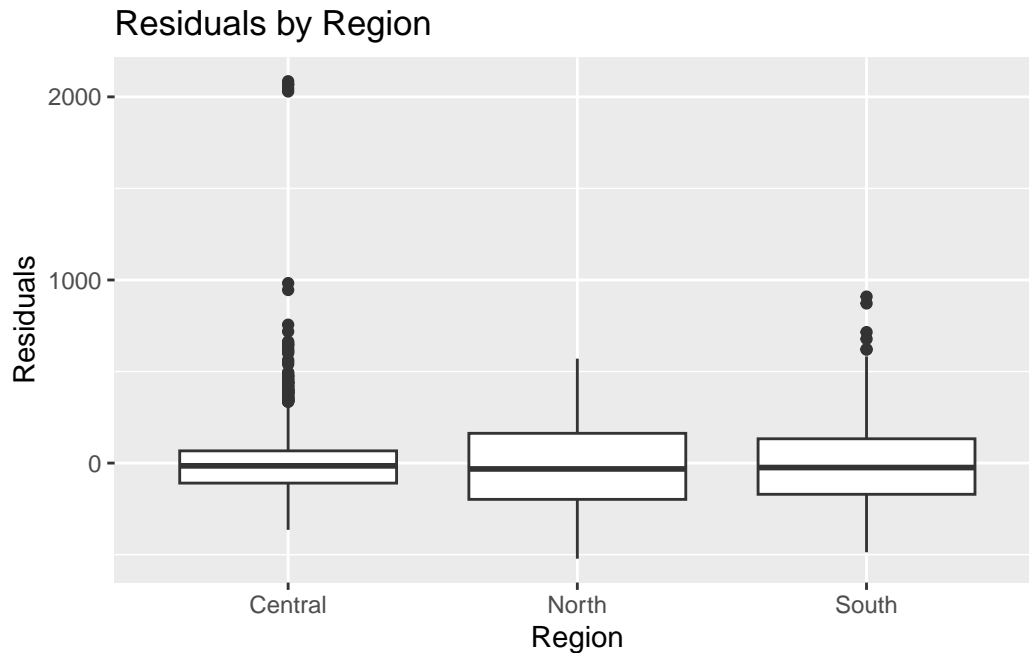
Residuals vs Fitted

Residuals by Scenario

Residuals by Region

Normal Q−Q Plot

Histogram of changed_df$ETT_sec

Residuals vs Fitted

Residuals by Scenario

Residuals by Region
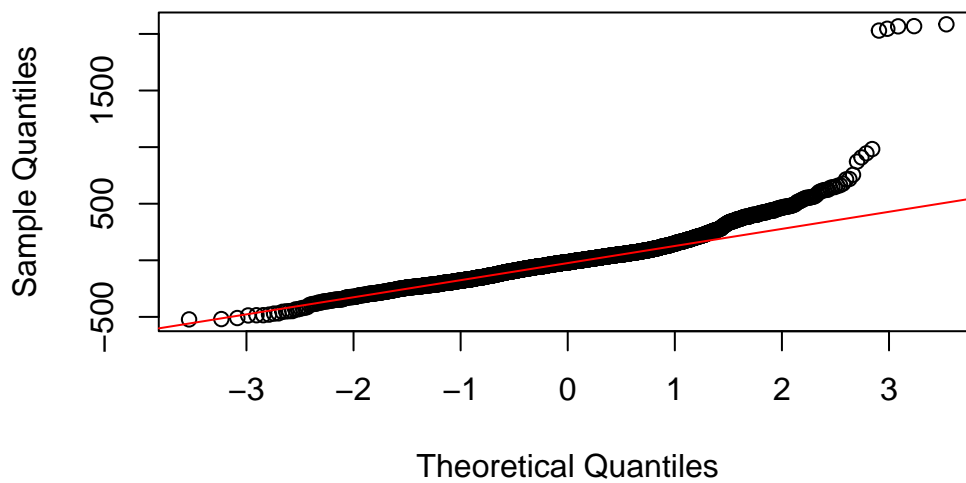

Normal Q–Q Plot


```
                Model df      AIC      BIC    logLik   Test  L.Ratio p-value
gls_model           1  9 33214.66 33266.90 -16598.33
gls_model_region    2 19 32977.35 33087.57 -16469.67 1 vs 2 257.3084  <.0001

               Model df      AIC      BIC    logLik   Test  L.Ratio p-value
gls_model           1  9 33214.66 33266.90 -16598.33
gls_model_pow       2 18 32954.17 33058.59 -16459.08 1 vs 2 278.4893  <.0001

Type III Analysis of Variance Table with Satterthwaite's method
                     Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
Scenario             4.2186 1.05465     4  1964 13.5242 7.053e-11 ***
region_category      5.7932 2.89661     2   488 37.1447 9.644e-16 ***
DISPATCH.PRIORITY.NAME 0.0472 0.04725   1   488  0.6059    0.4367
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
            df      AIC
lmm_model 10  2129.54
gls_model  9 33214.66

            df      BIC
lmm_model 10  2187.619
gls_model  9 33266.898

Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
Scenario              4.2186 1.05465     4  1964 13.5242 7.053e-11 ***
region_category       5.7932 2.89661     2   488 37.1447 9.644e-16 ***
DISPATCH.PRIORITY.NAME 0.0472 0.04725    1   488  0.6059    0.4367
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
              df      AIC
lmm_model     10  2129.54
gls_model_pow 18 32954.17

              df      BIC
lmm_model     10  2187.619
gls_model_pow 18 33058.592

Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
Scenario              3177443  794361     4   568 22.8310 <2e-16 ***
region_category        284976  142488     2   139  4.0953 0.0187 *
DISPATCH.PRIORITY.NAME  58272   58272     1   139  1.6748 0.1978
