Chapter 1 : Introduction

• If the random variable Y has the Normal distribution with mean μ, and variance σ^2, its probability density function is

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma^2}\right)^2\right]$$

• The **central chi-squared distribution** with n degrees of freedom is defined as the sum of squares of n independent random variables $Z_1, \ldots, Z_n$ each with the standard Normal distribution. It is denoted by

$$X^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)\cdot$$

• Let $Z_1, \ldots, Z_n$ be independent random variables each with the distribution $N(0,1)$ and let $Y_i = Z_i + \mu_i$, where at least one of μ_i's is non-zero. Then the distribution of

$$\sum Y_i^2 = \sum (Z_i + \mu_i)^2 = \sum Z_i^2 + 2\sum Z_i\mu_i + \sum \mu_i^2$$

has larger mean $n + \lambda$ and larger variance $2n + 4\lambda$ than $\chi^2(n)$ where $\lambda = \sum \mu_i^2$. This is called **non-central chi-squared distribution** with n degrees of freedom and **non-centrality parameter** λ. It is denoted by $\chi^2(n,\lambda)$.

• Suppose that the Y_i's are not necessary independent and the vector $y = [Y_i, \ldots, Y_n]$ has the multivariate normal distribution $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu},\ \mathbf{V})$ where the variance-covariance matrix $\mathbf{V}$ is non-singular and its inverse is V^{-1}. Then

$$X^2 = (\boldsymbol{y}-\boldsymbol{\mu})^T\ \mathbf{V}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \sim \chi^2(n)\cdot$$

• If $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu},\ \mathbf{V})$ where the variance-covariance matrix $\mathbf{V}$ is non-singular and its inverse is V^{-1} then $\boldsymbol{y}^T\ \mathrm{V}^{-1}\ \boldsymbol{y}$ has the non-central chi-squared distribution $\chi^2(n,\lambda)$ where $\lambda = \boldsymbol{\mu}^T\ \mathrm{V}^{-1}\ \boldsymbol{\mu}$.

• t-distribution

$$T = \frac{Z}{(X^2/n)^{1/2}}$$

where $Z \sim N(0,1)$, $X^2 \sim \chi^2(n)$ and Z and X^2 are independent. This is denoted by $T \sim t(n)$

• The **central F-distribution** with n and m degrees of freedom is defined as the ratio of two independent central chi-squared random variables each divided by its degrees of freedom

$$F = \frac{X_1^2}{n} / \frac{X_2^2}{m}$$

where $X_1^2 \sim \chi^2(n), X_2^2 \sim \chi^2(m)$ and X_1^2 and X_2^2 are independent. This is denoted by $F \sim F(n,m)$

• The **non-central F-distribution** is defined as the ratio of two independent random variables, each divided by its degrees of freedom, where the numerator has a central chi-squared distribution and the denominator has a central chi-squared distributoin, i.e.,

$$F = \frac{X_1^2}{n} / \frac{X_2^2}{m}$$

where $X_1^2 \sim \chi^2(n,\lambda)$ with where $\lambda = \boldsymbol{\mu}^T\ \mathrm{V}^{-1}\ \boldsymbol{\mu}$, $X_2^2 \sim \chi^2(m)$ and X_1^2 and X_2^2 are independent.

• The quadratic form $\boldsymbol{y}^T\mathbf{A}\boldsymbol{y}$ and the matrix $\mathbf{A}$ are said to be **positive definite** if $\boldsymbol{y}^T\mathbf{A}\boldsymbol{y} > 0$ whenever the elements of u are not all zero. What this basically means is that none of the roots of the quadratic form are complex. Thus the rank of the matrix $\mathbf{A}$ is called the degrees of freedom of the quadratic form $\boldsymbol{y}^T\mathbf{A}\boldsymbol{y}$.

• Cochran's theorem:
Suppose $Y_1, \ldots, Y_n$ are independent random variables each with the normal distribution $N(0,\sigma^2)$. Let $Q = \sum_{i=1}^{n} Y_i^2$, and let $Q_1, \ldots, Q_n$ be quadratic forms in the Y_i's such that

$$Q = Q_1 + \ldots + Q_k$$

where Q_i has m_i degrees of freedom $(i = 1, \ldots, k)$. Then $Q_1, \ldots, Q_n$ are independent random variables and $Q_1/\sigma^2 \sim \chi^2(m1), \ldots, Q_k/\sigma^2 \sim \chi^2(m_k)$ if and only if

$$m_1 + \ldots + m_k = n$$

Chapter 3 : Exponential Family and Generalized Linear Models

• Exponential family

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right]$$

If $a(y) = y$, the distribution is said to be in **canonical form** and $b(\theta)$ is sometimes called the **natural parameter** of the distribution. If there are other parameters, in addition to the parameter of interest θ, they are regarded as **nuisance parameters**.

