STA303/1002: Mid-semester Summary of Models

1 Components of a Generalized Linear Model

- Response variable: Y
- Explanatory variables: $X_1, \ldots, X_p$
- Link function: $g(\cdot)$
- Model: $g(E(Y)) = f(\boldsymbol{X}; \boldsymbol{\beta})$ where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and $f(\boldsymbol{X}; \boldsymbol{\beta})$ is a linear function of the β's

2 One-way and Two-way Analysis of Variance

1. Underlying probability distribution: Normal
2. Response variable: continuous
3. Explanatory variables: categorical
4. Model: $Y = f(\boldsymbol{X}; \boldsymbol{\beta}) + \epsilon$ or $E(Y) = f(\boldsymbol{X}; \boldsymbol{\beta})$ where the explanatory variables are indicator variables with coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and $f(\boldsymbol{X}; \boldsymbol{\beta})$ is a linear function of the β's.
5. Link function: identity
6. Conditions for valid inference (assuming correct form of model[1]):
   - independent observations
   - same variance
   - normally distributed error terms (so no outliers)
7. Estimation: least squares
8. Inference: t and F tests based on the Normal distribution

3 Binary or Binomial Logistic Regression

1. Underlying probability distribution: Bernoulli or Binomial
2. Response variable: binary or binomial counts out of m trials
3. Explanatory variables: anything
4. Model: $\log(\frac{\pi}{1-\pi}) = f(\boldsymbol{X}; \boldsymbol{\beta})$ where $f(\boldsymbol{X}; \boldsymbol{\beta})$ is a linear function of the β's
5. Link function: logit
6. Conditions for valid inference (assuming correct form of model):
   - independent observations
   - variance follows Bernoulli or Binomial distribution form
   - no outliers
   - large sample size
7. Estimation: maximum likelihood estimation
8. Inference: Likelihood ratio tests, Wald tests and confidence intervals based on large-sample properties of maximum likelihood estimators

[1] Correct form of model includes: necessary explanatory variables are in the model, unnecessary explanatory variables are not in the model, continuous explanatory variables are transformed as appropriate