STAT-437: APPLIED STATISTICAL METHODS
Assignment: 3
Due: October 17, 2025 at 10:00PM
Total=30 Marks

*

Instructions: All answers and relevant R outputs must be given in the main part. Include R code in the Appendix and submit the source code in the script(.r) or markdown(.rmd). Also, pdf version of Assignment is required to submit. Late submissions will not be accepted.

Ohio Children Wheeze Status:

"ohio" data set is available in the r-package "faraway". The description, Usage, Format and References of ohio dataset is as below:

- Description: The ohio data frame has 2148 rows and 4 columns. The dataset is a subset of the six-city study, a longitudinal study of the health effects of air pollution.

- Usage: data(ohio)

- Format:
  resp: an indicator of wheeze status (1=yes, 0=no).
  id: a numeric vector for subject id.
  age: a numeric vector of age, 0 is 9 years old.
  smoke: an indicator of maternal smoking at the first year of the study.

- References:
  (i)Fitzmaurice, G.M. and Laird, N.M. (1993) A likelihood-based method for analyzing longitudinal binary responses, Biometrika 80: 141-151.

Give the answers of the following:

1. Fit a Generalized Estimating Equations (GEE) model using the geeglm() function with a binary response variable **resp** as the dependent variable and **age** and **smoke** as covariates. Use four different correlation structures: "independence", "exchangeable", "AR(1)", and "unstructured". Based on the fitted models, interpret the coefficient of age under the exchangeable correlation structure and compare how the coefficients and significance levels for age and smoke differ across the four correlation structures.

2. Fit a Generalized Linear Mixed Effects Model (GLMM) with a random effect on the intercept, using a binary response variable **resp** as the dependent variable and **age** and **smoke** as covariates. Explain what the standard deviation of the random intercept represents in this model.

3. Fit a Generalized Linear Mixed Effects Model (GLMM) with random effects on both the intercept and **smoke**, using a binary response variable **resp** as the dependent variable and **age** and **smoke** as covariates. Based on the fitted model, interpret the results to explain how age and smoke influence the response variable.

*Instructor: Sharandeep Singh Pandher, Address: Department of Mathematics and Statistics, University of Alberta, Edmonton, AB,T6G 2G1, Canada

4. Fit a Generalized Linear Model (GLM) with a binary response variable **resp** as the dependent variable and **age** as the covariate. Then, fit a Generalized Additive Model (GAM) using a spline smoother for age, and compare the two models using an appropriate ANOVA test.