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                  df      AIC
lmm_model_changed 10  9632.967
gls_model          9 33214.656

                  df      BIC
lmm_model_changed 10  9678.69
gls_model          9 33266.90

Type III Analysis of Variance Table with Satterthwaite's method
                      Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
Scenario              3177443  794361     4   568 22.8310 <2e-16 ***
region_category        284976  142488     2   139  4.0953 0.0187 *
DISPATCH.PRIORITY.NAME  58272   58272     1   139  1.6748 0.1978
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                  df      AIC
lmm_model_changed 10  9632.967
gls_model_pow     18 32954.167
```
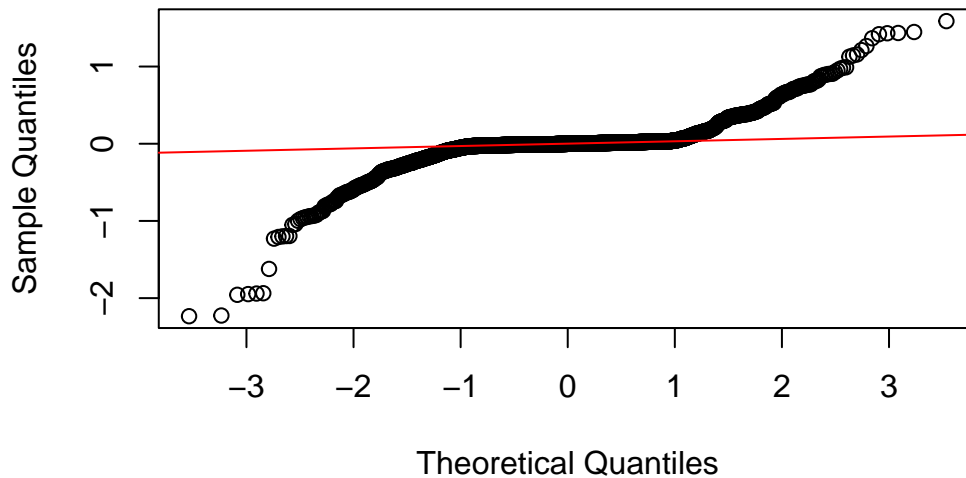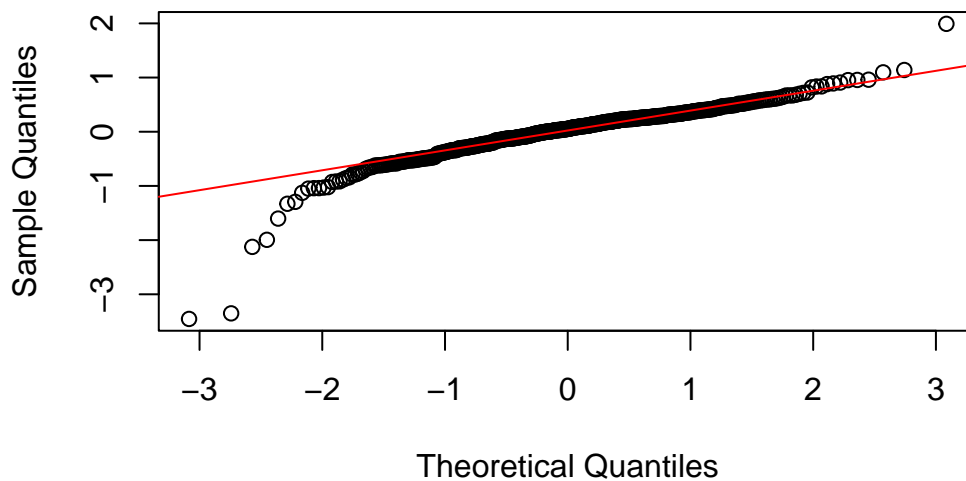
```
                     df       BIC
lmm_model_changed 10   9678.69
gls_model_pow     18 33058.59
```

Residuals

Random Effects


```
                Model df      AIC      BIC    logLik   Test  L.Ratio p-value
lmm_no_weights      1 10 2129.540 2187.586 -1054.770
lmm_model_ext       2 12 1170.544 1240.200  -573.272 1 vs 2 962.9957  <.0001
```

4. Assumptions and Shortcomings

Normality of residuals was poorly met Large deviances at tails Non-constant variance across fitted values External factors not taken into consideration Traffic / Rush hours/ Road closures Time of year EMT staffing patterns ETT inputs: based on single best-guess estimates (e.g., from Google), lacking nuance or uncertainty Variance modeling: only accounted for regional differences, while other factors may also influence variability Emergency vs non-emergency Simplicity in Dispatch

Rule Better for analysis Doesn't occur in real life

5. Conclusions

Scenario 3 results in the fastest overall response times Marginal mean travel times (ETTs) are reduced by ~16% compared with the baseline (Scenario 0)

$\exp(-0.173)$ $0.84 \rightarrow$ travel times in Scenario 3 are about 84% of baseline levels