QUESTION - 2(A)
Summary of "Regression Models for Ordinal Data" by Peter McCullagh

Ordinal Data

- It is a type of Categorical data with a natural ordering, but the difference between them is uneven. These categories have a clear order but are not represented by numerical values. For example, the survey might have responses like "Very Satisfied", "Satisfied", "Neutral", "Dissatisfied," and "Very Dissatisfied". Here the difference between those responses would be uneven.
- The purpose is to investigate structural models appropriate to measurements on a purely ordinal scale. All models advocated in this paper share the property that the categories can be thought of as contiguous intervals on some contiguous scale.

Proportional Odds Model

- This model extends the principles of logistic regression to ordinal response variables. Understanding the odds of an outcome falling into a particular category.
- In logistic regression for binary outcomes, the log-odds of an event happening are modeled as a linear combination of predictor variables. The proportional odds model extend this concept to ordinal outcomes. Also called as Ordinal Logistic Regression.

$P(Y <= j \mid X) = 1 / (1 + exp(- (\alpha_j + \beta X)))$

Here,
$P(Y <= j \mid X)$ represents the cumulative probability of the response variable Y being less than or equal to category j given predictor variable X.
α_j represents the category specific intercept for category j
β represents a vector of coefficients for the predictor variable X

Assumption of Proportional odds model

- The odds of an observation falling into a higher category versus a lower category do no depend on the specific category being considered.
- This model is appropriate when you believe that the relationship between the predictors and the ordinal response is consistent across the ordered categories.

- Using MLE we can determin the parameter (α_j and β) of ordinal logistic regression

Proportional Hazard Model

- Used in survival analysis, in which we are analyzing the expected duration of time until some event occur.
- The hazard function or instantaneous risk function $\lambda(t; x)$, of major importance in the analysis of survival data, is defined to be the instantaneous failure probability at time t conditional on survival up to time.
- The hazard function describes the likelihood of an event happening at a specific point in time, given that it hasn't happened yet.

  $$\lambda(t; x) \; = \; \lambda_0(t)exp(-\; \beta^t x)$$

- $\lambda_0(t)$ Hazard function when x = 0

  β vecotr of unknown parameter

  $S(t; x)$ survival function [probability of surviving beyond time t given covariant]

  $$log(S(t; x)) \; = \lambda_0(t)exp(-\; \beta^t x)$$

  $$\lambda_0(t) \; = \; \int_0^t \lambda_0(S)\, dS$$

- Proportional hazard function: it implies that the effect of certain variables on the risk of the event happening remains consistent over time.

The ratio of the log survival function depends on the difference between covariate value x2 - x1

For discrete data, The hazard function is given as below.

$$-\; log(1 \,-\, \gamma_j(x)) \; = \; exp(\theta_j \,-\, \beta^t x)$$

Here,

$1 \,-\, \gamma_j(x)$ represents complementary probability / probability of survival beyond category j

To obtain appropriate linear structure analogue to linear logistic model

$$log(-\; log(1 \,-\, \gamma_j(x))) \; = \; (\theta_j \,-\, \beta^t x)$$

Complementary log-log transform

$$logP_1(x) \; = \; exp\;(mean\; of\; difference\; of\; log\; log\; scale) \; * \; logP_2(x)$$

Assumptions of Proportional hazard model

- It assumes that the relative hazard (risk) of the event is constant over time. It means that the hazard curves for different groups or variable are proportional and can be compared.

Properties of Related Linear Model

- The proportional odds and the proportional hazard models have the same general form

  $link\{\gamma_j(x)\} = \theta_j - \beta^t x$

- Here,

  $link$ represents logit or complementary log log function or any monotone increasing function mapping (0,1) onto (-∞,∞) .

- Logit transformation helps to map probabilities to a continuous scale, which makes it easier to work with linear model

- However, for data where the order of categories is meaningful, it's more important that the model recognize this order. In these cases, it should still give meaningful results even if you rever the order of the categories, but it shouldn't be completely flexible to random shuffling. This makes more sense for data with a clear ranking or order.
- Some models, like logistic, probit and inverse Cauchy, don't mind if you reverse the order of the categories. They will still work fine, but the results might just flip in direction. For example if you were analyzing 'good', 'average' and 'poor' performance, switching the order won't confuse these models.
- Some models handle ordered data better when you change the category order, while others might not do as well with this kind of change.

Ordinal Regression

Likelihood - It represents the probability of observing a particular set of ordered categories given the predictor variables. It assesses how well the model predicts the ordinal responses.

Odd Ratio - It compares the odds of an observation being in a higher category to a lower category based on the predictor variables.

How is it different from Multi-class classification?

- In multi-class classification, the likelihood represents the probability of a specific class given the predictor variables. It measures how well the model predicts the actual class labels.
- Instead of odds ratio, softmax is more commonly used in multi-class classification to compute the probabilities for each class.
- The likelihood and odds ratio are not directly applicable in the same way as in ordinal regression.

How is it different from the Regression problem?

- Regression problems involve predicting numerical values. However, the focus is on modeling the relationship between predictors and a continuous outcome.