• Mean and Variance

$$E[a(Y)] = -c'(\theta)/b'(\theta)$$

$$Var[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

• Score Statistic : Mean and Variance

$$U(\theta;y) = \frac{dl(\theta;y)}{d\theta} = a(y)b'(\theta) + c'(\theta)$$

$$E[U] = 0, Var[U] = \mathfrak{I} = -E(U')$$

where $\mathfrak{I}$ is Fischer's information matrix

• GLM model has three components
  - Response variable $Y_1, \ldots, Y_n$ are assumed to share the same distribution from the exponential family, i.e., they have the canonical form and depend on a single parameter θ_i. So, the nuisance parameters are not needed for estimation purpose.
  - A set of parameters $\boldsymbol{\beta}$ and explanatory variables
  - A monotone linke function g such that $g(\mu_i) = \boldsymbol{x_i^T\beta}$ where $\mu_i = E(Y_i)$

Chapter 4 : Estimation

- Method of Scoring

$$\theta^m = \theta^{(m-1)} - \frac{U^{(m-1)}}{U'^{(m-1)}} = \theta^{(m-1)} + \frac{U^{(m-1)}}{\mathfrak{I}^{(m-1)}}$$

- Standard error of $\hat{\theta}$ is $\sqrt{1/\mathfrak{I}}$
- Method of Scoring for GLM

$$\boldsymbol{b^m} = \boldsymbol{b^{(m-1)}} + \left[\mathfrak{I}^{(m-1)}\right]^{-1} U^{(m-1)}$$

- $\mathfrak{I} = \mathrm{X}^{\mathrm{T}}\ \mathrm{W}\ \mathrm{X}$, where W is a diagonal matrix with

$$w_{ii} = \frac{1}{var(Y_i)} \left[\frac{\partial \mu_i}{\partial \eta_i}\right]^2$$

$$g(\mu_i) = \eta_i = x_i^T \boldsymbol{\beta}$$

- Iterative GLM Equation : $\mathrm{X^TWX\ b^{(m)} = X^TWz}$, where

$$z_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left[\frac{\partial \mu_i}{\partial \eta_i}\right]$$

For Generalized Linear Models, MLE estimators are obtained by an iterative weighted least squares procedure.

Chapter 5 : Inference

- If S is a statistic of interest, then

$$[s - E(s)]^T V^{-1} [s - E(s)] \sim \chi^2(p)$$

- Sampling Distribution of Score Statistic $U \sim N(0, \mathfrak{I})$

$$U^T \mathfrak{I}^{-1} U \sim \chi^2(p)$$

- Sampling Distribution of MLE $\boldsymbol{b} \sim N(\boldsymbol{\beta}, \mathfrak{I}^{-1})$.
- Wald Statisic

$$[\boldsymbol{b} - \boldsymbol{\beta}]^T \mathfrak{I}(\boldsymbol{b}) [\boldsymbol{b} - \boldsymbol{\beta}] \sim \chi^2(p)$$

- log likelihood (ratio) statistic

$$D = 2[l(\boldsymbol{b_{max}}; \boldsymbol{y}) - l(\boldsymbol{b}; \boldsymbol{y})]$$

- Deviance $\sim \chi^2(m - p, v)$ where v is the non-centraility parameter, m is number of parameters in the saturated model and p is the number of parameters in the model of interest
- Deviance of a binomial model

$$D = 2\sum_{i=1}^{N} [y_i log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i) log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)$$

- Deviance of a Normal model

$$D = \frac{1}{\sigma^2} \sum_{i=1}^{N} (y_i - \hat{\mu_i})^2$$

- Deviance of a Poisson model

$$D = 2\left[\sum y_i log\left(\frac{y_i}{\hat{y}_i}\right) - \sum (y_i - \hat{y}_i)\right] = 2\sum o_i\ log(o_i/e_i)$$

- We can test H_0 against H_1 using the difference of deviance statistics

$$\Delta D = D_0 - D_1 = 2[l(b_{max}; y) - l(b_0; y)] - 2[l(b_{max}; y) - l(b_1; y)]$$

$$D_0 \sim \chi^2(N - q);\ D_1 \sim \chi^2(N - p);\ \Delta D \sim \chi^2(p - q)$$

- In some cases where there are nuisance parameters , we eliminate nuisance parameters and form a new statistic. For example in the case of Normally distributed response variable

$$F = \frac{D_0 - D_1}{p - q} / \frac{D_1}{N - p} \sim F(p - q, N - p)$$

Chapter 6 : Normal Linear Models

- Form $Y_1, \ldots, Y_N$ are independent random variables. The link function is an indentity function, i.e., $g(\mu_i) = \mu_i$

$$E(Y_i) = \mu_i = x_i^T \beta\ ; Y_i \sim N(\mu_i, \sigma^2)$$

- Least Squares Estimate

$$b = (X^T X)^{-1} X^T y$$

- Variance-covariance matrix of the vector of residuals $\hat{\boldsymbol{e}}$

$$E(\hat{e}\ \hat{e}^T) = \sigma^2\ [I - X(X^T X)^{-1} X^T]$$

Hence standardized residuals are

$$r_i = \frac{\hat{e}_i}{\hat{\sigma}(1 - h_{ii})^{0.5}}$$

- Cooks distance

$$D_i = \frac{1}{p} (\boldsymbol{b} - \boldsymbol{b}_{(i)})^T X^T X (\boldsymbol{b} - \boldsymbol{b}_{(i)})$$

where $\boldsymbol{b}_{(i)}$ denotes the vector of estimates $b_{(i)}$, the estimate obtained by omitting the ith observation.

