Graphical user interface, diagram

Description automatically generated

Question 1.0 Please write down the Assumptions for simple linear regression from BIOS6611 (5’ or 0’)

Linearization

Independence

Identical distributed

Normality

Variance homoscedasticity

Question 1.1 Fill out the blanks 1-7 with the model we have learned in BIOS6643 (7’):

LM: simple linear model

GLS: general least square

GAM: generalized additive model

GEE: generalized estimate equation

GzLM: Generalized linear model

GzLMM: Generalized linear mixed model

LMM: Linear mixed model

Graphical user interface, application

Description automatically generated

Question1.2 From blank 1 to black 2 in the graph, which assumption is violated (2’)?

Outcome Normality and/or linearity

What are the link between 1 and 2 (three element from Gary’s lecture) (3’)?

exponential family

regression on mean function

link function

Graphical user interface, text, application

Description automatically generated

For 2, there is a practical problem for the variance, what is it (1’)? What can you do to solve this problem (3’)?

Under- and Over-dispersion

1. quasilikelihood

2. add a random error for the transformed mean function

3. use Gamma-Poisson (Negative-Binomial) for Poisson regression;

use Beta-Binomial for Logistic regression

Question1.3 (7 points) Write down the functions or algorithms you use (in either SAS or R) For blank 1-7 (7’)

as long as providing reasonable R or SAS functions used

Diagram

Description automatically generated

Question1.4 (4 points) From 1 to 4 which assumptions violated question1.0; From 1 to 5 which assumptions violated from question1.0 (2’)

Residual error term independence (correlated residuals) + equal variance

identical distribution + equal variance.

If I care about some individual’s trajectory (conditional model), which method should I choose? If I care population level (marginal model), which method should I choose? (2’)

conditional model with LMM with random terms

marginal model with GLS

Question1.5 (5 points) Why 7 is harder than 5 to apply in softwares? What are the major drawbacks in both 6 and 7? (5)

no closure form for marginal model with integration, has to use numerical methods.

Homework question, using no closure form of marginal model, and using quasilikelihood

Question1.6 (10 points) Consider 8, with both random terms and repeated measures with an AR1 covariance structure. Please write down the mean model at individual level (4’), including the distribution of random terms and error terms. Please write down the matrice of G, R, and V (2’ for each).

with Y, X, \beta, b, Z, sigma\_e, phi and other terms.

Homework question.

Question1.7 (9 points) Please list three types of Missing Pattern and summarize each one with one or two sentences or with simple graphes (6’). What are the missing assumptions for blank 2, 6, and 7 (3’).

MCAR: missing completely at random

MAR: missing at random

MNAR: missing not at random

Diagram

Description automatically generated

For likelihood based functions, like GzLM or GzLMM: MCAR or MAR are okay

GEE must under MCAR