With characteristics of a zip code to predict the enrollment of Medicare

1. Boxplot for average enrollment across zip code
2. F test across zip code
3. Income, tax, race, religion……with the control of zip code, to see if they are still significant
4. Predict with 10 years data

Inference

**1. Model Description:**

* **Formula**: State the formula used in the model, including both fixed and random effects.  
  Example:

The model used to predict enroll\_rate included fixed effects for AGI, marriage rate, elder rate, tax, charity, additional medical expenses (add\_med), education, retirement status (retire), and number of children (child). A random intercept was included for the STATE variable.

**2. Fit Information:**

* **REML criterion**: The log-likelihood statistic is an important measure for model fit (especially with REML estimation).  
  Example:

The model was fit using REML (Restricted Maximum Likelihood) with a criterion of -749.2.

**3. Residuals:**

* Report the distribution of residuals to assess model fit.  
  Example:

The residuals showed a reasonable distribution with a minimum of -4.6983, a first quartile (Q1) of -0.4187, a median of 0.0275, a third quartile (Q3) of 0.4053, and a maximum of 6.1039.

**4. Random Effects:**

* Provide the variance and standard deviation of random effects.  
  Example:

The variance of the random intercept for STATE was 0.3383, with a standard deviation of 0.5816. The residual variance was 0.00546, with a standard deviation of 0.07391.

**5. Fixed Effects:**

* List the fixed effects, their estimates, standard errors, and t-values.  
  Example:

The fixed effects estimates and their associated standard errors and t-values were as follows:

* + **Intercept**: Estimate = 2.919, Std. Error = 0.195, t = 14.96
  + **AGI**: Estimate = 0.155, Std. Error = 0.585, t = 0.27
  + **Marriage Rate**: Estimate = -7.036, Std. Error = 0.406, t = -17.33 (statistically significant)
  + **Elder Rate**: Estimate = 0.344, Std. Error = 0.236, t = 1.46
  + **Tax**: Estimate = 0.009, Std. Error = 0.049, t = 0.19
  + **Charity**: Estimate = 0.153, Std. Error = 0.192, t = 0.80
  + **Additional Medical Expenses**: Estimate = -0.314, Std. Error = 0.206, t = -1.52
  + **Education**: Estimate = -1.172, Std. Error = 0.429, t = -2.73 (statistically significant)
  + **Retirement Status**: Estimate = 0.912, Std. Error = 0.286, t = 3.20 (statistically significant)
  + **Children**: Estimate = -0.302, Std. Error = 0.131, t = -2.30 (statistically significant)

**6. Correlation of Fixed Effects:**

* Report correlations between fixed effects, which can help interpret potential multicollinearity in the model.  
  Example:

There was a high negative correlation between **marriage\_rate** and **education** (-0.813), and moderate positive correlations between **tax** and **charity** (0.440), and between **education** and **retirement status** (0.521).

**7. Statistical Significance:**

* Report which predictors are statistically significant. This can be based on the t-value, where predictors with t-values larger than approximately 2 (in absolute value) are typically considered significant at the 5% level.  
  Example:

Significant predictors (t-value > 2) include **marriage\_rate** (t = -17.33), **education** (t = -2.73), **retirement status** (t = 3.20), and **children** (t = -2.30).

**8. Conclusion:**

* Summarize the findings, emphasizing the impact of significant variables and explaining how they relate to the dependent variable (enroll\_rate).  
  Example:

The **marriage rate** was the strongest predictor of enroll\_rate, with a significant negative relationship. Additionally, **education**, **retirement status**, and **children** were significant predictors, with **education** and **children** negatively associated with enroll\_rate.