1. Consider the 1996 American National Election Study data. The response is "sPID" that can be Democrat, Independent, or Republican. We consider the effect from "nincome" (continuous variable in \$1000). Below is the summary of the fit of a multinomial regression model with the variable "nincome" only.

```
mmodi <- multinom(sPID ~ nincome, nes96)

summary(mmodi)
# Call:
# multinom(formula = sPID ~ nincome, data = nes96)

# Coefficients:
            # (Intercept)    nincome
# Independent  -1.1749331 0.01608683
# Republican   -0.9503591 0.01766457

# Std. Errors:
            # (Intercept)     nincome
# Independent   0.1536103 0.002849738
# Republican    0.1416859 0.002652532

# Residual Deviance: 1985.424
# AIC: 1993.424
```

   (a) (5 points) Provide the interpretation of the coefficient -1.1749331 under "(Intercept)".

   (b) (5 points) Provide the interpretation of the coefficient 0.01766457 under "nincome".

   (c) (10 points) Provide the predicted proportions of each category when the value of "nincome" is 42.

   (d) (5 points) Suppose we fit instead with the following code:

   ```
   polr(formula = sPID ~ nincome, data = nes96)
   ```

   How many coefficients (including intercept) will be in the output?

2. (20 points) Suppose $\boldsymbol{X} = (X_1, \dots X_p)$ and $Y \in \{0, 1\}$. We want to choose a decision function $h(\boldsymbol{X})$ to minimize the risk according to the $0-1$ loss Function:

$$\mathcal{L}(y, h(\boldsymbol{X})) = \mathbf{I}(y \neq h(\boldsymbol{X})),$$

where $\mathbf{I}$ is the indicator function. With this loss function, the risk is

$$\mathcal{R}(h) = \mathrm{E}[\mathcal{L}(y, h(\boldsymbol{X}))] = \mathrm{P}(y \neq h(\boldsymbol{X})).$$

Show that the minimizer of $\mathcal{R}(h)$ is

$$h^*(\boldsymbol{x}) = \begin{cases} 1 & \mathrm{P}(y=1|\boldsymbol{X}=\boldsymbol{x}) \geq 1/2 \\ 0 & \text{otherwise} \end{cases} .$$

3. Consider the GLM with

$$f(y_i|\theta) = exp\left\{ \frac{1}{a_i(\phi)}\big(y_i\theta - b(\theta)\big) + c(y_i, \phi) \right\} \tag{1}$$

and with linear predictors $\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta}, i = 1, \dots, n$. Suppose ϕ is known and $\boldsymbol{x}$ are fixed.

(a) (30 points) Suppose we use the canonical link function. Show that in the MLE of $\boldsymbol{\beta}$, the Fisher Scoring algorithm can be implemented by IRLS.

(b) (15 points) Show that the equivalence of Fisher Scoring algorithm and IRLS is generally true for any proper link function $g(\cdot)$.

(c) (10 points) When the IRLS converges, why shouldn't we use the SE estimates of $\boldsymbol{\beta}$ from the LS fit in the last step? How to correctly estimate the SEs?