- Variance Inflation Factor

$$VIF_j = \frac{1}{1 - R_{(j)}^2}$$

where $R_{(j)}^2$ is the coefficient of determination obtained from regressing the jth explanatory variable against all the other explanatory variables

Chapter 7 : Binary Variables and Logistic Regression

• Form : The proportion of successes, $P_i = Y_i/n_i$, in each subgroup can be modeled as a GLM.

$$E(Y_i) = n_i\pi_i;\ E(P_i) = \pi_i$$

• Linear Model

$$\pi = \boldsymbol{x}^T\boldsymbol{\beta}$$

• Probit

$$\pi = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

• Logit

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_1 + \beta_2 x$$

• Complementary Log Log / Extreme Value distribution

$$\pi = 1 - \exp\left[-\exp\left(\beta_1 + \beta_2 x\right)\right]$$

• Deviance of a binomial model = Deviance $_{Saturated}$ -Deviance $_{Fitted}$

$$D = 2\sum_{i=1}^{N}[y_i log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i) log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)$$

$$D \sim \chi^2(N-p)$$

• Pearson chi-squared statistic

$$X^2 = \sum \frac{(o-e)^2}{e}$$

• Pearson chi-squared residual

$$X_k = \frac{(y_k - n_k\hat{\pi}_k)}{\sqrt{n_k\hat{\pi}_k(1-\hat{\pi}_k)}}$$

• Deviance Residual has an alternate form. Always use this for model diagnostics

• Likelihood ratio chi-squared statistic

$$Deviance_{Fitted} - Deviance_{Minimal}$$

$$C = 2[l(\boldsymbol{b}) - l(\boldsymbol{b}_{min})] \sim \chi^2(p-1)$$

• psuedo R squared

$$\frac{l(\tilde{\pi};y) - l(\hat{\pi};y)}{l(\tilde{\pi};y)}$$

which represents the proportional improvement in the log-likelihood function due to the terms in the model of interest, compared to the minimal model.

Chapter 8 : Nominal and Ordinal Logistic Regression

• Multinomial distribution can be regarded as the joint distribution of Poisson variables, conditional upon their sum n.

• Nominal logistic regression models are used when there is no natural order among the response categories. One category is chosen as reference

$$logit(\pi_j) = \log\left(\frac{\pi_j}{\pi_1}\right) = x_j^T\beta_j\ , for\ j = 2, \ldots, J$$

$$\hat{\pi_j} = \frac{\exp x_j^T\beta_j}{1 + \sum_{j=2}^{J}\exp x_j^T\beta_j}$$

• Chi-Squared Statistic

$$X^2 = \sum_{i=1}^{N}\frac{(o_i - e_i)^2}{e_i}$$

• Deviance

$$D = 2[l(b_{max}) - l(b)]$$

• Likelihood ratio chi-squared statistic

$$C = 2[l(b) - l(b_{min})]$$

• Psuedo R^2

$$R^2 = \frac{l(b_{min}) - l(b)}{l(b_{min})}$$

• Ordinal logistic regression models are used when there is a natural order among the response categories.

•Cumulative logit model :

$$\log\frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J} = x_j^T\beta_j\ , for\ j = 2, \ldots, J$$

• Proportional odds model.

$$\log\frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J} = \beta_{0j} + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1}$$

Chapter 9 : Count Data, Poisson Regression and Log-Linear Models

• Form :

$$E(Y_i) = \mu_i = n_i\theta_i\ ; \theta_i = \exp x_i^T\beta\ ;\ \ Y_i \sim Poisson(\mu_i)$$

• Link function ($\log n_i$ is the offset term)

$$\log\mu_i = \log n_i + x_i^T\beta$$

- Deviance of a Poisson model

$$D = 2\left[\sum y_i log\left(\frac{y_i}{\hat{y}_i}\right) - \sum (y_i - \hat{y}_i)\right] = 2\sum o_i \; log(o_i/e_i)$$

- Log Linear Model

$$\log E(Y_i) = constant + x_i^T \beta$$

- Log Linear Saturated Model

$$\log E(Y_i) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

- Log Linear Additive Model

$$\log E(Y_i) = \mu + \alpha_j + \beta_k$$

- Log Linear Minimal Model

$$\log E(Y_i) = \mu$$